University of Cape Town

School of Economics

ECO5000W - Masters in Economics Full Dissertation

September 17, 2022

---

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. This essay is my own work.
3. I have used the author-date convention style for citation and referencing. Each significant contribution to, and quotation in, this paper from the work(s) of other people has been attributed, and has been cited and referenced.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as their own work.

Author:

Samantha Culligan

CLLSAM004

Supervisors:

Dr Nicola Branson

Professor Murray Leibbrandt

Using Census, Institutional and Geospatial Data to Estimate the Socio-Economic Profile of Post-School Students by Institutional Type

Abstract

The socio-economic profile of students who are participating in post-school education; and the distribution of their socio-economic characteristics between universities and colleges, between institutions of a similar type, and within particular institutions is not well understood. Part of the reason for this is because potential data sets that could be used to answer this fall short on dimensions needed to fully explore the extent of socio-economic differences amongst student bodies by institutional type. I, therefore, generate a data set that draws on institutional, census, and geospatial information to estimate the socio-economic background of students' home postal code. Using this data set, I compare the mean statistic and generalised entropy index of a range of individual and household socio-economic postal code indicators for student bodies by institutional type to descriptively analyse their socio-economic profile. I show student bodies at traditional universities and Unisa appear socio-economically similar and display higher socio-economic circumstances than that of student bodies at comprehensive universities, universities of technology and TVET colleges who appear socio-economically similar. Between 2008 and 2019, the mean socio-economic profile declined for all student bodies, whereas there was no uniform trend for whether socio-economic heterogeneity was increasing or decreasing over time by university type. Lastly, my findings suggest there is more evidence for horizontal stratification between particular universities (regardless of their institutional type) rather than between university types, or between universities and TVET colleges.

Funding for this research is acknowledged from the National Research Fund (NRF) Chair in Poverty and Inequality Research, the Kresge Foundation (via the Siyaphambili Post-school Research Group within the Southern African Labour and Development Research Unit (SALDRU)), and the Spencer Foundation.

The GitHub repository contains the code (do-files written in Stata version 16) and documentation used for this research project.

Contents

1 Introduction **8**

2 The state of post-school education in South Africa under apartheid and why institutions merged in the early 2000s **12**

3 Conceptualising student institution choice **21**

4 Profiling universities and FET/TVET colleges from 2005 to 2019 **23**

4.1 Institutional characteristics, from 2005 to 2019 23

4.2 Student characteristics, from 2005 to 2019 34

4.3 Discussion of profiling observations 40

5 Measuring socio-economic circumstances of university and FET/TVET college students **42**

6 Generating a student-level data set with socio-economic information from the 2011 census **50**

6.1 Constructing the data set 50

6.2 Analysis of where students in the sample reside 55

6.2.1 Defining the sample 55

6.2.2 Provincial representation 58

6.2.3 Postal code representation 62

6.3 Caveats and considerations of the data set 67

7 Results and discussion **69**

7.1 Socio-economic profile of student bodies at different institutions 69

7.2 Heterogeneity estimates of students' socio-economic backgrounds within and between institutions 80

8 Concluding thoughts and reflections **89**

8.1 Summary 89

8.2 Implications for policy 92

8.3 Implications for future research 93

References **97**

Appendices **99**

A Terminology of education institutions in South Africa **99**

B Reform of higher education institutions in South Africa **100**

C Provincial representation **102**

D Constructing the socio-economic indicators **104**

D.1 Indicator for number of years of education 104

D.2 Indicator for annual household income 106

E Comparison of socio-economic measures **107**

F Choice of the generalised entropy index to measure socio-economic heterogeneity of student bodies **109**

F.1 Inequality axioms 109

F.2 Why GE(1) is best suited for my analysis 109

F.3 Decomposition of GE(1) estimates for socio-economic indicators by university type 111

F.4 Alternative inequality measures 113

G Ethics approval and plagiarism declaration **114**

List of Figures

1 Student headcount enrolment trends between 1990 and 2000 . . . 14
2 Student enrolment in universities and technikons between 1995 and 2000 . . . 15
3 Predicted and actual totals of matriculation exemptions between 1995 and 2000 . . . 15
4 Chapman model – Factors influencing students' choice of post-school institution . . . 21
5 Student headcount enrolment at universities and FET/TVET colleges . . . 24
6 Student headcount enrolment at FET/TVET colleges by province . . . 25
7 Student headcount enrolment by institutional type . . . 26
8 UGFTEN student headcount enrolment by institutional type . . . 27
9 Number of students with an NSC bachelors pass (year n-1) and number of UGFTEN students enrolled for a degree programme . . . 28
10 Number of UGFTEN students enrolled for a degree programme and number of UGFTEN students enrolled for a certificate or diploma programme . . . 29
11 Student headcount enrolments in qualification programmes by institutional type . . . 30
12 Students headcount enrolments at FET/TVET college by qualification programme . . . 31
13 Student headcount graduates in CESM fields by institutional type . . . 32
14 Student headcount enrolments by gender . . . 35
15 Share of enrolled students who identify as female by institutional type . . . 35
16 Student headcount enrolments by population group . . . 36
17 Population group composition of university types . . . 38
18 Nested geography hierarchy for the 2011 census . . . 51
19 Small Area Layer (SAL) polygons for South Africa . . . 53
20 Postal code polygons for South Africa . . . 53
21 Small Area Layer (SAL) polygons do not cover the full area of South Africa . . . 54
22 Number of university students by home province . . . 58
23 Distribution of university type enrolment by students' home province in 2019 . . . 59
24 Where are the university and TVET college campuses located? . . . 60
25 Distribution of university UGFTEN and TVET college students' home postal code in 2019 63
26 Mean socio-economic characteristics of student bodies' home postal code by university type from 2008 to 2019 . . . 76
27 Generalised Entropy (GE(1)) estimates for the socio-economic indicators of students' home postal code by university type from 2008 to 2019 . . . 84

List of Tables

1 Number of universities and technikons by responsible authority, prior 1994 . . . 13
2 University institutions by cross tabulation of historical type and institutional type . . . 18
3 Number of students and institutions in the FET college sector by province in 2001 . . . 19
4 University student enrolment by age group . . . 40
5 TVET college student enrolment by age group . . . 40
6 Number of university and FET/TVET college students recorded by DHET and the 2011 census . . . 43
7 Mean socio-economic characteristics using 2011 census . . . 46
8 Number of students with valid and invalid postal codes . . . 56
9 Share of university students that are South African . . . 57
10 Share of university students with invalid postal codes by university type . . . 57
11 Number of students by home province and number of students attending an institution within each province . . . 61
12 Number of years a postal code is the home postal code of at least one university student . 62
13 Number of students per postal code . . . 63
14 Number and share of postal codes represented by the number of students at university types 65
15 Share of postal codes represented by university from 2008 to 2019 . . . 66
16 Mean socio-economic characteristics of student bodies' home postal code by institutional type in 2019 . . . 70
17 Mean annual household income of student bodies' home postal code by university . . . 79
18 Generalised Entropy (GE(1)) estimates for the socio-economic indicators of students' home postal code by institutional type in 2019 . . . 82
19 Decomposition of the Generalised Entropy (GE(1)) estimates for the mean annual household income in students' home postal code by university from 2008 to 2019 . . . 88
B1 List of total 36 higher education institutions (prior merges) by historical type . . . 100
B2 Higher education institution merges between 2002 and 2005 . . . 101
C1 Estimated number of students by home province . . . 102
C2 Number of campuses per province by institutional type . . . 102
C3 Distribution of university type enrolment by students' home province in 2019 . . . 103
C4 Postal code range for each province . . . 103
D1 Highest level of education categories in 2011 census community profile data set and assigned number of years of education . . . 105
D2 Annual household income categories in the 2011 census community profile data set and imputed midpoint values . . . 106
E1 Comparison of socio-economic estimates using different geospatial weights for the analysis sample of university students in 2019 . . . 107
E2 Mean socio-economic characteristics by type of student body at universities in 2019 . . . 108

F1 GE(1) estimates for the share of individuals with a post-school qualification by university type 111
F2 GE(1) estimates for the share of individuals employed by university type 111
F3 GE(1) estimates for the mean annual household income by university type 111
F4 GE(1) estimates for the share of households with access to the internet by university type 112
F5 GE(1) estimates for the share of households with access to a computer by university type 112
F6 GE(1) estimates for the share of households with access to electricity by university type . 112
F7 Inequality measures for students' postal code mean annual household income in 2019 . . . 113

Acronyms

CESM Classification of Educational Subject Matter

CHE Council on Higher Education

DHET Department of Higher Education and Training

FET Further Education and Training

GIS Geographic Information System

GPS Global Positioning System

HEMIS Higher Education Management Information System

MIS Management Information System

NDP National Development Plan

NIDS National Income Dynamics Study

NSC National Senior Certificate

SAL Small Area Layer

TVET Technical and Vocational Education and Training

TVETMIS Technical Educational Management Information System

UGFTEN Undergraduate, First-Time Entering

USAf Universities South Africa

1 Introduction

In South Africa, learning that takes place outside of basic education is referred to as post-school education.[1] Post-school education tends to be positioned in policy discussions as a vital component to tackling some of South Africa's biggest socio-economic challenges, such as alleviating poverty and reducing inequality (National Planning Commission of South Africa, 2012). Empirical evidence suggests unemployment rates are lowest, and wages earned are highest, for individuals with at least some years of post-school education (Lam et al., 2015). For a country like South Africa, with a high rate of unemployment (30.1% Q1:2020, Statistics South Africa) and a large discrepancy in wages earned,[2] attainment of post-school education is arguably one approach to help more people compete for quality jobs in the labour market and, in turn, lead a better quality of life.

Yet, a relatively small share of South Africa's population reach post-school education[3] and there lacks a thorough understanding of who these students are — what is their socio-economic profile and at which type of post-school institution they enrol. Furthermore, little is known about the distribution of their socio-economic characteristics between universities and colleges, between institutions of a similar type, and within particular institutions. The Department of Higher Education and Training (DHET) typically categorises students enrolled in its institutions by population group, gender and age as a way to measure and compare equitable enrolment over time. By taking into account students' socio-economic background using a more extensive set of measures, the DHET can broaden their understanding of the profile of students within and across institutions in the post-school system. This will provide empirical evidence to better inform policy development and responses, student application decisions, admission policies, affirmative action, financial aid structures, tuition fees and the distribution of government funding across institutions.

However, assessing post-school students' socio-economic background is not a straightforward exercise because a single data set with sufficient information to undertake this analysis does not exist. Student-level institutional data does not collect socio-economic information. Alternatively, South Africa has nationally representative survey data (such as the Census), which collects information on individuals and their households, but lacks information on the specific post-school institution enrolled students attend. Hence, in this paper, I describe how I build a novel data set to overcome this challenge and I offer an empirical analysis of how I use this data set to profile the socio-economic characteristics of post-school students at different institutional types.[4]

[1] See Appendix A for a discussion of the terminology used for education institutions in South Africa.

[2] South Africa has persistently experienced high levels of income inequality with a household income per capita Gini coefficient of 0.66 in 2014, down from 0.68 in 1993 (Leibbrandt et al., 2018).

[3] Attainment of post-school qualifications for the working age population (15 to 64 years of age) in the country has risen to 14% in 2020, up from 7% in 1994 (Branson et al., 2020).

[4] The institutions of focus in this paper are public universities (particularly undergraduate, first-time entering students in the analysis) and FET/TVET colleges. It is acknowledged that there are other types of post-school institutions in the post-school system, such as community colleges, private colleges and private universities. However, public universities and FET/TVET colleges are the institutions of focus in this analysis since they account for the largest share of enrolments in the post-school education system (77% in 2019, Department of Higher Education and Training (2021)), their public redistributive element (rather than pure privately financed education), and data availability.

To construct this data set, I first calculate socio-economic indicators for each Small Area Layer (SAL), which is the lowest publicly available geographic unit area in the 2011 census. I then link these SAL socio-economic indicators to student-level institutional data via students' home postal code. This involves calculating the area of a postal code and its share that falls within each SAL. The shares are then used to weight the SAL socio-economic indicators to calculate postal code socio-economic indicators. Lastly, I assign the postal code socio-economic indicators to the student-level institutional data as a proxy for the students' home socio-economic background. From there, the research questions I set out to answer with this data set are:

1. Where do undergraduate, first-time entering public university and FET/TVET college students reside, according to their home postal code? Has this changed between 2008 and 2019?

2. What is the socio-economic profile of post-school institutions, based on the home postal codes of their student body? How different are student bodies by institutional type?

3. Have post-school institutions' student bodies become more heterogeneous by socio-economic profile since 2008? Which types of institutions are contributing to increased levels of heterogeneity?

The analysis for these research questions contributes to the knowledge of South Africa's post-school education system in three respects. First, this research offers an approach to estimating post-school students' socio-economic background. I estimate the socio-economic profile of students' home postal code by weighting the 2011 census information at the SAL level by the share of the postal code area that falls within each SAL area. Second, the analysis offers a time dimension with 12 years of data from 2008 to 2019 for public university students. By restricting the sample to undergraduate, first-time entering students, I am able to isolate each new cohort of students entering post-school education and assess whether the socio-economic profile of student bodies has changed over time – given that I hold the socio-economic data fixed (representative of the 2011 census). With this time dimension, my findings suggest that the mean socio-economic profile of students across all university types has declined over time and that there is no uniform trend in whether socio-economic heterogeneity is increasing or decreasing. Third, I compare both university and FET/TVET college student bodies for the year 2019 (the latest year of available data). Research that collectively assesses universities (as well as their types) and FET/TVET colleges together is rare as research on these two distinct institutional types tend to be studied in isolation. This enables me to observe that student bodies at FET/TVET colleges are actually relatively similar to student bodies at comprehensive universities and universities of technology.

Part of building this descriptive analysis of the socio-economic profile of student bodies at different university types and FET/TVET colleges involves recognising that there are a host of factors at play that affect why a student may enrol at one institution over another. Furthermore, South Africa's history of apartheid, and prior, has undeniably shaped the relationship between institutional and student characteristics. What became evident was that the fragmented higher education and FET/TVET college sectors formalised under apartheid legislation were no longer sustainable. During this period there was a wave of policy documents, which articulated the need for the post-school education system to

increase access and broaden participation by undoing the entrenched inequalities of the past post-school system under apartheid, and to produce trained graduates who would help build the social and economic infrastructure of the newly democratic state. An aspect of the path forward involved the merging of institutions to reduce the number of institutions and giving rise to new institutional identities. All of which, in turn, influence student enrolment patterns at institutions. Even if post-school institutions are more accessible today for all population groups, the extent to which institutions are stratified by students' socio-economic circumstances in the reformed post-school era (post-2005) is not well documented.

Cooper (2015) argues that the categorisation of institutions under the reformed post-school education system in the 2000s disguises socio-economic inequalities based on which institutions did and did not merge. Those that did not merge were predominantly well resourced, historically White institutions. Hence, while the implementation of mergers appears as if there has been a systemic change, it is not clear to what extent it has shifted the socio-economic profile of students bodies relative to the explicit segregation under apartheid. To test various hypotheses about students' socio-economic background and ascertain to what extent social justice is being fulfilled, Cooper (2015) urges for a national survey of all students. However, it is difficult to gather a nationally representative sample of students as survey response rates tend to be low. Surveys can also be costly to conduct. Therefore, I choose to offer the next best alternative by drawing on the strengths of a collection of existing South African data sets – namely census, institutional and geospatial data.

The significance of this research is not only about understanding the socio-economic segregation in the post-school system but also about the long term implications of this segregation. There is a need for this documentation because, ultimately, the inequalities that play out in post-school institutions are later magnified in the labour market. Pellicer and Ranchhod (2012) note that an inequality trap works through post-school education as it is post-school education (explicitly 12 or more years of education) where attainment is low but wage returns are substantially greater relative to wage returns to incomplete secondary school. Thus, this induces high levels of income inequality. Given that research has found a positive correlation between the income of parents and the income of their adult offspring, these high levels of income inequality are passed down from one generation to the next and persist over time (Solon, 1999; Black and Devereux, 2011; Corak, 2013). This translates into low rates of educational and economic mobility. Rather than post-school institutions ameliorating the high levels of socio-economic inequality weighing on the country, as positioned in national policy discussions,[5] they are to some extent perpetuating it if post-school institutions are admitting students primarily from a more advantaged socio-economic background. Therefore, an empirical analysis of the profile of post-school student bodies will provide evidence to better inform post-school education policy planning and responses in an effort to achieve more equitable outcomes in society.

The following section discusses the repercussions apartheid had for the country and its education system. Thereafter, Section 3 discusses the conceptual framework, namely the student institution choice literature. I choose to draw on the combined model approach which factors in both: (1) characteristics of

[5] The National Planning Commission of South Africa (2012) highlights skills development and innovation as one of the top 3 priorities to eliminate poverty and reduce inequality in the National Development Plan by 2030.

the institution; and (2) characteristics of the student. Collectively, these factors contextualise who enrols in post-school education and in which institution they enrol. I profile these institutional and student characteristics for different institutional types and their students in Section 4. What is absent from this profiling is a discussion of the current methods, and respective findings, of measuring post-school students' circumstances. I discuss this in Section 5, acknowledging the prior work by Van Broekhuizen et al. (2016), Kerr et al. (2016), and Branson and Kahn (2018). Section 6 and 7 discuss how I build the student-level data set with socio-economic information in order to profile the socio-economic circumstances of student bodies at different institutional types. Section 8 closes with a summary, what this research means for policy, and where this research is being applied in future work.

2 The state of post-school education in South Africa under apartheid and why institutions merged in the early 2000s

The Extension of University Education Act (Act No. 45 of 1959), which augmented the 1953 Bantu Education Act, segregated post-school institutions by establishing separate higher education institutions for individuals of colour. This act formed part of the apartheid legislation and made it illegal for a student to register at an institution that was designated for a different population group and ethnicity to that of their own without the written consent of the Minister of Internal Affairs (Lapping, 1987).

A new constitution for the Republic of South Africa, introduced in 1984, further entrenched the racial segregation (Bunting, 2006b) by dividing South Africa into five separate legislative and geographic entities: The Republic of South Africa and four "independent republics" (the Republic of Transkei, the Republic of Bophutatswana, the Republic of Venda and the Republic of Ciskei, namely the TBVC homelands). The TBVC homelands were predominantly poor with limited employment opportunities, whereas White individuals living in the Republic of South Africa experienced a relatively higher standard of living. These authorities were each responsible for a number of higher educational institutions, as outlined in Table 1 (Bunting, 2006b). White individuals had access to 19 out of the total 36 higher education institutions – despite White individuals making up approximately 13.43% of the population (inclusive of the TBVC homelands) according to the 1991 census estimates (Khalfani and Zuberi, 2001). Not only did Black, Coloured and Indian population groups have access to fewer higher education institutions, but the institutions were also more likely to be located in remote outlying areas (Chisholm, 1992). Hence, the incoming democratic government inherited a fragmented higher education sector where access to institutions, and quality of learning within these institutions, was inequitably distributed amongst racial population groups.

From the late 1980s, the weakened apartheid state no longer had the political will to enforce the racial barriers to entry at higher education institutions (Gibbon and Kabaki, 2006). There was no formal legislation to mark this transition, yet the number of Black student enrolments started to increase at institutions previously designated White, Indian, and Coloured. Hence, between 1990 and 1994, the higher education sector was met with a new wave of optimism and student enrolments at universities and technikons grew by more than 130,000 (33%) in 1994 relative to 1990, as depicted in Figure 1 (Bunting, 2006a).

Amongst the universities, historically Black universities grew by 28,000 (37%) and the historically White universities grew by 10,000 (8%) between 1990 and 1994. The technikons also experienced rapid growth, especially in the distance technikon, namely Technikon South Africa. Between 1990 and 1994, student enrolments in Technikon South Africa grew by 38,000 (126%), historically Black technikons increased enrolments by 11,000 (60%), and historically White technikons expanded by 19,000 (41%) (Bunting, 2006a). This growth in student enrolments, between 1990 and 1994, in the higher education sector shaped the context on which the policy recommendations by the National Commission on Higher Education (NCHE) in their report, *A Framework for Transformation* (1996), and by the Department of

Education in *Education White Paper 3* (1997) were based.[6]

Table 1: Number of universities and technikons by responsible authority, prior 1994

Responsible authority	Universities	Technikons	*Total*
Republic of South Africa			
House of Assembly (for White individuals)	11	8	19
House of Representatives (for Coloured individuals)	1	1	2
House of Delegates (for Indian individuals)	1	1	2
Department of Education and Training (for Black individuals)	4	2	6
Independent republics			
Republic of Transkei	1	1	2
Republic of Bophuthatswana	1	1	2
Republic of Venda	1	0	1
Republic of Ciskei	1	1	2
Total	**21**	**15**	**36**

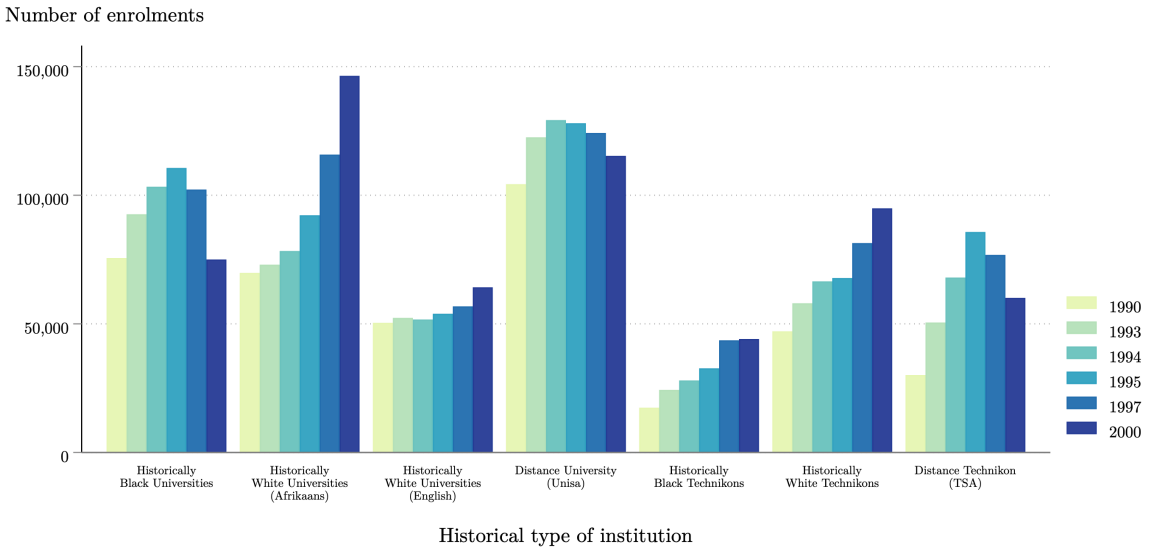
Source: Bunting (2006a).
Notes: Table B1 contains the list of the total 36 higher education institutions.
The 1984 constitution for the Republic of South Africa divided the national parliament into three chambers (the "tricameral" parliament): one house for representatives of White voters (the House of Assembly), one for representatives of Coloured voters (the House of Representatives), and one for representatives of Indian voters (the House of Delegates) (Bunting, 2006a). As the Black population group did not have its own separate parliamentary house, education for Black individuals was considered a general affair and the responsibility of the Department of Education and Training.

One of the predominant recommendations made by the National Commission on Higher Education was to "massify" the higher education system (Maassen and Cloete, 2006). In other words, higher education should be widely accessible. It was believed that this expansion in participation would undo the past inequalities in access to institutions and produce the high-skilled citizens necessary for South Africa's economic growth in the post-apartheid era. Given the growth in student enrolments between 1990 and 1994, this seemed like a reasonable recommendation at the time.

Government funding for higher education institutions was, and still is, largely based on the number of students enrolled (Bunting, 2006a). The increase in student enrolment between 1990 and 1994, therefore, produced expectations that government funding would grow in the upcoming years. However, as the Department of Education started to adopt the policy recommendations made by the National Commission on Higher Education, the system was met with an unanticipated plateau, and even decline, in enrolment numbers. From 1995 to 2000, the only institutions that continued to experience growth in student enrolment numbers were historically White universities (Afrikaans- and English-speaking) and historically White technikons, as shown in Figure 1. The historically Black technikons experienced growth

[6] The NCHE set in motion the policy process to reform the fragmented, racially divided post-school education system into a co-ordinated and equitable one. What followed was a series of draft papers and subsequent Acts for each respective sector. For the higher education sector this included the Green Paper on Higher Education in December 1996, the Draft White Paper on Higher Education in April 1997, and the Education White Paper 3 (A Programme for Higher Education Transformation) in August 1997 - which collectively led to the Higher Education Act 101 of 1997.

Figure 1: Student headcount enrolment trends between 1990 and 2000



Source: Bunting (2006a).
Notes: This figure shows historically Black universities, the distance university and the distance technikon experienced a decline in the number of headcount student enrolments post 1995. Years of data presented are subject to data availability. See Table B1 for a list of the universities and technikons in each historical type category.

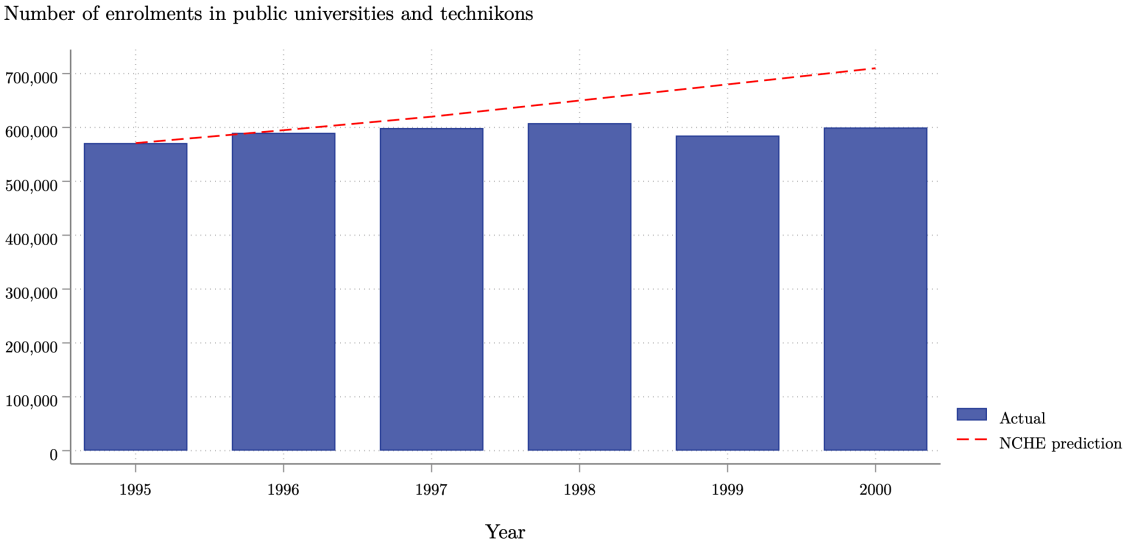
between 1995 and 1997, but this growth stagnated between 1997 and 2000. Thus, higher education institutions competed with one another to increase their number of student enrolments and secure government funding (Bunting, 2006a). Simultaneously, the Higher Education Act (1997) also provided for the existence of private post-school institutions, which further increased competition for student intake. The increased competition was not something the government policies, nor the institutions themselves, foresaw emerging because they expected a larger number of students to be entering the system, based on the growth experienced between 1990 and 1994, than what actually occurred.

Figure 2 compares the actual enrolment numbers in the higher education sector against the National Commission on Higher Education's forecast of enrolment numbers between 1995 and 2000. By 1998, it became evident that the student enrolments in universities and technikons were not meeting the National Commission on Higher Education's growth predictions, which were over-optimistic. Not only were there almost 140,000 fewer student enrolments in 1998 and 1999 compared to what the National Commission on Higher Education predicted; but enrolments fell between 1998 and 1999 by 23,000 (4%). Enrolments then rose by 15,000 (3%) in 2000 due to the sharp growth in distance programme enrolments at some of the historically White universities with Afrikaans as the language of instruction.

One of the factors that explains why the student enrolment growth model was so poorly predicted by the National Commission on Higher Education was due to the schooling system not producing the expected number of grade 12 school-leavers with a matric exemption.[7] The National Commission on

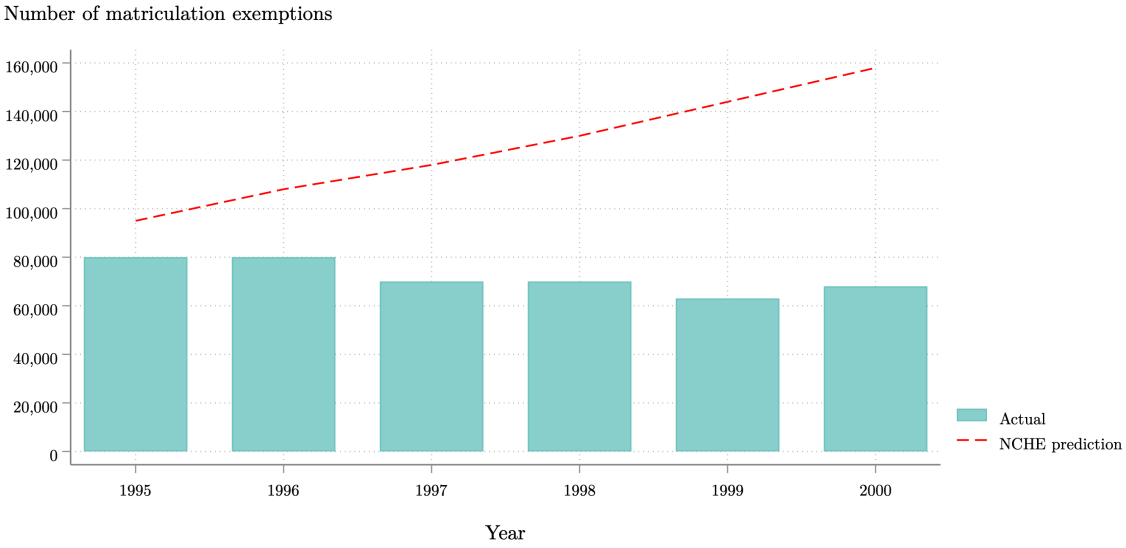
[7] "A matric exemption is a legal requirement for the first-degree study at a South African university." – Universities South Africa (USAf) (2020).

Figure 2: Student enrolment in universities and technikons between 1995 and 2000



Source: Cloete and Bunting (2000).
Notes: This figure show the actual number of headcount enrolments for public universities and technikons, compared to the predicted number of head count enrolments by the National Commission on Higher Education (NCHE). The NCHE predicted a higher number of enrolments than what was realised. Notably, the number of enrolments even declined in 1999.

Figure 3: Predicted and actual totals of matriculation exemptions between 1995 and 2000



Source: Cloete and Bunting (2000).
Notes: This figure shows the actual number of matriculation exemptions awarded, compared to the predicted number of matriculation exemptions by the National Commission on Higher Education (NCHE). Not only did the NCHE over estimate the number of matriculation exemptions awarded, but the actual number of exemptions declined from 1995 to 2000.

Higher Education had anticipated an increase in the number of matriculation exemptions given the end of apartheid (Bunting, 2006a). Instead, the number of matriculation exemptions declined between 1995 and 2000 (Cloete and Bunting, 2000). Not only did the number of matriculation exemptions decline, but the schooling system also produced 320,000 fewer matriculation exemptions than the forecast calculated by the National Commission on Higher Education predicted for this period. These data are presented in Figure 3, which compares the actual number of matriculation exemptions against the National Commission on Higher Education's forecast of matriculation exemption numbers between 1995 and 2000. The number of matriculation exemptions decreased the most in 1999 (by 17,000) relative to 1995. It is evident in Figure 3 that the National Commission on Higher Education grossly over estimated the number of matriculation exemptions – anticipating more than double the number of matriculation exemptions in 2000 than was actually realised.

Therefore, overall enrolment numbers were not meeting expectations – despite this commitment to massification of the higher education system. The number of enrolments at historically Black universities particularly struggled between 1995 and 2000, declining back to the same number of enrolments they had in 1990. The policy aspirations of increased and broadened participation for these universities had not materialised. Students who would have traditionally attended the historically Black universities now had the choice to enrol at historically White universities. This was to the particular advantage of historically Afrikaans-speaking White universities who continued to experience steady growth in enrolments throughout the 1990s, as depicted in Figure 1. Cloete (2006) argues that students of colour embraced these "privileged" universities, which they were previously excluded from, in rejection to the universities established for them under apartheid legislation. Thus, the post-apartheid era (between 1994 and 2000) brought about a significant improvement in student institution choice but lacked the necessary systemic redress as the gap between enrolments at historically Black institutions and historically White institutions expanded (Cloete, 2006).

During this period of unforeseen enrolment challenges emerging in the system, the Higher Education Act (No. 101 of 1997) provided the legal legislation for the merging of university and technikon institutions, amongst a number of other policy changes outlined in the act. One of the other notable contributions of the act was the establishment of the Council on Higher Education (CHE), which is an independent statutory body.[8] Under the Higher Education Act (1997), the Minister of Education - after consulting the CHE - was allowed to merge two or more universities and/or technikons into a single, university institution. The Higher Education Act (1997) also allowed for the closure of any universities or technikons.

By 2000, the CHE established a *Size and Shape Task Team* to develop a set of concrete strategies for the reform of the higher education sector (Council on Higher Education, 2000). The CHE went on to advise a reduction in the "present number of institutions through combining institutions" (Council on Higher Education, 2000). This recommendation by the CHE was based on their view that the structure of the higher education system developed under apartheid legislation was no longer sustainable. The sector

[8] The CHE also serves as the Quality Council for Higher Education as amended by the National Qualifications Framework Act (No. 67 of 2008).

lacked the necessary financial resources; leadership, management and administrative roles were absent in parts of the system; and student enrolment trends were declining at historically Black universities, historically Black technikons, and the distance university and technikon.

Furthermore, the National Plan for Higher Education (2001) provided the framework on how to implement the vision for the reform of the post-school education system, as outlined in the Education White Paper 3 (1997). This report called attention to the importance of institutional collaboration in the post-apartheid era. A key underlying purpose of the reformed post-school system is to no longer have institutions whose identity is based on a population group (Black or White) or language (English- or Afrikaans-speaking). Therefore, in transforming the system to a nationally co-ordinated one, building new institutional identities and cultures is imperative.

Collectively, these policy discussions and developments led to the restructuring of the higher education sector via merging institutions, incorporating campuses, and institution name changes between 2002 and 2005.[9] This process consolidated the number of higher education institutions (universities and technikons) from 36 to 23 universities. The merging of universities and technikons gave rise to a new way of categorising higher education institutions in the post-school education system, which has become the official government categorisation of the public universities (Cooper, 2015). The three distinct university types are: (1) comprehensive universities, (2) universities of technology, and (3) traditional universities (Republic of South Africa, 2002).

Comprehensive universities were formed based on the merging between universities and technikons; or due to the expansion of existing universities in rural areas where there were previously no technikons to offer vocationally orientated programmes (Essop, 2020). In principle, comprehensive universities offer a combination of theoretically and vocationally orientated programmes (Gibbon, 2004). Technikons that did not merge with a university are no longer called technikons. Rather, technikons are now called universities of technology. Universities of technology offer vocationally orientated programmes, and traditional universities offer theoretically orientated programmes.

Although an intention of this new way of categorising universities is meant to spark new institutional identities, Cooper (2015) argues that some of the structures of inequality in the reformed post-school education system are already built in, based on which institutions did and did not merge. For example, some institutions that were excluded from the higher education mergers include: four large historically White institutions, two smaller historically White institutions, and one historically Coloured institution. By comparison, the institutions included in the mergers were mainly technikons and historically Black institutions, as displayed in Table B2.

Table 2 presents the cross tabulation of the historical type and the institutional type classification. Under the institutional classification, there are 8 comprehensive universities, 6 universities of technology, and 12 traditional universities (inclusive of the new universities introduced in 2014/5). Ten of the existing 26 universities are the result of a merge between institutions and are distributed evenly amongst the institutional types (plus the distance university, Unisa, which is classified as a comprehensive university). Half of the traditional universities were historically White; whereas only two universities of technology

[9] Table B2 outlines the higher education merges that took place between 2002 and 2005.

(and no comprehensive universities) were historically White.

Table 2: University institutions by cross tabulation of historical type and institutional type

	Comprehensive universities	**Technology universities**	**Traditional universities**
Historically White (historically advantaged)		Central University of Technology Vaal University of Technology	Rhodes University University of Cape Town University of Pretoria University of Stellenbosch University of the Free State University of Witwatersrand
Historically Black (historically disadvantaged)	University of Venda University of Zululand	Mangosuthu University of Technology	University of Fort Hare University of Western Cape
Merger			
	Nelson Mandela University University of Johannesburg Walter Sisulu University	Durban Institute of Technology Tshwane University of Technology Cape Peninsula University of Technology	North West University University of KwaZulu-Natal University of Limpopo
Distance	University of South Africa		
New[a]			
	Sol Plaatje University University of Mpumalanga		Sefako Makgatho Health Science

Source: Bunting (2006b) and Wildschut et al. (2018).
Notes: Table B2 lists which institutions merged with each other and the university that they formed.
[a] Sol Plaatjie University and the University of Mpumalanga were established in 2014. Sefako Makgatho Health Science University was established in 2015. Institutional type classification for these new institutions are as per categorised by Wildschut et al. (2018).

Higher education institutions were not the only branch of post-school education to merge institutions. The White Paper on Further Education and Training (1998a) characterises the Further Education and Training (FET) sector (commonly referred to at the TVET sector from 2013) as fragmented, poorly co-ordinated and inefficient. Programmes within the FET sector experienced high failure and repetition rates, which placed stress on the financial sustainability and physical infrastructure of the sector. Given these challenges, the technical colleges underwent a process of consolidation too (Department of Education, 2003).[10]

In 2001, the 152 technical colleges were officially merged into 50 larger, multi-campus, FET colleges (Department of Education, 2005a). Table 3 outlines the number of students and the number of

[10] Similarly to the NCHE, the National Commission on Further Education and Training (NCFE) was set up in 1996 to investigate the state of the FET sector. The FET sector had their own Green Paper on Further Education and Training (1998), and Education White Paper 4 (A Programme for the Transformation of Further Education and Training) in September 1998a. These papers gave rise to the Further Education and Training (FET) Act 98 of 1998 for FET colleges, analogous to the Higher Education Act (1997) for universities and technikons. The FET Act (1998b) legally allowed for the merge of two or more public FET institutions into a single FET institution; as well as the closure of any FET institutions. The FET Act (1998b) was later repealed by the Continuing Education and Training Act (previously known as the FET Colleges Act) 16 of 2006. The National Plan for FET colleges (2008) then served as the mechanism to carry out the governance framework discussed in the Education White Paper 4 (1998a).

institutions (before and after the merging of technical colleges) by province in 2001. Unlike the universities (and technikons), which can be grouped by historical type (historically White or Black) or university type (comprehensive universities, universities of technology, or traditional universities), FET colleges do not have any similar categorisation. Rather, they tend to be grouped by province in government reports.

Before the merging of technical colleges, the number of institutions in each province ranged from 7 to 33 institutions. Thereafter the merging of technical colleges, this range was reduced to between 2 to 9 institutions per province. This wide variation in the number of institutions per province reflects the spatial segregation that occurred under apartheid and the deprivation of FET colleges in former homelands or predominantly Black areas (Department of Education, 2003).

In 2001, almost a third (32.40% or 115,377 students) of all FET college students attended an FET college based in Gauteng – the province which is home to South Africa's largest city (Johannesburg) and the capital (Pretoria). KwaZulu-Natal is the province with the second highest number of students attending an FET college within the province but is less than half that of Gauteng. Furthermore, the mean number of students per FET college in Gauteng resembles somewhat of an outlier at 14,422 students per FET college. This is more than double the approximate mean of 4,000 to 7,000 students per FET college amongst the other provinces. By consolidating the number of FET colleges, there are naturally more students per FET college (across all provinces) than there would have been otherwise. The intension here is that fewer FET colleges, despite administrating more students, will improve co-ordination and efficiency between institutions. In addition, each reformed FET college has a chance to strengthen the programmes they offer.

Table 3: Number of students and institutions in the FET college sector by province in 2001

Province	Number of students	Number of technical colleges (before merging)	Number of FET colleges (after merging)	Mean number of students per FET college
Eastern Cape	40,499	28	8	5,062
Free State	29,169	11	4	7,292
Gauteng	115,377	33	8	14,422
KwaZulu-Natal	55,259	24	9	6,139
Limpopo	29,082	15	7	4,155
Mpumalanga	20,529	10	3	6,843
North West	20,036	11	3	6,679
Northern Cape	8,231	7	2	4,116
Western Cape	37,867	13	6	6,311
National	356,049	152	50	

Source: Department of Education (2003).
Notes: This table shows the number of Further Education and Training (FET) students in 2001. Prior to the consolidation of institutions, these institutions were called technical colleges and then became known as FET colleges. The table lists the respective number of these institutions (before and after merging), as well as the mean number of students per FET college.

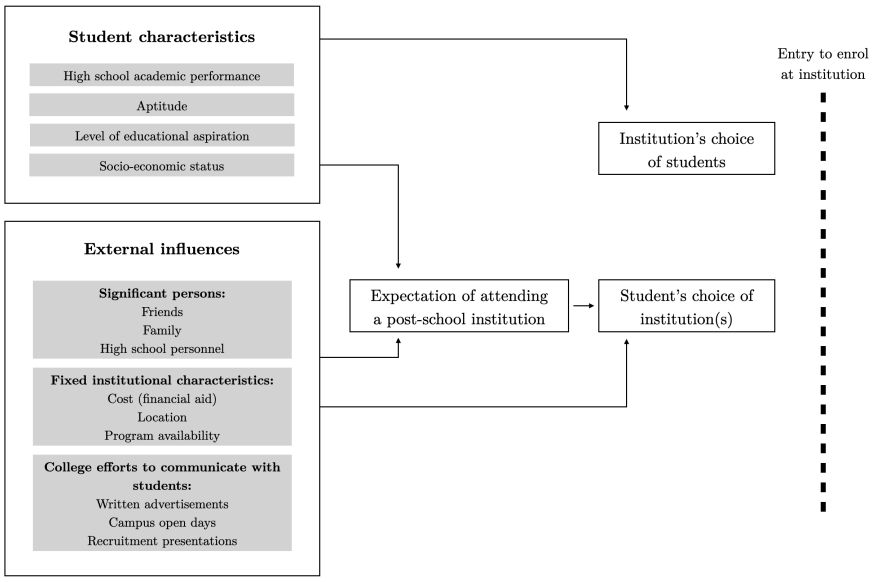
This section has highlighted how higher education and FET institutions under apartheid legislation explicitly segregated the post-school system along racial lines. Towards the fall of apartheid, this legislation could no longer be enforced and students gained greater institution choice. However, overall enrolment levels fell short of exceptions and the gap between enrolments at historically Black institutions and historically White institutions widened, as was shown in Figure 1. In addition, the FET White paper (1998a) referred to FET colleges as fragmented, poorly co-ordinated and inefficient. Despite students no longer being excluded from institutions based on the colour of their skin, both of these institutional types lacked the necessary systemic change to function as a truly national system of education and training. This prompted the merging of institutions in policy documents to consolidate the number of institutions, and build a more coherent and efficient system. Along with these expectations, the merging also allowed new institutional identities, in theory, to form. To what extent have these new institutional identities influenced student institution choice in the reformed post-school system?

3 Conceptualising student institution choice

There are an array of factors that affect how students sort into different types of institutions and impact the end enrolment point I assess. Various models have attempted to provide an understanding of the choices students make when deciding whether to attend a post-school institution and which institution to attend. This includes: (1) economic models, (2) status-attainment models, and (3) combined models such as the Jackson model, the Chapman model, and the Hanson and Litten model (Vrontis et al., 2007). These models are based on factors that influence the decisions of prospective students entering post-school institutions straight out of secondary school (age 15 to 24). Certain factors may also influence students who are of an older age but variables that are more likely to apply to older students, such as career advancement and family opportunity trade-offs, are not reflected in these models.

Economic models weigh up the costs and benefits, which have a measurable value, of attending a post-school institution. For students who choose to enrol in a post-school institution, they will rationally pick the institution which offers the highest value to them. Whereas the economic models are based on rational decision making of the highest utility value, status-attainment models assess the interaction of the student's socio-economic characteristics to understand their choice of post-school institution. These background variables are developed over the course of students' lives. Combined models use both the rational decision-making aspect of the economic models, as well as the sociological aspect of the status-attainment models. It is suggested that combined models offer more explanatory power over either the economic or status-attainment model (Hossler et al., 1999).

Figure 4: Chapman model – Factors influencing students' choice of post-school institution



Source: Chapman (1981).
Notes: The Chapman model was designed at a time when there was intense competition for students across post-school institutions. It is stated that institution administrators were operating from a point of incomplete understanding of the complex set of factors at play influencing students' institution choice (Chapman, 1981).

One example of a combined model is the Chapman (1981) model, presented in Figure 4. For a student to enrol at a post-school institution, the student needs to apply to that institution and the institution needs to accept the student's application. Whether and where a student decides to apply is influenced by their expectation of attending that institution. The Chapman model suggests that a student's expectation of attending a post-school institution, as well as their final choice of institution, is influenced by a set of student characteristics and external factors. Whether an institution accepts or rejects a student's application is only based off of the student's characteristics. Student characteristics include the student's socio-economic status, aptitude, level of educational aspirations, and academic performance. On the other hand, external influences include significant persons in the student's life, fixed institution characteristics, and institutions' efforts to communicate with prospective students. The comments and advice from friends and family weigh on students' choice of institution. Students' choice may also be influenced by the institution friends and family have attended themselves. Location, tuition fees, campus environment, type of institution and programmes offered by the institution are relatively fixed by the institution. Hence, the student institution choice models acknowledge that the sorting of students into different types of institutions is not random.

I am most interested in the group of students who are enrolled at a post-school institution and at which institution they have enrolled. Hence, this is a select group of students who have chosen to apply to at least one post-school education institution, the institution has admitted them (and was possibly offered a financial aid award), and the student has accepted the offer to enrol. Amongst this select group of students who reach this point in the education pipeline, these students sort themselves across the institutional types. Each of these students are defined by a unique set of characteristics and interests that influence their choice of institution. I am then curious about analysing the collective socio-economic profile of these enrolled students at different institutional types once they have enrolled. To what extent do they differ? Has this changed over time?

4 Profiling universities and FET/TVET colleges from 2005 to 2019

Guided by the holistic set of factors outlined in the combined model, I profile the characteristics of institutions and the characteristics of the students enrolled at different types of institutions from 2005 to 2019. This period represents a reformed era for the post-school education system. I draw on existing institutional data published in the public domain to note the shifts and patterns in the number of enrolments, the types of qualifications that are offered, and the fields of specialisation at different institutions; as well as how the composition of enrolled students has changed over time by gender, population group, and age group at these institutions. Collectively, I am observing the systemic changes in the institutional structure and how greater student choice of institutions has influenced the profile of student bodies (in terms of their fixed characteristics) at these newly established institutional identities.

4.1 Institutional characteristics, from 2005 to 2019

By 2005, the number of universities and FET/TVET colleges had been consolidated but the intension remained to grow the number of students enrolled within these institutions. As shown in Figure 5, between 2005 and 2019, the number of students enrolled at universities increased. Since 2015, the number of students enrolled at FET/TVET colleges decreased. Figure 5 also shows more students are enrolled at universities relative to FET/TVET colleges, even though there are only 26 universities (as of 2015) and 50 FET/TVET colleges. Hence, universities have larger student bodies within a particular institution.

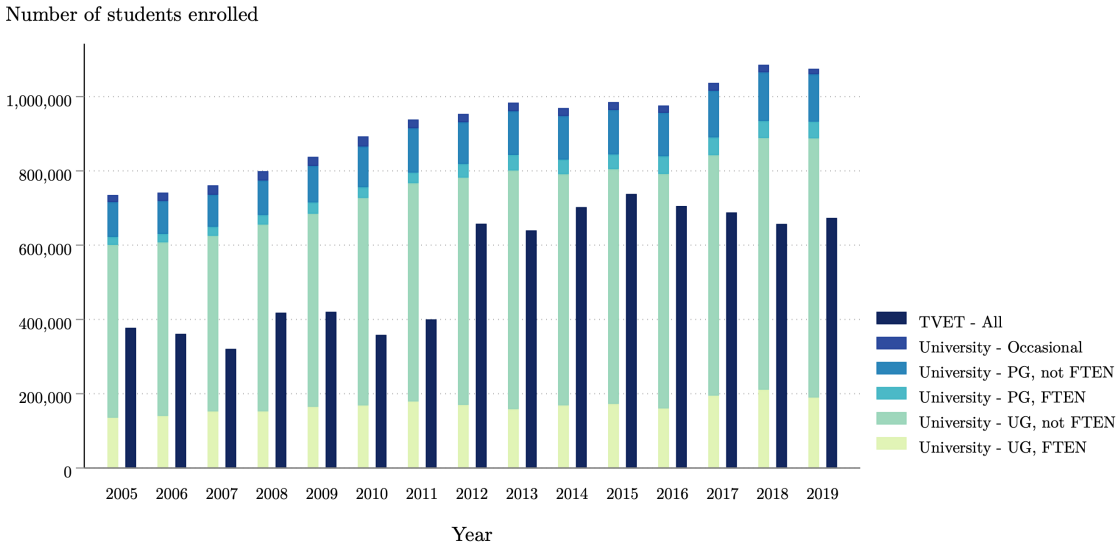
This balance is the inverse of what the White Paper for Post-School Education and Training (2013) initially set out to achieve by 2030. FET/TVET colleges were seen as the cornerstone of skills development in the country and the DHET foresaw building FET/TVET colleges into a post-school institution "of choice" for a large proportion of students leaving the schooling system (Department of Higher Education and Training, 2013).[11] As such, the DHET set an ambitious enrolment target of 2.5 million students in FET/TVET colleges by 2030. FET/TVET colleges did experience growth in the number of enrolled students – almost doubling the number of students enrolled between 2005 and 2015. Much of this rapid growth was due to the expansion of bursary opportunities allocated to FET/TVET colleges, which went from R300 million in 2010 to R1.988 billion in 2013 (Department of Higher Education and Training, 2013). It was expected that this budget would substantially increase over the following two decades to support the ambitious goal of 2.5 million students in FET/TVET colleges by 2030.

However, the number of FET/TVET college enrolment numbers have stalled since 2015, and even declined due to budget constraints and the quality of FET/TVET education being called into question (van der Berg et al., 2020). The DHET, thus, re-evaluated the 2.5 million TVET student enrolment

[11] The DHET explains that students who are not accepted at university (or students who are unlikely to be accepted) tend to enrol at a FET/TVET college. Rather, the DHET wants students to pick FET/TVET colleges as their first institution of choice and not see it as a second best alternative (Department of Education, 2020).

target down to 1.25 million for 2030 (van der Berg et al., 2020). Given that the target enrolment number for university students is 1.6 million, the DHET anticipates that universities will continue to attract more students than FET/TVET colleges, at least until 2030.

Figure 5: Student headcount enrolment at universities and FET/TVET colleges



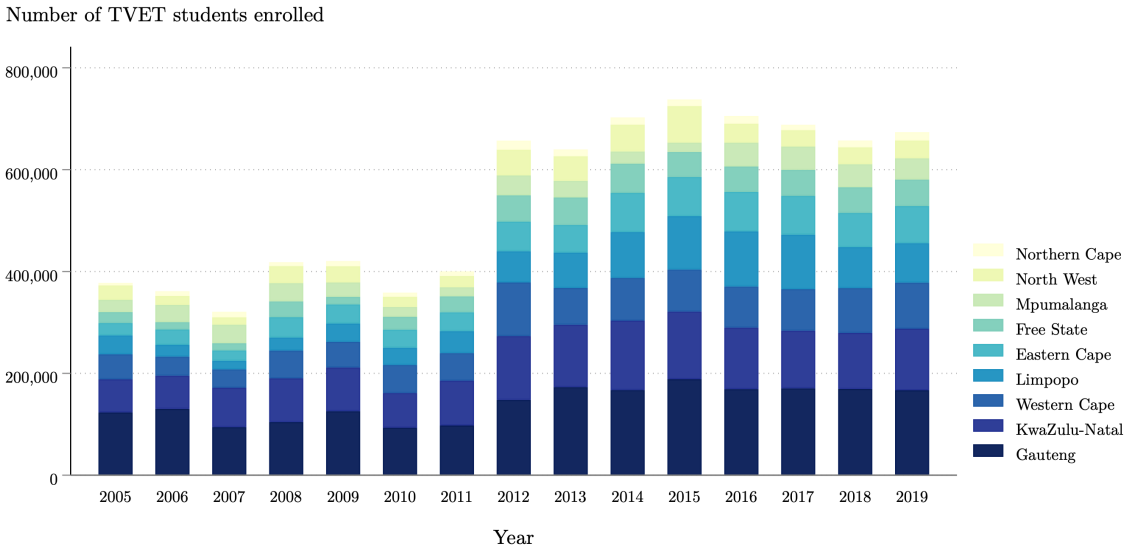
Source: Power Higher Education Data Analyser (PowerHEDA). Department of Education (1999-2009). Department of Higher Education and Training (2010-2019).

Notes: This figure shows the number of students enrolled at universities and FET/TVET colleges by qualification level and entrance category. The qualification levels displayed are occasional, postgraduate (PG), or undergraduate (UG). The entrance categories displayed are first-time entering (FTEN), or not FTEN.

Between 2005 and 2019, students enrolled at FET/TVET colleges in Gauteng, KwaZulu-Natal and the Western Cape accounted for more than 50% of all FET/TVET college enrolments in a given year, as shown in Figure 6. These provinces are also South Africa's most populated and home to three of the country's major metropolitan areas (Johannesburg, Durban and Cape Town). When FET/TVET college enrolments declined from 2015, it was due to enrolments declining in these provinces and the North West. FET/TVET colleges in the North West experienced the largest year-on-year decline (by 48%) from 2015 to 2016 whereas FET/TVET colleges in Gauteng, KwaZulu-Natal and the Western Cape experienced smaller declines of 10%, 9% and 2%, respectively. FET/TVET colleges in the remaining provinces all experienced a growth in student enrolments from 2015 to 2016 – especially FET/TVET colleges in Mpumalanga, which increased student enrolments by 157%.

Growth in the number of students enrolled at universities (as shown in Figure 5) was particularly driven by the increase in the number of university students enrolled at Unisa. Between 2005 and 2019, the number of university students enrolled at Unisa rose from 207,931 students to 342,797 students (65% increase) – representing the largest university type contribution to the growth in the number of total university students. Traditional universities, universities of technology and comprehensive universities experienced smaller growths in the size of their student bodies. University enrolments rose from:

Figure 6: Student headcount enrolment at FET/TVET colleges by province



Source: Department of Education (1999-2009). Department of Higher Education and Training (2010-2019).
Notes: This figure shows students enrolled at FET/TVET colleges in Gauteng, KwaZulu-Natal and the Western Cape accounted for more than 50% of all FET/TVET college enrolments. The number of FET/TVET colleges are as follows: 8 in Gauteng, 9 in KwaZulu-Natal, 6 in the Western Cape, 7 in Limpopo, 8 in the Eastern Cape, 4 in the Free State, 3 in Mpumalanga, 3 in North West, and 2 in the Northern Cape.

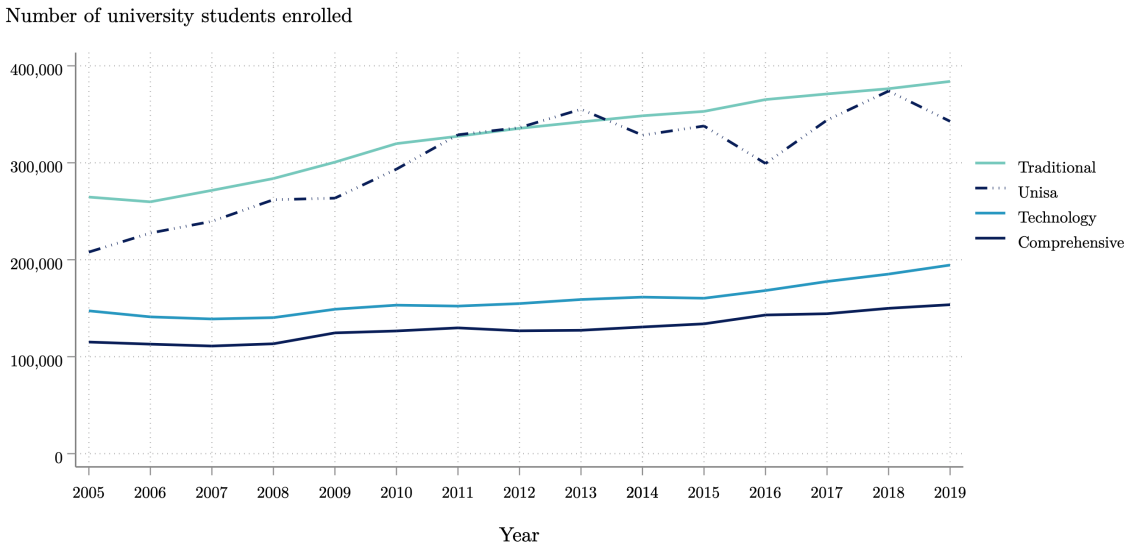
264,673 to 383,941 students (45% increase) at traditional universities; 147,340 to 194,536 students (32% increase) at universities of technology; and 115,129 to 153,638 students (33% increase) at comprehensive universities. These data, showing the number of students enrolled from 2005 to 2019 at each of the four university types, is presented in Figure 7.[12]

Historically, Unisa provided higher education for older and working individuals given the flexibility of distance learning. Unisa was, therefore, not designed to cater for students leaving the schooling system (Essop, 2020). However, Essop (2020) argues as academic pressure for a spot at predominantly contact learning universities increase, first-time entering students are considering distance learning as a viable option, thus driving up the number of enrolments at Unisa. Distance learning also tends to come at a lower cost relative to contact learning, which makes Unisa an attractive choice of institution (Essop, 2020). Unisa's growth in the number of enrolled students meant that, by 2019, student enrolments at Unisa accounted for 32% of all students enrolled at universities, up from 28% in 2005. This rise came at the expense of reducing the share of university enrolments for traditional universities (36 to 35 percent of enrolments), universities of technology (20 to 17 percent of enrolments), and comprehensive universities (16 to 14 percent of enrolments).

Figure 8 focuses on the number of undergraduate, first-time entering students by university

[12] It is noted that growth in enrolments at the institutional types has more to do with the constraints of DHET's enrolment planning processes than the choice of students and the number of institutions in each category.

Figure 7: Student headcount enrolment by institutional type



Source: Power Higher Education Data Analyser (PowerHEDA).
Notes: This figure shows Unisa experienced the highest growth rate (65%) in the number of students enrolled between 2005 and 2019, followed by a growth rate of 45% at traditional universities, 33% at comprehensive universities, and 32% at universities of technology.

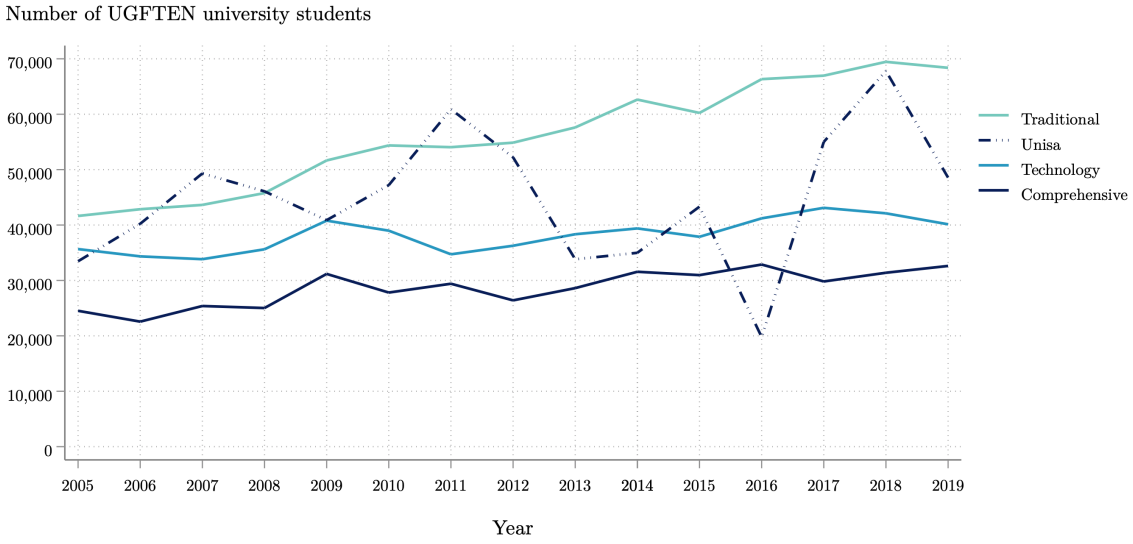
institutional type.[13] Each year represents the number of students enrolling at a type of university for the first time. In other words, each year is a brand new cohort of students. This restriction is useful because trends in the number of new students enrolled are not conflated by the number of existing students in the system – some of whom may be spending more years in the system due to longer qualification programmes (e.g. a degree versus a diploma), extending the duration of their programme (e.g. completing their degree over 5 years rather than 3 years), or enrolling for an additional year to repeat a course. I am interested in which type of university is attracting the highest number of new undergraduate enrolments; and which university types are growing at a faster rate than others.

Traditional universities attracted the highest number of undergraduate, first-time entering students between 2005 and 2019; except for 2007 and 2011. In these years, Unisa enrolled the highest number of undergraduate, first-time entering students. Over the course of this period, traditional universities increased their number of undergraduate, first-time entering students by 26,741 students (65%). This was the greatest growth rate of all the types of universities. The number of undergraduate, first-time entering students at universities of technology grew by 4,473 students (13%, the lowest growth rate of the university types) and by 8,101 students (33%) at comprehensive universities. In addition, the gap between the number of undergraduate, first-time entering students at traditional universities and universities of technology has widened. In 2005, this gap was a difference of 5,980 students and by 2019 this gap reached a difference of 28,248 students.

It is apparent in Figure 8 that the number of undergraduate, first-time entering students for Unisa

[13] Data on first-time entering FET/TVET students is not available for this analysis.

Figure 8: UGFTEN student headcount enrolment by institutional type



Source: Power Higher Education Data Analyser (PowerHEDA).
Notes: This figure shows traditional universities experienced the greatest growth rate (65%) in the number of undergraduate, first-time entering (UGFTEN) students enrolled between 2005 and 2019. Over this period, there was a 33% growth rate at comprehensive universities and 13% growth rate at universities of technology. Unisa showed the most volatile fluctuation in its number of UGFTEN students but grew by 45% between 2005 to 2019.

is more volatile than the other three types of universities. Unisa saw the number of undergraduate, first-time entering students sharply rise in 2007, 2011, 2015, and 2018; and fall in 2009, 2013, 2016 and 2019. The fluctuation in the number of undergraduate, first-time entering students at Unisa is large enough to cause noticeable shifts in the same direction for these years in Unisa's aggregate population trend, displayed in Figure 7. Despite these fluctuations, Unisa still managed to increase its number of undergraduate, first-time entering students by 15,102 students (45%) between 2005 and 2019.

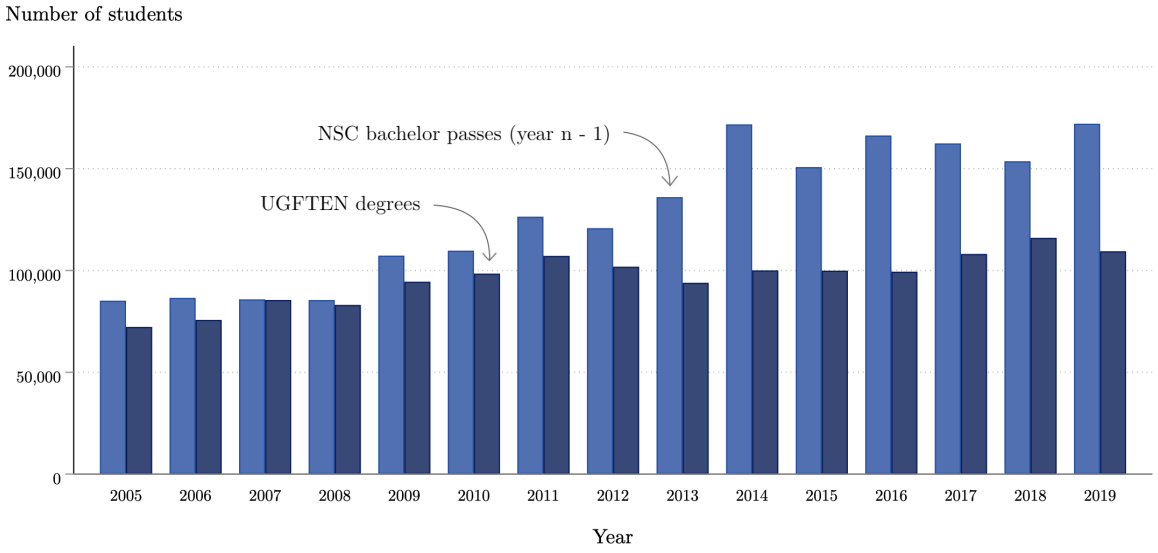
Essop (2020) explains that the growth in university enrolments at Unisa (and its volatility) is largely driven by Unisa enrolling students leaving the schooling system who meet the minimum academic entry requirements for university but whose academic marks are not high enough to secure a spot in a contact learning programme at one of the other universities.[14] A National Senior Certificate (NSC) bachelors pass, previously referred to as a matriculation exemption, is the minimum academic eligibility requirement for students to apply for an undergraduate degree programme at any of the university types.

Figure 9 shows the number of undergraduate, first-time entering degree enrolees per year alongside the

[14] Van Broekhuizen et al. (2016) find substantial differences across universities in the average grade 12 performance of their students. The University of Cape Town, Stellenbosch University, and University of Pretoria enrolled students with an average grade 12 mark above 70%, while the majority of South Africa's universities accepted students with an average grade 12 mark between 50% to 60%. It is possible that part of the variation in average grade 12 marks across universities is due to the qualification type (degree, diploma or certificate) for which students are enrolling. For example, more than 99% of all undergraduate enrolments are for degree programmes at 11 out of the 26 universities; whereas degree programmes account for less than 50% of undergraduate enrolments at 7 out of the 26 public universities. Unsurprisingly, the 11 universities which have predominantly enrolees for degree programmes are a mixture of traditional and comprehensive universities and the 7 universities with a lower proportion of degree enrolees are universities of technology. However, even when Van Broekhuizen et al. (2016) restrict the analysis to degree programmes, there is still sizeable variation in students' average grade 12 performance across universities.

number of NSC bachelor passes for year n-1. For example, in 2011 there were 107,138 undergraduate, first-time entering degree enrolees. Displayed next to this bar is the number of students who wrote the NSC in 2010 and achieved a bachelors pass (126,371 students). Hence, despite this growth in the number of NSC bachelor passes, the number of undergraduate, first-time entering students enrolling for an undergraduate degree programme have not kept pace.[15]

Figure 9: Number of students with an NSC bachelors pass (year n-1) and number of UGFTEN students enrolled for a degree programme



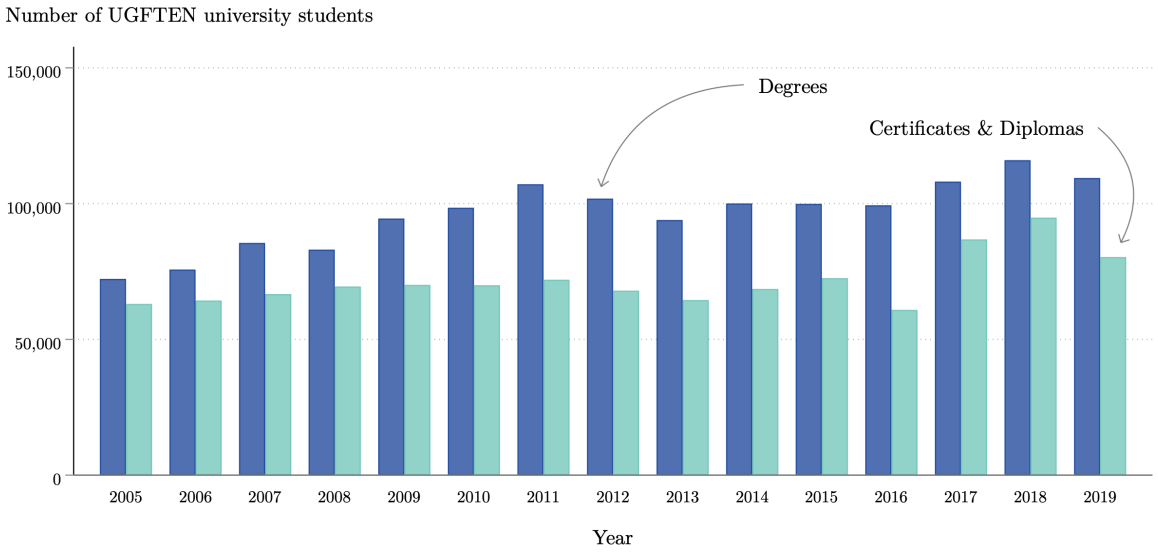
Source: Power Higher Education Data Analyser (PowerHEDA). Department of Education (1999-2009). Department of Basic Education (2009-2019).
Notes: This figure shows the number of undergraduate, first-time entering (UGFTEN) degree enrolments have not kept pace with the rise in the number of National Senior Certificate (NSC) bachelor passes.

Undergraduate, first-time entering university students can also enrol for undergraduate certificates or diplomas; not only undergraduate degrees.[16] Figure 10 shows more undergraduate, first-time entering students enrol for degree qualifications than certificate and diploma qualifications between 2005 and 2019. In 2005, 72,261 (53%) undergraduate, first-time entering students enrolled for an undergraduate degree and the remaining 63,032 (47%) students enrolled for an undergraduate certificate or diploma. Over time, a larger share of undergraduate, first-time entering students enrolled in degree programmes. The biggest difference was seen in 2016 when degree programmes accounted for 62% – and certificate

[15] Branson and Kahn (2018) find, using the first four wave of the National Income Dynamic Study (NIDS), that only 32% of students who successfully completed grade 12 enrolled in a post-school institution within two years of their grade 12 NSC examinations. This group is made up of 13% attending a university, 8% attending a FET/TVET college, and 11% attending a variety of post-school institutions. The remaining 68% of students who complete grade 12 opt not to enrol in a post-school institution.

[16] Students who achieve a NSC bachelor pass are not restricted to applying to undergraduate degree programmes only – they may also choose to apply for a certificate or diploma programme at any university or FET/TVET college. Students who achieve a NSC diploma pass are not eligible to apply for a degree programme but they may apply for an undergraduate diploma or certificate programme at a university or FET/TVET college. Lastly, students who achieve a NSC higher certificate pass are only eligible to apply for a higher certificate programme at a university or FET/TVET college. Hence, the type of programme and the type of university undergraduate, first-time entering students' choose to enrol in is constrained by their basic education schooling performance.

Figure 10: Number of UGFTEN students enrolled for a degree programme and number of UGFTEN students enrolled for a certificate or diploma programme



Source: Power Higher Education Data Analyser (PowerHEDA).
Notes: This figure shows more undergraduate, first-time entering (UGFTEN) students enrol for a degree programme relative to a certificate or diploma programme at university.

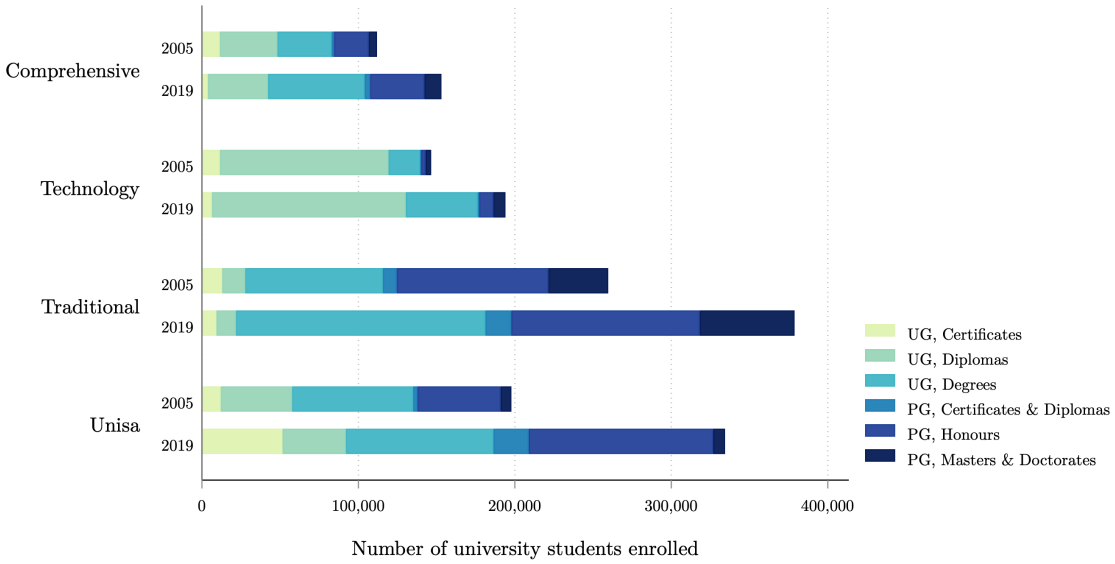
and diploma programmes accounted for 38% – of undergraduate, first-time entering students. This was the same year where Unisa experienced its lowest number of undergraduate, first-time entering students (19,785 students) for the period. The significance of this is that Unisa is one of the institutional types with a larger offering of undergraduate certificate and diploma programmes.[17] In 2017 and 2018, undergraduate, first-time enrolments in certificate and diploma programmes started to pick up before dropping again in 2019 when Unisa experienced another decline in its number of enrolled undergraduate, first-time entering students. In 2019, 109,412 (58%) undergraduate, first-time entering students enrolled for an undergraduate degree and the remaining 80,298 (42%) students enrolled for an undergraduate certificate or diploma.

The qualification programmes offered by universities and FET/TVET colleges differ. Universities offer undergraduate certificates, undergraduate diplomas, undergraduate degrees, postgraduate certificates and diplomas, postgraduate honours, and postgraduate masters and doctorates; whereas FET/TVET colleges offer undergraduate certificates and diplomas. FET/TVET colleges do not offer undergraduate degrees or postgraduate qualifications. Degree programmes are typically longer in minimum duration (approximately 3 years) compared to certificate and diploma programmes (approximately 1 to 3 years).

Furthermore, the learning orientation of the qualification programmes differ by institutional type. Universities of technology tend to offer more teaching-led programmes (undergraduate certificates and

[17] Figure 11 shows Unisa has a larger number of students enrolled for undergraduate certificates and diplomas relative to comprehensive and traditional universities.

Figure 11: Student headcount enrolments in qualification programmes by institutional type



Source: Power Higher Education Data Analyser (PowerHEDA).
Notes: This figure shows the volume of students enrolled in qualification programmes differ by institutional type. For example, traditional universities have the greatest number of students enrolled in postgraduate, masters and doctorate qualifications; and universities of technology have the greatest number of students enrolled in undergraduate, diploma qualifications.

diplomas), traditional universities tend to offer more research-led programmes (undergraduate degrees; and postgraduate honours, masters and doctorates), and comprehensive universities tend to offer a combination of both teaching-led and research-led programmes.[18] FET/TVET colleges offer either technical, vocational, or occupational orientated qualification programmes. These include: the National Certificate (Vocational) (NC(V)), Report 191/NATED (N1 to N6), occupational qualifications, and other skills and short courses.[19] These data showing the number of students enrolled by qualification programme are presented in Figure 11 (for universities) and Figure 12 (for FET/TVET colleges).

From 2005 to 2019, the number of enrolments for undergraduate certificates at comprehensive universities, universities of technology, and traditional universities declined. However, the number of enrolments for undergraduate certificates at Unisa rose from 12,024 to 51,500 – an increase of 328%. Unisa also experienced substantial growth in postgraduate certificates and diplomas (2,729 to 22,548), and postgraduate honours qualifications (53,178 to 117,922), which increased by 726% and 122%, respectively. Growth in enrolments for undergraduate degrees from 2005 to 2019 was high for comprehensive universities (78%), universities of technology (124%) and traditional universities (81%). Unisa only grew undergraduate degree enrolments by 22% from 2005 to 2019 – much lower than the other university types.

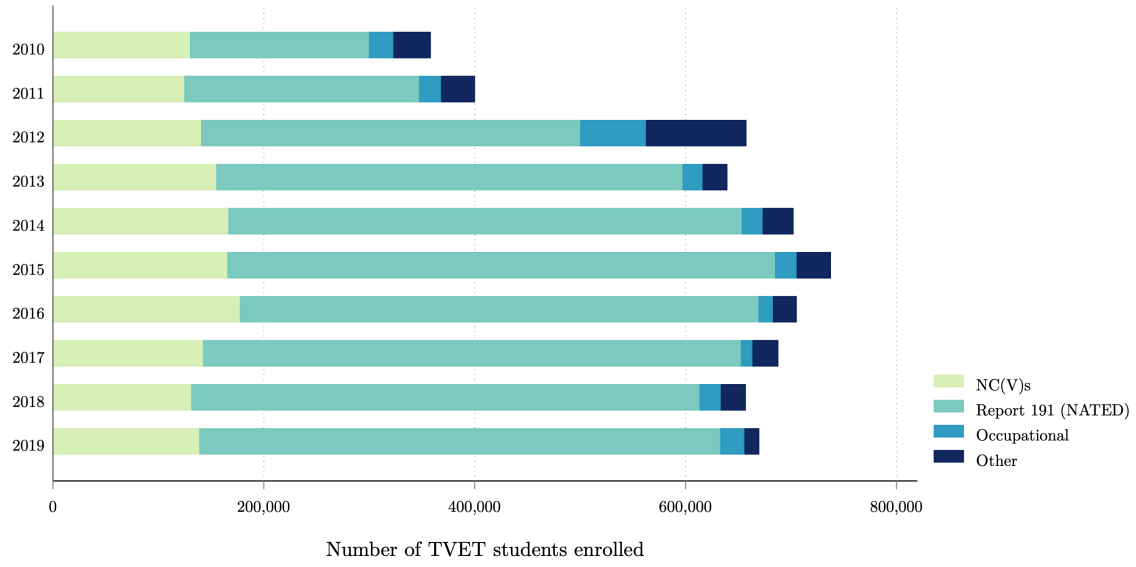
[18] Unisa is classified as a comprehensive university.

[19] The Report 191/NATED (N1 to N6) programmes consist of theory learnt at colleges and relevant practical application in the workplace. The NC(V) programmes combine both theory and practical skills within specific industry fields. The occupational programmes are associated with a trade, occupation or profession and consists of knowledge and theory, practical skills and workplace experience (Department of Higher Education and Training, 2021).

In 2019, 73% of undergraduate certificate enrolments were at Unisa. Universities of technology accounted for the majority (58%) of undergraduate diploma enrolments in 2019. This was followed by Unisa making up 19%, comprehensive universities making up 18%, and traditional universities making up 6% of undergraduate diploma enrolments. Although traditional universities account for the smallest share of undergraduate diploma enrolments, these universities account for the largest share of undergraduate degree enrolments (43%) and postgraduate masters and doctorates enrolments (70%) in 2019. These enrolment shares by qualification type are inline with expectations based on what is known about the learning orientation of the programmes on offer at these institutional types.

As shown in Figure 12, Report 191/NATED qualification programmes were the most popular, followed by the NC(V) qualification programmes. In 2010, 47% of FET/TVET college enrolments were for Report 191/NATED qualification programmes. This jumped to 73% by 2019. NC(V) qualification programmes have slowly been accounting for a smaller and smaller share of FET/TVET college enrolments year after year, despite only being introduced in 2007. In 2010, NC(V) programmes accounted for 36% of student enrolments. By 2019, NC(V) programmes only accounted for 21% of student enrolments. From 2013 to 2019, occupational qualifications accounted for, on average, 3% of enrolments and the Other programmes accounted for, on average, 4% of enrolments.

Figure 12: Students headcount enrolments at FET/TVET college by qualification programme



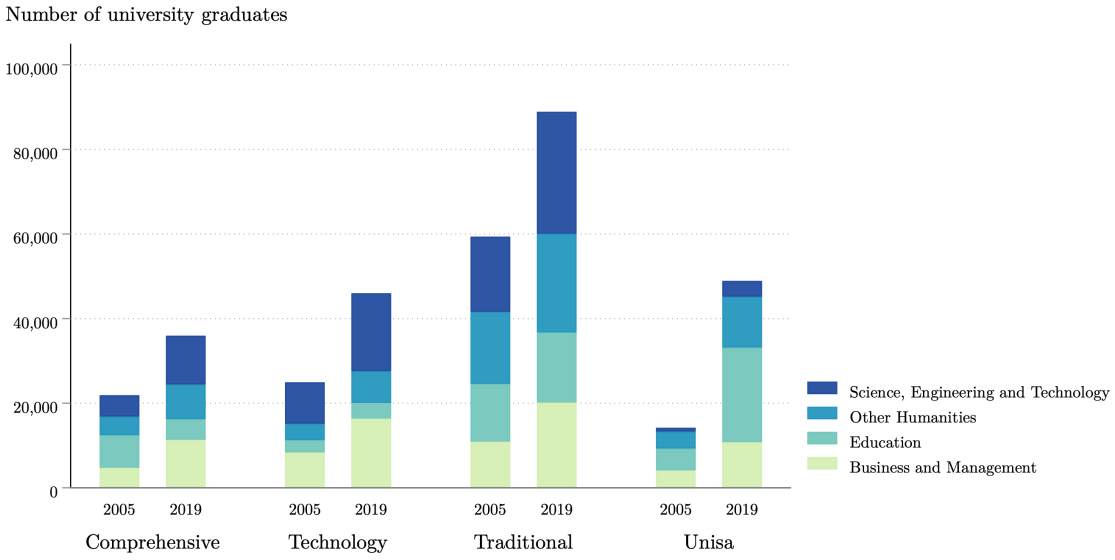
Source: Department of Higher Education and Training (2010-2019).
Notes: This figure shows Report 191 (NATED) qualification programmes are the most popular at TVET colleges.

Universities and FET/TVET colleges also offer different fields of specialisation. To assess these differences, I compare the number of graduates for each major field of study. Universities aggregate fields of study based on the Classification of Educational Subject Matter (CESM). These major fields are business and management; education; humanities; and science, engineering, and technology. The fields of specialisation at FET/TVET colleges are not based on the CESM. Rather, fields of study are

discussed in terms of the NC(V)4 and N6 programme examinations. Programmes with the highest number of students completing provide a sense of the fields of specialisation at FET/TVET colleges.

Figure 13 shows Unisa produced 3.4 times more graduates in 2019 relative to 2005. All CESM fields for Unisa graduates increased by more than 2-fold: the number of education graduates increased from 5,148 to 22,363 graduates (334% increase), the number of science, engineering and technology graduates increased from 894 to 3,718 graduates (316%), the number of other humanities graduates increased from 4,016 to 12,037 graduates (200%), and the number of business and management graduates increased from 4,128 to 10,789 graduates (161% increase). One possible explanation for Unisa's substantial growth in graduates is due to the rise in its enrolments for undergraduate certificate and postgraduate honours qualification programmes (see Figure 11). These programmes typically take only one year to complete. Universities of technology, comprehensive universities and traditional universities produced 84%, 64% and 50% more graduates in 2019 compared to the number of graduates they produced in 2005.

Figure 13: Student headcount graduates in CESM fields by institutional type



Source: Power Higher Education Data Analyser (PowerHEDA).
Notes: This figure shows Unisa experienced a greater than 3-fold increase in graduates from 2005 to 2019. In 2019, Unisa was the largest producer of education graduates, and traditional universities produced the largest number of graduates in the remaining three Classification of Educational Subject Matter (CESM) fields – namely science, engineering, and technology; other humanities; and business and management.

In 2005, traditional universities produced the most education graduates (13,639 graduates). By 2019, Unisa took the lead in producing the most education graduates (22,363 graduates) – thus, producing 47% of all education graduates. The second largest producer of education graduates in 2019 was traditional universities (35%). In addition, traditional universities produced the largest share of graduates in the remaining three fields – business and management (34%), other humanities (46%), and science, engineering and technology (46%). Universities of technology accounted for the smallest share of education graduates (8%); and 29% of science, engineering and technology graduates; 28% of

business and management graduates, and 15% of other humanities graduates. Comprehensive universities produced 19% of business and management graduates, 18% of science, engineering and technology graduates, 16% of other humanities graduates, and 10% of education graduates.

As detailed in the *Statistics for Post-School Education and Training 2019* report (Department of Higher Education and Training, 2021), the NC(V)4 qualification examinations had 10,920 FET/TVET college students complete the examinations out of a total 22,126 students who wrote the examinations (49% completion rate). These students were enrolled across 18 different programmes, ranging from engineering to agriculture.[20] Office Administration had the highest number of students (3,170) who completed the NC(V)4 qualification examinations, making up 29% of those who completed. This was followed by Tourism with 909 students who completed (8% of total competed), and Engineering and Related Design with 826 students who completed (8% of total competed). Drawing Office Practice (6 students completed), Process Plant Operations (25 students completed), and Mechatronics (61 students completed) had the lowest highest number of students who completed – each programme making up less than 1% of the students who completed.

The N6 qualification examinations (part of the Report 191 (NATED) programmes), in 2019, had 113,393 FET/TVET college students who completed out of a total 117,641 students who wrote the examinations (96% completion rate). The N6 programmes, like the NC(V) programmes, offer a broad range of fields across 20 programmes – including public relations, interior decorating, and farming management to name a few.[21] The N6 programmes with the largest number of completed students are Engineering Studies with 27,343 students (24%), Management Assistant with 22,368 students (20%), Human Resources Management with 12,777 students (11%), Business Management with 11,535 students (10%), and Public Management with 10,745 students (9%). Hence, engineering and management fields of specialisation are the most popular N6 programmes. Interior Decorating and Popular Music: Composition were the programmes with the lowest number of completed students – each with 3 students.

[20] The 18 NC(V) programmes in 2019 are: Civil Engineering and Building Construction; Drawing Office Practice; Education and Development; Electrical Infrastructure Construction; Engineering and Related Design; Finance Economics and Accounting; Hospitality; Information Technology and Computer Science; Management; Marketing; Mechatronics; Office Administration; Primary Agriculture; Primary Health; Process Plant Operations; Safety in Society; Tourism; and Transport and Logistics (Department of Higher Education and Training, 2021).

[21] The 20 N6 programmes in 2019 are: Art and Design; Business Management; Clothing Production; Educare; Engineering Studies; Farming Management; Financial Management; Hospitality and Catering Services; Human Resources Management; Interior Decorating; Legal Secretary; Management Assistant; Marketing Management; Medical Secretary; Popular Music: Composition; Popular Music: Performance; Popular Music: Studio Work; Public Management; Public Relations; and Tourism (Department of Higher Education and Training, 2021).

4.2 Student characteristics, from 2005 to 2019

A higher number of females, relative to males, are enrolled in universities and TVET colleges.[22] This is plotted in Figure 14. In 2005, there were 401,042 female university students and 334,030 male university students.[23] The gap between the two gender groups for university students has widened overtime as the enrolment of female students increased at a higher rate relative to male students. From 2005 to 2019, the enrolment of female students at universities grew by 60%, whereas the enrolment of male students grew by 30%. By 2019, there were 640,333 female students and 434,514 male students enrolled at universities.

Enrolment numbers by gender for TVET colleges is only available from 2013 but the data tells a similar story to that of the gender enrolments at universities. There is a larger number of female students enrolled at TVET colleges than male students, and growth in the number of female student enrolments outpaced the growth in the number of male student enrolments between 2013 and 2019. In 2013, the difference in enrolment numbers between female (327,091 students) and male (312,527 students) is relatively close but gradually expands overtime. Noticeably, growth in the overall number of TVET enrolments stems from growth in female enrolments, which grew by 21%. Male enrolments declined by 11% from 2013 to 2019. Male enrolments also took a larger hit in 2015 when overall TVET college enrolments started to decline. From 2015 to 2019, male enrolments at TVET colleges declined by 17%, whereas female enrolments only declined by 2%. Therefore, there were 396,559 female students and 276,931 male students enrolled in TVET colleges in 2019.

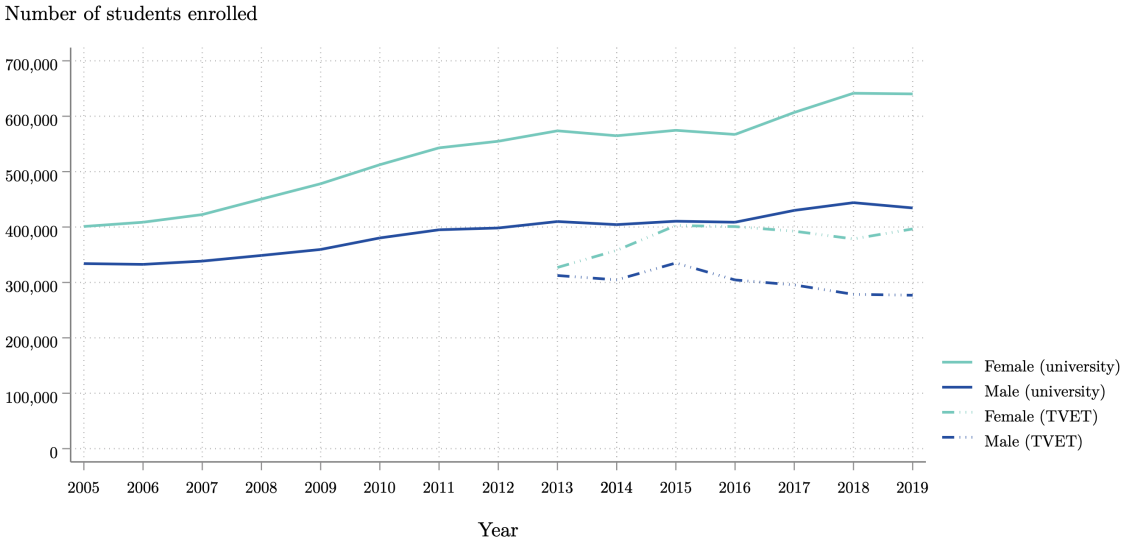
University enrolment patterns have not always been female-leaning. In 1993, 43% of university students were female (Bunting, 2006a). Since then, the higher education sector (universities and technikons) has made strides in female representation. By 1997, 50% of university students were female. Historically Black universities and historically White universities (Afrikaans speaking) had the highest share of female students – each with 55% of their student enrolments identifying as female. Historically White technikons and historically White universities (English speaking) had the lowest share of females students, at 43% and 47% respectively. Historically Black technikons equalled the overall average of the higher education sector at 50%; while less than half (48%) of student enrolments at the distance university (Unisa) and the distance technikon (Technikon South Africa) were female. By 2000, the majority (53%) of university enrolments were female (Bunting, 2006a). All higher education institutions, by historical classification, had 50% or more of their enrolments identifying as female in 2000, except for the historically White technikons at 46%. However, this was still an improvement of 3 percentage points since 1997.

Post the merging of institutions, universities continued to increase their share of female representation, as was depicted in Figure 14. In 2005, the share of female enrolments at universities was 54%; and by 2019 the share of female enrolments reached a high of 60%. However, the extent of female representation differs across the university types. These data are plotted in Figure 15, along with the available TVET data from 2013 to 2019. Literature by Van Broekhuizen and Spaull (2017) show females are more likely

[22] Data on student characteristics enrolled at TVET colleges is only available from 2013 – the year FET colleges became known as TVET colleges. Hence, in this section I use the term TVET colleges rather than FET/TVET colleges.

[23] There was 1 student of unknown gender at universities in 2005.

Figure 14: Student headcount enrolments by gender



Source: Power Higher Education Data Analyser (PowerHEDA).
Department of Higher Education and Training (2013-2019).
Notes: This figure shows both universities and TVET colleges have a higher number of female students enrolled than male students.

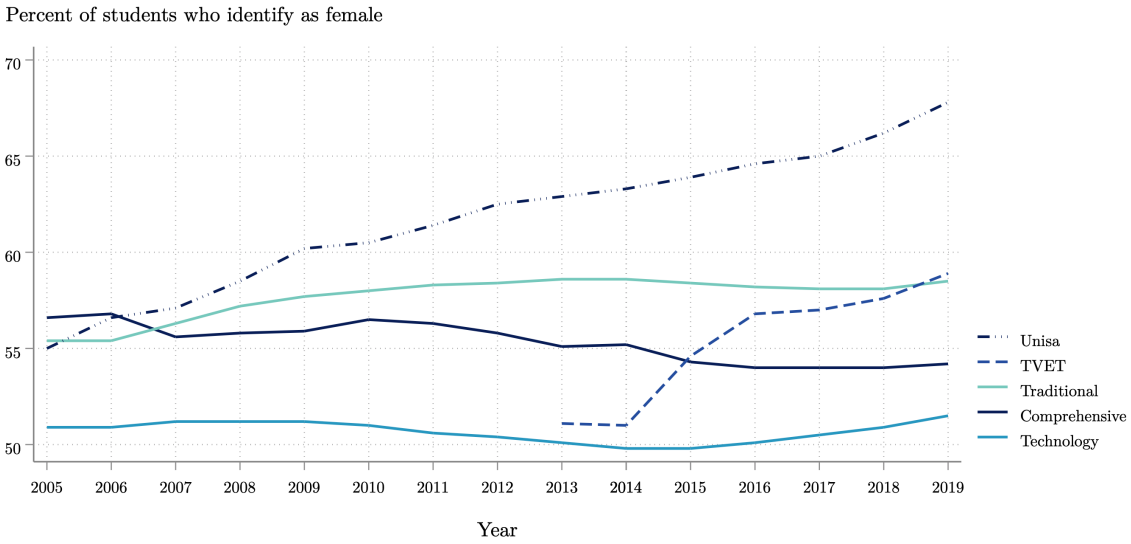
to enrol in, and graduate from, post-school institutions.

All institutional types have upwards of 50% of enrolments identifying as female. Universities of technology have the lowest representation of females, ranging between 50% and 51% from 2005 and 2019.[24] Traditional universities and Unisa both had females accounting for 55% of enrolments in 2005. While the traditional universities managed to increase their share of female enrolments by approximately 3 percentage points in 2019, Unisa increased its share of female enrolments by approximately 13 percentage points to reach a share of 68% – the highest female share of all the institutional types. TVET colleges also experienced expansion in their share of female enrolments, which grew from 51% in 2013 to 59% in 2019. Comprehensive universities, on the other hand, are the only institutional type to see their share of female enrolments decline post 2005. The share of female enrolments for comprehensive universities fell from 57% in 2005 to 54% in 2019. Although these overall improvements in female representation are encouraging, they mask underlying inequalities, which may still persist in qualification types and fields of study.

Figure 16 shows the number of enrolments for universities and FET/TVET colleges by population group. The number of Black students enrolled at universities has grown from 446,946 students in 2005 to 830,797 students in 2019 – a growth rate of 86%. Despite this growth, the share of Black students at universities (77% in 2019) is still less than the share of Black people in the population (81%) (Statistics South Africa, 2020). The second highest number of students are from the White population group. In 2005, there were 185,847 White students at university. This number continuously declined throughout

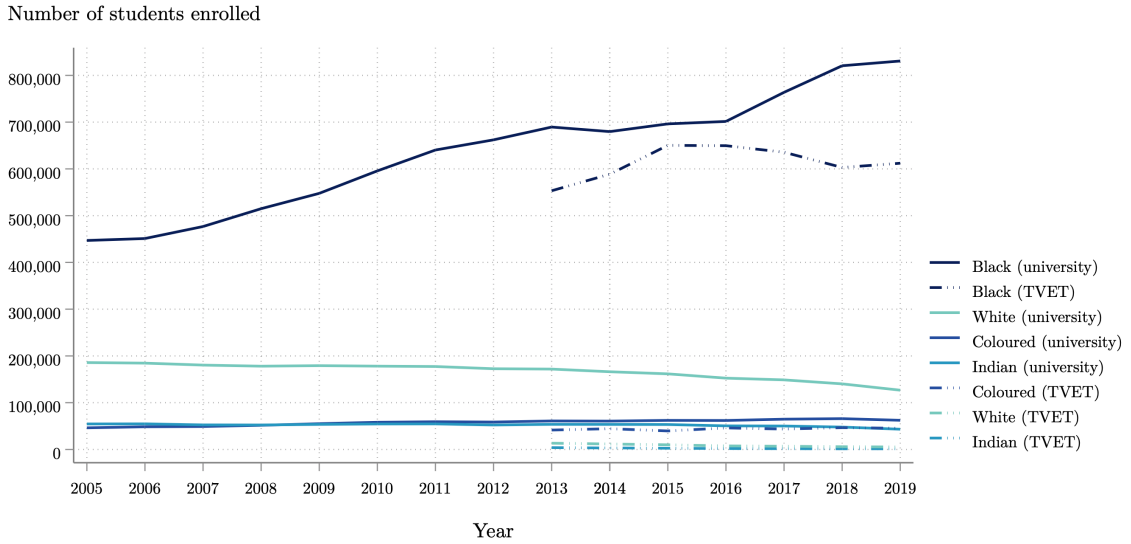
[24] The percent of students identifying as female at universities of technology dipped below 50% to 49.8% in 2014 and 2015.

Figure 15: Share of enrolled students who identify as female by institutional type



Source: Power Higher Education Data Analyser (PowerHEDA).
Department of Higher Education and Training (2013-2019).
Notes: This figure shows all institutional types increased their share of enrolled students who identify as female from 2005 to 2019, except for comprehensive universities.

Figure 16: Student headcount enrolments by population group



Source: Power Higher Education Data Analyser (PowerHEDA).
Department of Higher Education and Training (2013-2019).
Notes: This figure shows both universities and TVET colleges have a higher number of Black students relative to the other population groups.

the post-merge period, reaching 126,755 students in 2019. The number of Coloured students surpassed the number of Indian students at university in 2009. Between 2005 and 2019, the number of Coloured students grew from 46,302 to 62,362 students (35% increase); and the number of Indian students declined from 54,611 to 43,330 students (26% decrease).

The vast majority (approximately 90%) of students enrolled at TVET colleges are Black students. Between 2013 and 2019, the number of Black students enrolled at TVET colleges rose from 553,248 to 612,187 students. The number of Coloured, White and Indian students at FET/TVET colleges is substantially smaller. The number of Coloured students increased from 41,557 to 45,374 students; whereas the number of White and Indian students shrunk. Between 2013 and 2019, the number of White students declined by 61% (from 13,370 to 5,243 students) and the number of Indian students also declined by 61% (from 3,985 to 1,535 students). This reduction in the number of White student enrolments caused the share of White students enrolled in TVET colleges to drop from 2.2% in 2013 to 0.8% in 2019. Indian students consistently made up less than 1% of enrolments and Coloured students accounted for between 5% and 7% of enrolments over the period.[25]

The share of students enrolled by population group at universities has also shifted over time. In 2005, 61% of students enrolled at universities were Black students and, by 2019, the share of Black students increased to 77%. Hence, the share of Black students enrolled at TVET colleges (91% in 2019) is greater than at universities (77% in 2019). However, there is variation in the share of students enrolled by population group for each type of university.

Recall 6 out of the 8 historically White institutions are now categorised as traditional universities since the merging of institutions.[26] Figure 17 shows, of all the university types, traditional universities have the lowest share of Black students (49% in 2005 to 64% in 2019) and the highest share of Coloured, Indian and White students. On the other hand, in 2019, 90% of students enrolled at comprehensive universities and at universities of technology were Black students – a similar composition to that at TVET colleges. No comprehensive universities were historically for White students and only two universities of technology were historically White. Noting this historical reference, it seems that traditional universities are making progress in becoming more representative of the the South African population rather than being skewed towards historically White students. Unisa made the most gains in broadening its share of Black students by 24 percentage points – from 56% in 2005 to 80% in 2019.[27]

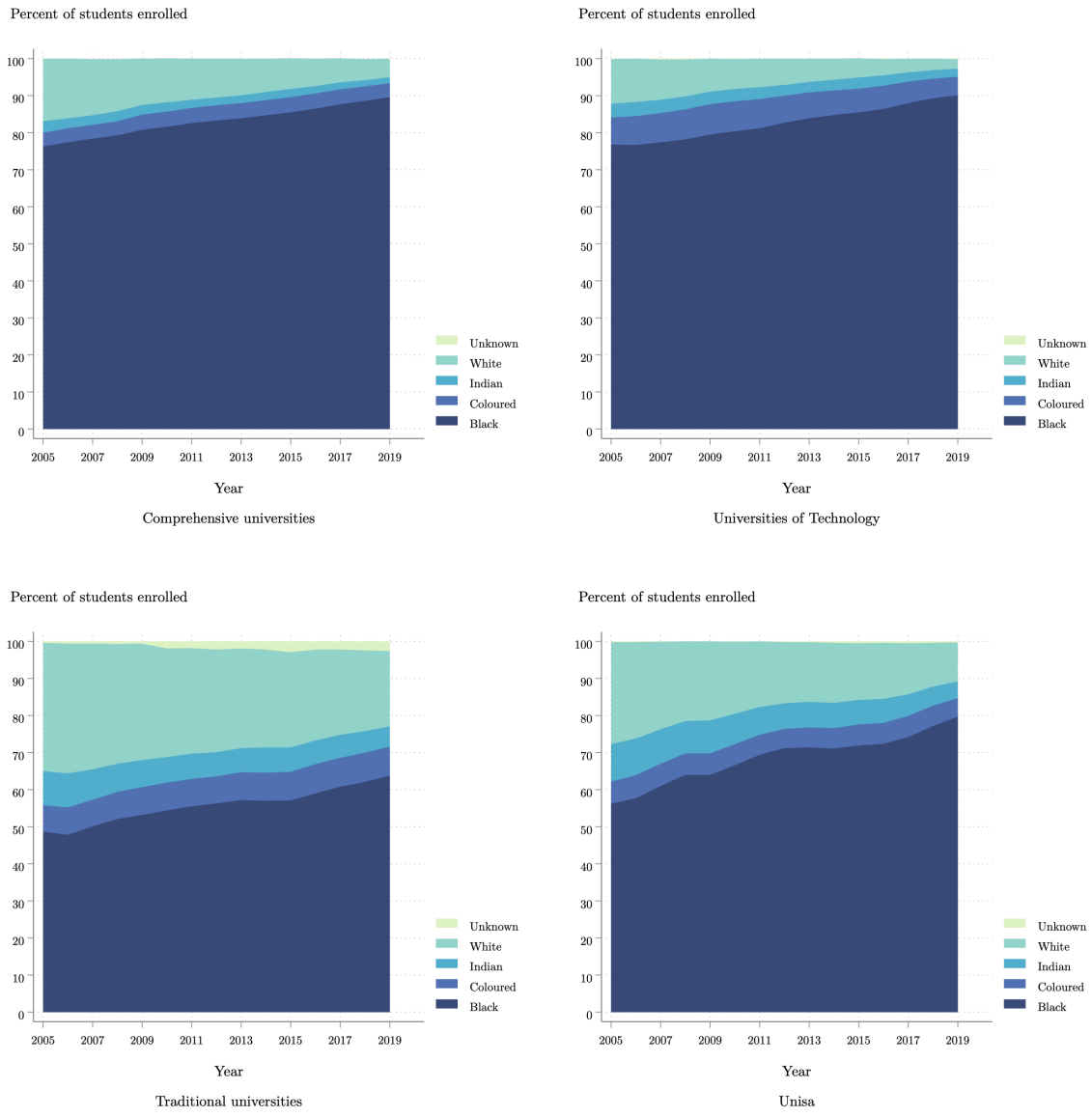
This progress of increasing participation by previously disadvantaged population groups in the reformed post-school institution era is encouraging as this was not the case in the early 1990s. Despite the higher education sector (universities and technikons) experiencing high growth rates in student

[25] Theses percentages add up to less than 100% due to unknown and reconciliation categories.

[26] See Table 2.

[27] How do these enrolment population group trends relate to NSC outcomes? Van Broekhuizen et al. (2016) find overall enrolment at university, based on NSC results, is not biased based on population group. The proportion of White students with a bachelors NSC pass from the 2008 cohort who went on to study at a university within the next six years was 63.4%. By comparison, 71.2% of Black students with a bachelors NSC pass studied at a university. Yet, only 75% of Black students in this group enrolled for a degree programme; whereas 93% of the White student group chose to pursue a degree programme. Black students with a bachelors NSC pass are, therefore, more likely to enrol for certificate or diploma programmes at a university relative to White students with a bachelors NSC pass. These findings show there are differences by population group when comparing enrolments in undergraduate degree programs at universities.

Figure 17: Population group composition of university types



Source: Power Higher Education Data Analyser (PowerHEDA).
Notes: This figure shows the share of Black students at universities increased from 2005 to 2019 for all university types. The share of Black students reached as high as 90% of the student body at comprehensive universities and universities of technology.

enrolments between 1990 and 1994, one noticeable challenge was that racial inequalities in the sector were still evident. Bunting (2006a) notes, in 1993, approximately 13% of the South African population were White and 75% were Black. Yet, the participation rate by population group in universities and technikons for individuals age 20 to 24 years was 70%, 40%, 13%, and 9% for White, Indian, Coloured, and Black population groups respectively (Bunting, 2006a).

Only towards the late 1990s did historically White universities and historically White technikons start to noticeably increase their share of Black students (Bunting, 2006a). From 1997 to 2000, historically White universities (Afrikaans-speaking) increased their share of Black students from 49% to 58%; historically White universities (English-speaking) increased their share of Black students from 55% to 62%; and historically White technikons increased their share of Black students from 60% to 77%. Distance institutions (Unisa and Technikon South Africa) had the lowest share of Black enrolments (37%) in 1997 and increased this share to 69% by 2000. Historically Black universities and historically Black technikons retained a 99% share of Black students. Hence, they did not reduce their share of Black students in favour of absorbing a higher share of, for example, White students.

Bunting (2006a) notes that this shift in the distribution of population group enrolments does not stem primarily from a major increase in Black student enrolments. Although the number of Black student enrolments did rise, there was also a sharp fall in the number of White student enrolments in the higher education sector from 215,000 in 1995 to 164,000 in 2000 – making up a decline of 41,000 (19%) White students in the sector. Hence, this had an effect on the distribution of student enrolments by population group.

The final student characteristic I assess from the publicly available institutional data is students' age. These data are a recent addition (since 2018) to the *Statistics on Post-School Education and Training in South Africa* reports for university students. Institutional data on age groups for students enrolled at TVET colleges have been incorporated since 2013. Compulsory schooling in South Africa is until the student reaches grade 9 or turns 15 years old, as per the South African Schools Act (1996). Therefore, the age group categories assessed start from 15 years of age. If students continue through the basic education stream, and incur no grade repetitions, students are expected to complete grade 12 at age 18. Therefore, if the majority of students enrol shortly after completing their basic education, it is anticipated that they will fall within the 19 to 24 age group. This aligns with what is shown in Table 4 for university students, and Table 5 for TVET college students.

According to the institutional data for universities, most enrolled students (approximately 50%) are 19 to 24 years old. Likewise, the majority of enrolled TVET college students are 20 to 24 years. This age group has accounted for a growing majority of TVET students – starting at 51.5% in 2013 and reaching approximately 60% from 2016 to 2019. Table 4 shows a small share of students (9.2% to 9.1%) between the age of 15 to 18 enrol at university. Given that completion of basic education is required to enrol at university, it is rare to see students between 15 to 18 years old enrolled here. On the other hand, students are allowed to enrol at TVET colleges after grade 9 (or before reaching grade 12). Besides the outlier of 24.7% in 2013, between 9.9% and 11.8% of TVET college enrolments are for students from the 15 to 19 year age group. This is a relatively small share of students compared to the 20 to 24, and 25

to 29 age groups. This implies that students who leave the schooling system before completing grade 12 may not be optimising this opportunity at TVET colleges to continue their studies.

Over the course of 2013 to 2019, between 14.5% and 19.6% of TVET college students have come from the 25 to 29 years age group. The 30 to 34, and 35 plus age groups account for the smallest share of students (between 4.2% and 6.4%). TVET colleges do not offer postgraduate qualification options so it is unlikely to see students from older age groups enrolling at TVET colleges. By comparison, universities (where there are postgraduate qualification programmes) see a higher share of enrolled students falling within the 30 to 39 (15.4% to 16.1%), and 40 plus (8.8% to 8.5%) age group.

Table 4: University student enrolment by age group

	15 - 18 years		19 - 24 years		25 - 29 years		30 - 39 yeasrs		40+ years		**Total**
	n	%	n	%	n	%	n	%	n	%	
2018	99,811	9.2	540,983	49.8	175,010	16.1	174,720	16.1	95,044	8.8	1,085,568
2019	97,377	9.1	545,299	50.7	176,082	16.4	165,199	15.4	90,955	8.5	1,074,912

Source: Department of Higher Education and Training (2018-2019).
Notes: This table shows most university students are 19 to 24 years of age.

Table 5: TVET college student enrolment by age group

	15 - 19 years		20 - 24 years		25 - 29 years		30 - 34 years		35+ years		**Total**
	n	%	n	%	n	%	n	%	n	%	
2013	140,590	24.7	293,616	51.5	82,525	14.5	29,226	5.1	23,761	4.2	569,718
2014	83,056	11.8	368,797	52.5	116,679	16.6	43,358	6.2	35,301	5.0	702,383
2015	86,659	11.7	379,625	51.4	141,290	19.1	47,202	6.4	34,949	4.7	737,880
2016	70,139	9.9	423,451	60.0	138,203	19.6	42,075	6.0	31,518	4.5	705,397
2017	72,059	10.5	413,489	60.1	131,025	19.0	39,900	5.8	31,545	4.6	688,028
2018	71,882	10.9	387,254	58.9	125,461	19.1	40,558	6.2	31,945	4.9	657,133
2019	77,867	11.6	395,064	58.7	127,558	18.9	41,693	6.2	31,304	4.6	673,490

Source: Department of Higher Education and Training. (2013-2019).
Notes: This table shows the majority of TVET college students are 20 to 24 years of age. The Department of Higher Education and Training statistical reports for 2013 and 2014 provide a reconciliation count of students of 55,192 for 2013 and 48,155 for 2014 to reach the total count of students for the respective years.

4.3 Discussion of profiling observations

This section has profiled the institutional and student characteristics of universities (along with its respective types) and TVET colleges. While the system is designed to be differentiated based on its institutional characteristics (such as qualification offerings), it is no longer designed to be differentiated by students' characteristics. Cloete (2006) had previously argued that in the 1990s the system lacked the

necessary systemic changes. Now that the system has restructured itself, to what extent is it succeeding at redressing past inequalities?

I have shown that university types and TVET colleges differ in their number of enrolments, the types of qualifications they offer, and their fields of specialisation. Furthermore, a general pattern I observe is the student profile of traditional universities and Unisa tend to be more similar. In the same way, the profile of the other two university types, namely universities of technology and comprehensive universities, are more alike in terms of the student characteristics discussed. Universities of technology and comprehensive universities also appear comparable to the TVET colleges in terms of the types of qualification programmes and population group composition, but these institutional types differ in terms of their female representation. While TVET colleges have experienced a steady growth in raising its share of female students, universities of technology and comprehensive universities have seen a decline.

Yet, what is absent from this descriptive overview is a profiling analysis of students' socio-economic circumstances. It is important to understand this because of its significance for post-school education's role in overcoming some of South Africa's biggest socio-economic challenges – such as the high rates of poverty, inequality, and unemployment. If there are high levels of socio-economic heterogeneity in post-school student bodies at all institutions, it signals that students are coming from a diverse set of socio-economic backgrounds. On the other hand, if there are low levels of socio-economic heterogeneity within an institution or institutional type but high levels of socio-economic heterogeneity between different institutions or institutional types, then this may suggest evidence of horizontal stratification. Horizontal stratification at the post-school education level by institutional selectivity, field of study or type of program has the potential to reproduce socio-economic advantage in society (Torche, 2011). This is due to placement into a particular type of post-school institution being increasingly dependent on an individual's socio-economic origin.

While investment in education (inclusive of the quality of learning, access to educational resources, and the number of years of schooling) is viewed as an intervention to improve individuals' quality of life, their inherited socio-economic circumstances directly affect their level of educational attainment and the types of post-school institutions they select into, if they make it that far. As the number of students enrolling at post-school institutions expands and the system becomes more horizontally stratified, families who are relatively more socio-economically advantaged will use their resources to ensure their child attains a quantitatively similar but qualitatively superior post-school qualification (Lucas, 2001). This superior qualification is deemed financially rewarding and well-respected in the labour market. In this way, if qualitatively distinct institutional identities emerge at the same educational level (in this context, the post-school education level), then post-school institutions effectively reproduce the socio-economic class structure in the country.

Therefore, in the following section I discuss the current methods of measuring post-school students' socio-economic circumstances, along with their respective findings and shortfalls in being able to soundly address my research questions of interest.

5 Measuring socio-economic circumstances of university and FET/TVET college students

A single data set with sufficient information to measure the socio-economic profile of post-school student bodies from different institutional types does not exist. Universities and FET/TVET colleges naturally collect institutional data on student characteristics (such as gender, population group, and age) and hold comprehensive information on students' current education (such as the programme they are enrolled in, and which institution they are attending). However, the collection of student socio-economic information in this data set is limited. Information on students' household infrastructure, asset ownership, parental education, and income that is lacking in the institutional data would be of value to understand the socio-economic profile of different groups of students.

Alternatively, South Africa has nationally representative survey data that collects both educational and socio-economic information. An example of this data set is the census. It collects individual and household socio-economic information, based on where the individual is currently living. Furthermore, the data set has information on whether an individual is currently attending an educational institution, which type of institution they are attending, and whether it is a public (government) or private institution. This means, in theory, I am able to profile the mean socio-economic characteristics of students who are currently enrolled at a public university or a public FET/TVET college (but not by individual institutions or university types) for a fixed point in time using the census.

A constraint of using the census for this analysis is that it does not collect detailed information on students who live in residence. Therefore, it is anticipated that the sample of enrolled university and FET/TVET college students captured by the census will be undercounted. It then becomes a question of whether the students who are captured are representative of the true population of enrolled post-school students. To investigate this, I first compare the number of students enrolled at (public) universities and (public) FET/TVET colleges (as recorded by the institutional data and presented in the Department of Higher Education statistical reports for 2011) to the estimated total enrolment counts of the 2011 census (10% sample). These data are presented in Table 6.

The 2011 census data overcounts the number of enrolled university and FET/TVET college students when no public or private restriction is applied, both by 22%. However, when the 2011 census enrolled estimates are restricted to public institutions only, the number of university and FET/TVET college students are undercounted relative to the values recorded in the institutional data. The recorded number of public university students enrolled in 2011 was 938,201 students but the census total estimate was 773,794 students. This is an undercount of 164,407 students (or 18%). The recorded number of FET/TVET college students enrolled in 2011 was 400,273 students. The census total estimate for the number of enrolled FET/TVET students was 295,655 students; an undercount of 104,618 students (or 26%). Hence, both the number of public university and FET/TVET college students estimated in the 2011 census are undercounted. Only 82% of the number of public university students and 74% of the number of FET/TVET college students are, theoretically, captured in the 2011 census.

Table 6: Number of university and FET/TVET college students recorded by DHET and the 2011 census

	DHET 2011	2011 Census			
		Number	Lower CI	Upper CI	*Undercount*
University	**938,201**				
University, no public or private restriction		1,144,159	1,136,933	1,151,386	
University, public restriction		**773,794**	767,833	779,755	164,407
University, private restriction		206,527	203,430	209,625	
FET/TVET college	**400,273**				
FET/TVET, no public or private restriction		487,101	482,371	491,832	
FET/TVET, public restriction		**295,655**	291,963	299,347	104,618
FET/TVET, private restriction		125,849	123,441	128,258	

Source: Department of Higher Education and Training (Post-school Education and Training Statistics report, 2011). Statistics South Africa (Census 2011, 10% sample).

Notes: This table compares number of students enrolled university and FET/TVET college students as recorded in the institutional data versus the total counts of the 2011 census (10% sample). 95% Confidence Intervals (CI) are provided. Public and private restrictions do not add up to the no public or private restriction because of a third "don't know" option offered in the census questionnaire.

Although it was anticipated that the 2011 census total counts of enrolled post-school student would be lower than the true count, the undercount is larger than expected.[28] There are three potential reasons for this. First, it is possible that respondents are interpreting the enrolled institution categories differently to what the questionnaire is intending – something that Branson (2018) also noted. Second, respondents may not understand the difference between public and private institutions or do not know whether their institution is public or private. Third, some post-school students not captured by the census may fall within households who did not respond to the census. Given that wealthier households are less likely to respond to surveys (Lustig, 2020), this is particularly a problem if there is reason to believe that wealthier households are more likely to have individuals who enrol at post-school institutions.

To assess to what extent the sample of enrolled university and FET/TVET college students in the 2011 census are representative of the true population, I compare the individual characteristics (such as gender, population group, and age) to that recorded in the institutional data (previously discussed in Section 4.2). In addition, I estimate these enrolled students' household characteristics (type of geographical area, infrastructure, asset ownership, education, and annual household income). I present the 2011 census (10% sample) estimates of individual and household characteristics for the enrolled sample of public university and public FET/TVET college students for this year in Table 7, as well as the respective characteristics for the full South African population.

For the South African population, Table 7 shows South Africa has a predominantly young population (66.61% are younger than 35 years) and where over half (51.61%) of its population identify as female. Most of the country's population is from the Black population group (79.21%), with the remaining proportion of the population falling in the Coloured (8.98%), White (8.77%), Indian (2.52%), and Other

[28] For example, approximately 10.30% of university students spent at least some portion of the year in residence, according to the HEMIS data in 2011. Therefore, it is anticipated that the 2011 census total count of enrolled university students would be approximately 10% lower but instead it is 18% lower.

(0.52%) population group categories.

It is difficult to definitively determine whether the 2011 census individual characteristics align with the individual characteristics recorded in the institutional data. While some measures are similar, others are far off or lack the corresponding data to compare. The 2011 census estimates that 54.50% of university students identify as female. This is only 3.38 percentage points shy of the share of students who identify as female in the institutional data in 2011, at 57.88%. In terms of population group categories, the 2011 census reported 68.91% of university students are Black, 19.61% White, 5.44% Coloured, 5.25% Indian, and 0.79% Other. These estimates align almost perfectly with the shares reported in the institutional data, which were as follows: 68.26% Black, 18.90% White, 6.32% Coloured, 5.83% Indian, and 0.68% Other. There is no publicly reported data on the age distribution of public university students until 2018, so no direct comparison can be drawn here. Regardless of this missing comparison, the 2011 census estimates that the median age was 28. It is estimated that 29.86% of university students were 35 years or older in 2011, which is a relatively high share of students considering that students typically enrol at university around 18 years of age, if they enrol soon after exiting the basic education stream in grade 12. In the lower age categories, 7.25% of university students fell within the 15 to 19 year age category and 28.34% of students were 20 to 24 years old (the second highest share of students, after the 35 years or older age category). Hence, I would argue that the university sample is representative on gender and population group but I cannot validate whether it is representative on age.

No institutional data on gender, population group, and age has been publicly reported for FET/TVET college students in 2011. The earliest data point reported is 2013. Comparing the 2011 census individual characteristics estimates to these 2013 institutional values, the gender and population group estimates are in plausible reason of each other considering the 2 year difference between the different data sets. The 2011 census estimates 53.40% of FET/TVET college students identify as female; and in terms of population groups, 84.70% Black, 7.18% White, 5.96% Coloured, 1.83% Indian, and 0.33% fall within the Other population group category. The share of FET/TVET students recorded as White in the institutional data for FET/TVET colleges in 2013 was 2.2%, which is approximately 5 percentage points lower than what the censes estimates for 2011. The share of Black FET/TVET students recorded in the institutional data was approximately 5 percentage points higher in 2013 (90.3%), relative to the 84.7% estimated in the census for 2011. However, the share of students reported in the different age groups are vastly different. The share of students that fall within each age category for the 2011 census (relative to the 2013 institutional data in brackets) was as follows: 12.75% (24.7%) were 15 to 19 years; 38.16% (51.5%) were 20 to 24 years; 23.76% (14.5%) were 25 to 29 years; 10.27% (5.1%) were 30 to 34 years; and 15.05% (4.2%) were 35 years or older. Hence, the gender and population group estimates from the 2011 census and 2011 institutional data are within plausible reason of each other, but the age group estimates are not.

Despite the age group information in the institutional data being absent for university students or providing different estimates for the FET/TVET college students, I assume these samples of enrolled students in the 2011 census are closely representative of the true population as the gender and population group estimates are similar (although not perfect). Holding this assumption, the remaining discussion for Table 7 compares the population group profile of university and FET/TVET college students to the South African population, before discussing the household characteristics of these sample groups.

Table 7 shows FET/TVET colleges are more nationally representative; whereas universities have more than double the share of Indian and White students, and a lower representation of Black and Coloured students relative to the population. In terms of household characteristics, the mean household with an enrolled university or FET/TVET college student are more socio-economically advantaged than the mean household in the South African population. Furthermore, the mean household with an enrolled university student fairs even better, in terms of the household characteristics presented in Table 7, than the mean household with an enrolled FET/TVET college student.

The mean number of people living in households with either a university student (3.76 people per household) or FET/TVET college student (4.30 people per household) is smaller than the mean number of people living in the average household in South Africa (4.91 people per household). In addition to households with post-school students having a lower number of people per household, they also have larger households (in terms of the number of rooms) than the average household in South Africa which has 4.52 rooms per household. The mean number of rooms for a household with a university student is 5.20 rooms and the mean number of rooms for a household with a FET/TVET college student is 4.62 rooms.

A higher share of households with university enrolees reside in urban areas (82.86%) relative to the share of households with FET/TVET college enrolees that reside in urban areas (75.25%). Given that the share of households in urban areas is high, household access to electricity and water is also expected to be relatively high. Table 7 shows the vast majority of households have access to electricity for light: 85.49% of households in the population, 93.98% of household with university enrolees, and 91.88% of households with FET/TVET college enrolees have access to electricity for light. Access to piped water in the household is not as common as access to electricity. Less than half of households (45.20%) in the population have access to piped water. Households with university and FET/TVET college enrolees are better-off, whereby 74.13% and 57.61% of these households have access to piped water.

Assets such as a computer and internet connection have become increasingly essential materials in post-school education learning. These assets are often required to access study materials and complete assignments. While the percent of university or FET/TVET college enrolee households with a computer in their household is low (56.65% and 31.85% respectively), it is substantially higher than the national average (21.21%). Only 8.05% of households in the population have access to the internet. While the percent of university or FET/TVET college enrolee households with access to the internet is higher (24.23% and 9.78%, respectively) than the population average, these percentages are still small. The share of households with access to these assets emphasises the importance of on-campus resources in "level the playing-field" in access to resources for students.

Table 7: Mean socio-economic characteristics using 2011 census

	2011 Census		
	Population	University	FET/TVET college
Individual characteristics			
Gender (%)			
Female	51.61	54.50	53.40
Population group (%)			
Black	79.21	68.91	84.70
Coloured	8.98	5.44	5.96
Indian	2.52	5.25	1.83
White	8.77	19.61	7.18
Other	0.52	0.79	0.33
Age (%)			
Younger than 15 years	30.26	0.00	0.00
15 to 19 years	9.94	7.25	12.75
20 to 24 years	9.80	28.34	38.16
25 to 29 years	9.24	21.18	23.76
30 to 34 years	7.37	13.36	10.27
35+ years	33.39	29.86	15.05
Median age	25	28	24
Household characteristics			
Household size	4.91	3.76	4.30
Number of rooms	4.52	5.20	4.62
Type of geographical area (%)			
Urban	62.92	82.86	75.25
Infrastructure (%)			
Electricity (light)	85.49	93.98	91.88
Piped water	45.20	74.13	57.61
Asset ownership (%)			
Car	30.36	57.46	35.84
Cellphone	90.87	96.73	95.48
Computer	21.21	56.65	31.85
DVD player	63.71	73.99	68.28
Internet	8.05	24.23	9.78
Satellite television	27.41	49.87	34.97
Radio	69.35	77.44	72.23
Washing machine	33.94	53.97	37.79
Education (%)			
Post-school qualification	5.25	34.75	15.92
Annual household income (2011 ZAR)			
Mean income	R114 919.00	R279 200.40	R121 525.90
Median income	R28 900.50	R115 100.50	R28 900.50

Source: Statistics South Africa (Census 2011, 10% sample).
Notes: This table shows the mean socio-economic characteristics of the South African population and households with at least one public university or FET/TVET college student using the 2011 census (10% sample). The median income values of the population and FET/TVET college households are exactly the same given I worked with imputed mid-points (see Table D2 for details on the imputed mid-point values used).

Lastly, there are vast differences in education and income levels within households. In 2011, 5.25% of all households in South Africa had at least one person with a post-school qualification as their highest level of education. Relative to households with a FET/TVET college enrolee, this is almost 3 times higher (15.92%). Moreover, 34.75% of households with a university enrolee have at least one person with a post-school qualification as their highest level of education. Whereas the mean annual household income of FET/TVET college enrolee households (R121,525.90) is higher but close to the mean annual household income in the population (R114,919.00), the mean annual household income of university enrolee households (R279,200.40) is more than double that of the population and FET/TVET college enrolee households. Thus, according to the 2011 census data presented in Table 7, university enrolee households show the highest mean socio-economic advantage; and both university enrolee households and FET/TVET college enrolee households fair better than the mean population household.

Hence, with the 2011 census (10% sample) I am able to estimate the socio-economic characteristics of enrolled university and FET/TVET college students for a fixed point in time. Nevertheless, the 2011 census alone is not a perfect solution to measure the socio-economic background of post-school student bodies from different institutional types over time for four key reasons. First, I have shown censuses do not have detailed information on students who live in residence; thus, questioning to what extent the sample of students are representative of the true population of students. Second, the household information collected in the census reflects a student's current living arrangement. In other words, socio-economic information is collected based on where the student is living while attending a post-school institution, which may not be representative of the socio-economic circumstances of the household in which the student grew up. Third, the census does not collect the name of the institution where the student is enrolled, nor the type of university (i.e. comprehensive university, university of technology, traditional university, or Unisa). The closest restriction I can make with the 2011 census is knowing whether a student is enrolled at a public versus private university or FET/TVET college. This means it is impossible to compare student body socio-economic heterogeneity between institutions of a similar type (for example, between different universities or between different types of universities) or within a particular institution. Fourth, I can only assess enrolled students for fixed point in time, namely 2011. Hence, I cannot assess how the composition of university or FET/TVET students, based on their socio-economic circumstances, has changed over time. Other studies that have attempted to address aspects of the four caveats outlined above have either mapped students to the Global Positioning System (GPS) coordinates of their secondary school (Van Broekhuizen et al., 2016) or home address (Kerr et al., 2016). The researcher is then able to use the census information to determine the socio-economic profile of the surrounding area of the coordinates. Alternatively, Branson and Kahn (2018) used longitudinal data.

Van Broekhuizen et al. (2016) investigates university access, throughput, and dropout rates for the 2008 South African grade 12 cohort. To proxy for students' socio-economic background, Van Broekhuizen et al. (2016) construct a wealth index for each school among the 2008 grade 12 cohort.[29] The wealth index

[29] The authors use 12 questions which relate to the household's socio-economic status to build the wealth index. The paper does not document which 12 questions are used. The proportion of households in each small area with access to the 12 items is calculated as a percentage. This percentage is then converted into an index, referred to as the wealth index in the study, using a multiple correspondence analysis.

is constructed using the 2011 census community profile data at the SAL, whereby each SAL is assigned a wealth index value. The GPS coordinates of each school is then mapped to the appropriate SAL and allocated the respective wealth index value. The findings from Van Broekhuizen et al. (2016) show that a higher wealth index is associated with a higher university access rate. This suggests that students who attend schools in wealthier socio-economic areas are more likely to enrol in university relative to students who attend schools in poorer socio-economic areas, which is largely driven by students who attend schools in wealthier areas performing better than their peers who attend schools in poorer areas.

Although this wealth index does not necessarily reflect the socio-economic status of each student that attends a particular school, it presumably provides an accurate description of the socio-economic status of the immediate area surrounding the school. Thus, Van Broekhuizen et al. (2016) argue the wealth index is a better measure for students' socio-economic background compared to the school poverty quintile classification that is commonly used in the literature. However, the method adopted by Van Broekhuizen et al. (2016) fails to account for the socio-economic distribution present within a school as all students who attend a school within a given SAL are assigned the same wealth index value. There is also an underlying assumption here that students attend a school near where they reside, which may not be the case.

Kerr et al. (2016) also use the 2011 census community profile data but rather than assessing the socio-economic status of all grade 12 students, they assess the socio-economic status of applicants who apply to the University of Cape Town in an effort to gauge whether affirmative action is well targeted. From the 2011 census community profile data, Kerr et al. (2016) calculate the average income of the SAL as a proxy for socio-economic status of this geographic unit. In the application data, individuals provide their physical address. Kerr et al. (2016) use this address information and match it to the SAL using the gpsbound Stata package by Brophy et al. (2015). Here, each student in the University of Cape Town application data will be assigned an estimate of the average income of the SAL in which they reside as a proxy for their household's income level. Having the students' physical address allows Kerr et al. (2016) to offer an arguably improved socio-economic measure because they know where the student resides. However, much like Van Broekhuizen et al. (2016) could not account for the variation of students' socio-economic circumstances within a school; Kerr et al. (2016) do not know where a particular applicant lies in the SAL's socio-economic distribution.

Van Broekhuizen et al. (2016) and Kerr et al. (2016) have adopted methods that involve linking the 2011 census community profile data to student-level institutional data to estimate post-school students' socio-economic backgrounds, given that socio-economic information is limited (if even collected) in the student-level institutional data. Alternatively, Branson and Kahn (2018) used longitudinal data, namely the first four waves of the National Income Dynamics Study (NIDS), to estimate the socio-economic characteristics of grade 12 students prior to their decision to enrol in a post-school institution. The comparison of post-school enrolment options were restricted to public universities, FET/TVET colleges or students who choose not to enrol in a post-school institution (referred to as non-enrolees).

Branson and Kahn (2018) find students who enrol in FET/TVET colleges have similar socio-economic characteristics to those who do not enrol in a post-school institution at all, relative to the characteristics of

students who enrol in a public university. FET/TVET students and non-enrolees are from households of similar size, reside in similar geographical areas (rural versus urban) and attend secondary schools which are close to home. However, FET/TVET students are from households with marginally (but statistically significant) higher incomes and their parents have higher levels of educational attainment when compared to students who do not enrol in post-school institutions. Public university enrolees, on the other hand, are from households with more than three times the household income level than FET/TVET students and non-enrolees. Similarly, my estimates in Table 7 showed the mean annual household income of households with a university enrolee were more than twice the mean annual household income of households with FET/TVET college enrolees.

Furthermore, Branson and Kahn (2018) find the parents of public university enrolees have 2.5 more years of education and are more likely to reside in urban areas. Families who send their children to public universities also spend substantially more on school fees (based on the secondary school attended in grade 12), suggesting that they are capable of investing in education deemed to be of a higher quality because of its price tag. Conversely, school fees for FET/TVET students and non-enrolees are similar and notably lower than the average spent on school fees by university enrolees. Therefore, these findings by Branson and Kahn (2018) reveal that grade 12 students who go on to attend a public university are more socio-economically advantaged than grade 12 students who enrol at a FET/TVET college.

This section has acknowledged the possible methods and existing findings of measuring the socio-economic circumstances of university and FET/TVET college students. However, these approaches and data sets fall short on being able to assess the socio-economic profile of these student bodies from different institutional types and individual institutions; as well as not being able to offer a time trend analysis. This is where my research contributes to the growing understanding of the students who enrol at post-school institutions. By offering a complementary measurement approach of their socio-economic circumstances, I assess these aspects that have yet to be explored in previous research. I draw on both the institutional data to offer the best student-level information and the census data to offer the best socio-economic information – with the help of geospatial data to link the two sets of data – to improve the measurement of students' socio-economic characteristics one step further.

6 Generating a student-level data set with socio-economic information from the 2011 census

6.1 Constructing the data set

Given that I set out to descriptively examine differences in post-school students' socio-economic backgrounds, I require a student-level data set with socio-economic information. Institutional data from South African universities and Technical and Vocational Education and Training (TVET) colleges are recorded in centralised management information systems, namely the Higher Education Management Information System (HEMIS) and the Technical Educational Management Information System (TVETMIS). The information collected and compiled in the HEMIS and the TVETMIS is published in annual reports by the Department of Higher Education. However, the raw data collected is not publicly available. These databases contain information at the level of the student by year. Although these databases provide a comprehensive overview of students enrolled in public post-school institutions, this institutional data has limited information on the students' socio-economic backgrounds.

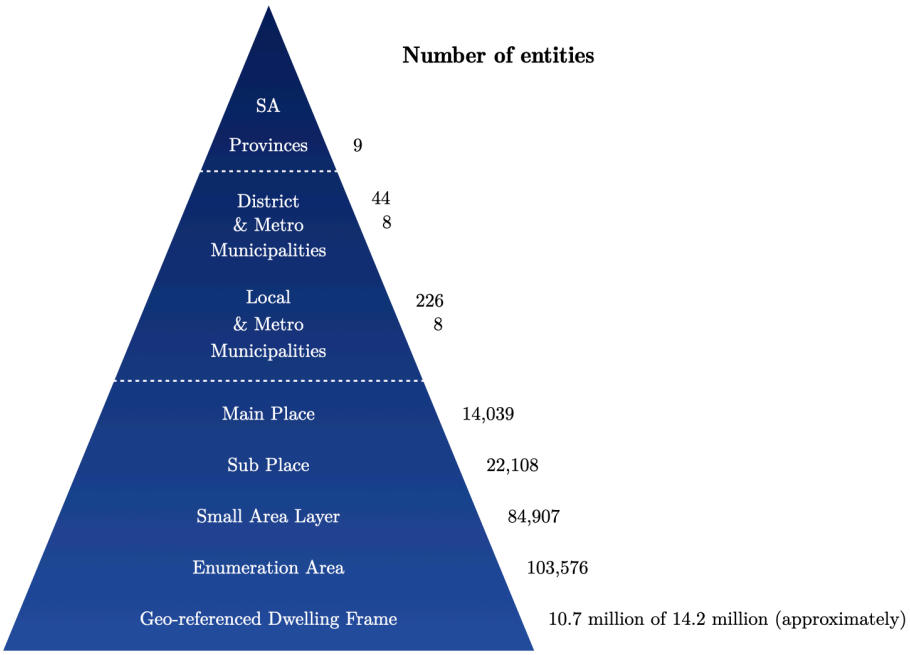
Therefore, I build on the method of linking institutional data to census data, as employed by Van Broekhuizen et al. (2016) and Kerr et al. (2016). This requires that I draw on the following data sets: the South African Census Community Profiles 2011 from Statistics South Africa, the HEMIS 2008 to 2019 from the DHET, the TVETMIS 2019 from the DHET, a SAL shapefile[30] 2013 from Statistics South Africa, and a postal code shapefile 2012 from STGroup. The data sets sourced from Statistics South Africa are the only publicly available data I use; the other data sets are not publicly available. This is the first known study to collectively draw on these data sets, particularly the use of a postal code shapefile. Thus, my approach to estimating post-school students' socio-economic circumstances adds a new dimension to the research being done in this area.

The latest census in South Africa was conducted in 2011 and collected information on both household characteristics (dwellling type, home ownership, household assets, access to services, and energy sources) and individual characteristics (age, population group, language, citizenship, migration, fertility, general health, mortality and economic characteristics of individuals, including employment status and employment activities). Given that the census holds nationally representative data on population, demographic, social, economic, and housing characteristics, this makes the data set ideal for estimating the socio-economic status of a geographical region. The South African census geography hierarchy is nested from the lowest level (dwelling unit) up to the highest level (province), as shown in Figure 18.

The 2011 census is made publicly available by DataFirst in two formats. One format of the 2011 census is a 10% sample. The data entries of the 10% sample are uniquely identifiable on the serial number of the household or person. The second format of the 2011 census is a database of community profiles based on the full record of respondents. The data is uniquely identifiable on the small area layer, the lowest geographic level of publicly released 2011 census data. Data at the enumeration area or dwelling

[30] A shapefile is commonly used to store geometric Geographic Information System (GIS) data.

Figure 18: Nested geography hierarchy for the 2011 census



Source: Statistics South Africa.
Notes: I use 2011 census community profile data at the Small Area Layer (SAL). This is the lowest geographic level of publicly released 2011 census data.

frame is not publicly available to preserve the confidentiality of the census respondents. The 2011 census community profile database consists of multiple data sets (in other words, one data set per selected questions from the 2011 census) that hold the number of individuals (or households) that responded to each question's category by the SAL. Due to rounding at the SAL (up from the dwelling unit), the total number of individuals (or households) per SAL may not consistently add up across data sets.

As identified in the previous section, the 10% sample format of the 2011 census is not appropriate for what I need to carry out my analysis. Furthermore, the lowest geographic level in the 10% sample is municipality. Instead, I opt for the community profile format of the 2011 census because it offers the lowest geographic level of the data that is publicly available, namely the SAL. The community profile format was also used by Van Broekhuizen et al. (2016) and Kerr et al. (2016), from whom I have drawn inspiration for my approach. The individual-level community profile data sets I assess are: home language, highest level of education, employment, type of geographical area.[31] The household-level community profile data sets I assess are: access to electricity for light, access to piped water, access to a car, access to a cellphone, access to a computer, access to a DVD player, access to the internet, access to a satellite television, access to a radio, access to a washing machine, and annual household income.

For each of these community profile data sets, I divide the number of individuals (or households) that fall within the response of interest by the total number of individuals (or households) that responded to

[31] For the community profile data, the type of geographical area data is only available at the individual-level and not the household-level.

the question for each SAL. This gives me the share of individuals (or households) that have access to the socio-economic variable of interest, which in turn form the indicators that proxy students' socio-economic background at the SAL.[32] I assume constant socio-economic characteristics across the area of each SAL polygon.

From the raw HEMIS (2008 to 2019) and TVETMIS (2019) databases, I am most interested in using the following variables: institution attended, qualification type, and entrance category; in addition to student demographic information on gender, population group and – most critically – home postal code.[33] With the students' home postal code, I can identify the community region where each student resides. However, postal codes do not form part of South Africa's census geography hierarchy and, instead, overlap with the SAL boundaries. Given that the socio-economic indictors are first calculated at the level of the SAL, I need to take into consideration the proportion of each postal code that falls within the different SALs. These proportional areas are then used to weight the SAL socio-economic indicators and, in turn, form socio-economic indicators at the level of the postal code. These weights are calculated using both the SAL shapefile and postal code shapefile.

The the SAL shapefile produces a map of South Africa with the SAL geography boundary lines, as depicted in Figure 19. For the context of this explanation, each SAL unit making up a map of South Africa can be thought of as a unique polygon area. Likewise, the postal code shapefile produces a map of South Africa with the postal code boundary lines, as depicted in Figure 20. While the SAL shapefile is publicly available by Statistics South Africa on DataFirst, spatial postal code data is not easy to come by.[34] I, therefore, source a South African postal code shapefile from STGroup, a Spatial Asset Management company. This is a file compiled by STGroup using the formal suburb boundaries from TomTom, one of their main data vendors. The company then matches these boundaries to the four-digit postal code using the suburb name, as per the list of South African postal codes on the official South African Postal Office website. Outside of these formal boundaries, STGroup employs a Voronoi technique. This process involves taking each known point outside the formal suburb boundaries and drawing a boundary line at the midpoint. Access to the postal code shapefile compiled by STGroup made the method I carry out viable because the postal code shapefile makes it possible to take into account the area of postal codes.

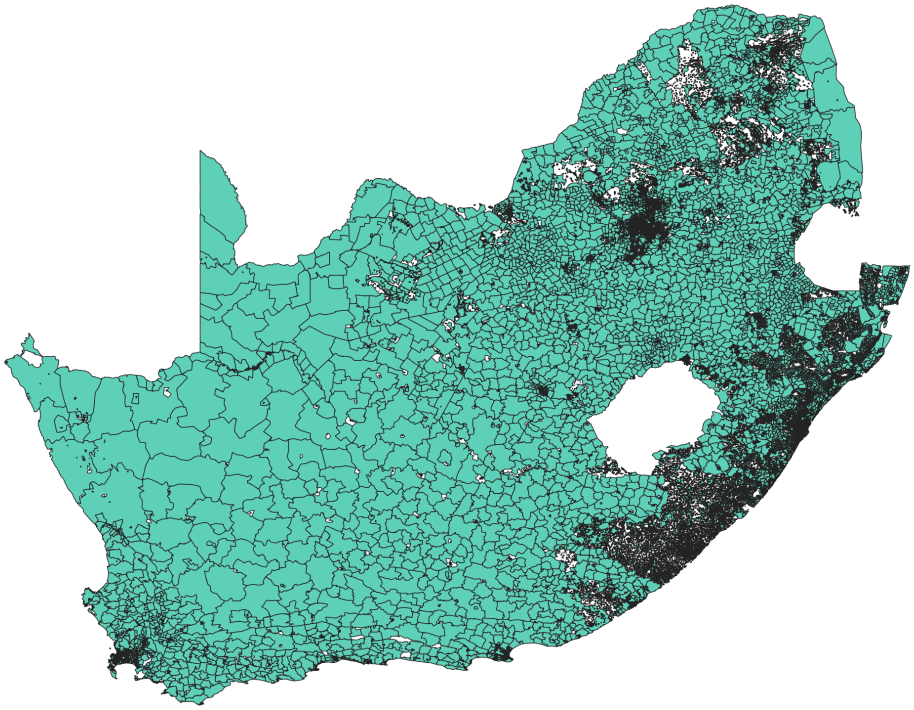
South African postal codes are made up of a four-digit number and represent a particular area in the country. The postal codes range from 0001 to 9999 and are sorted into postbox and street codes. For this research, I am only interested in using the street codes, otherwise referred to as the postal code of the

[32] For example, one of the community profile data sets I use captures access to the internet at the household level. It is expressed as the number of households that mainly access the internet from: (1) home, (2) cellphone, (3) work, (4) elsewhere, or (5) no access to the internet. As the question asks what is the main form of access, respondents can only select one category. I divide the first category, the number of households that mainly access the internet from home, by the total number of household responses to calculate an estimate for the share of households with household access to the internet within each small area layer. As a hypothetical example, I may calculate that 15% of households within a particular small area layer have household access to the internet. This then becomes one of my SAL characteristic estimates. Please see Appendix D for further details on the construction of the indicators.

[33] This research forms part of the Siyaphambili project, which has been granted access to use selected variables from the raw HEMIS (2008 to 2019) and TVETMIS (2019) databases. Please see Appendix G for the attached ethics approval.

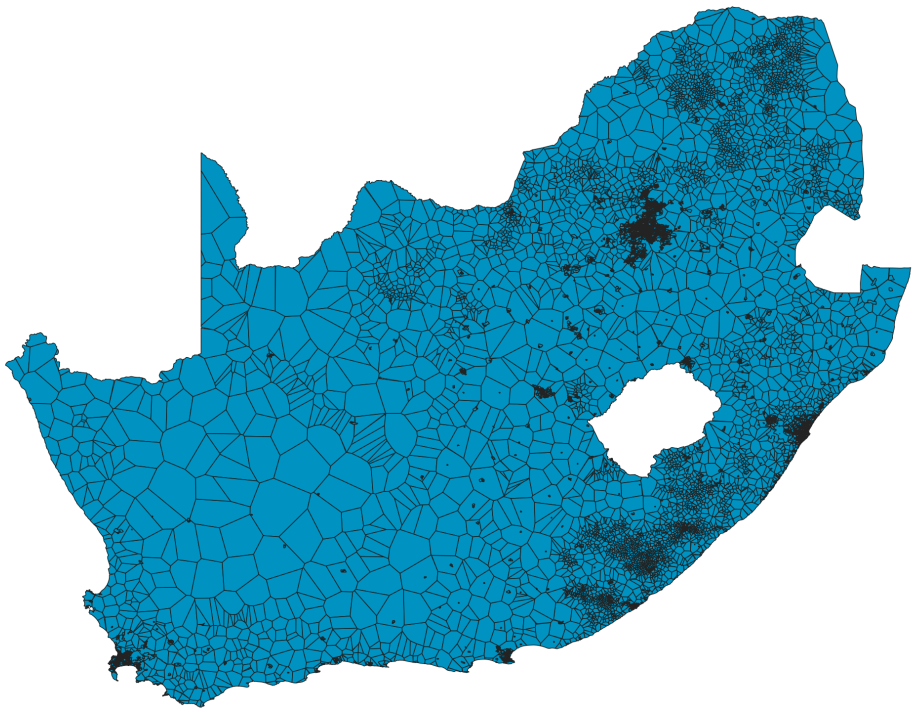
[34] Although a list of South Africa's domestic postal codes can be downloaded from the South African Postal Office website, no spatial data set is released by them.

Figure 19: Small Area Layer (SAL) polygons for South Africa



Source: Statistics South Africa (2013).
Notes: There are 84,907 unique SAL polygons. Areas that are left blank have no households, according to the 2011 census. There are, on average, 609.66 individuals per SAL and 170.16 households per SAL.

Figure 20: Postal code polygons for South Africa



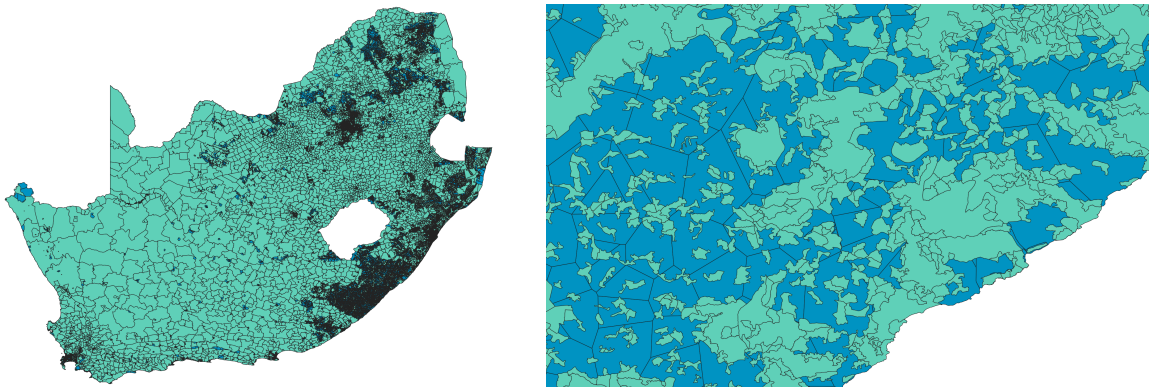
Source: STGroup (2012).
Notes: There are 20,861 unique postal code polygons and 2,833 unique street (home) postal codes, according to the shapefile requested from STGroup.

home street address. Postal codes were historically designed for the sorting and delivery of mail. This means postal codes are linked to the nearest city with a post office distribution centre, causing postal code boundaries to cross the other geographic area boundaries used in the census (Lombaard, 2009).

Some postal codes correspond to multiple postal code polygons. For example, there may be a polygon on the South African map and a separate polygon somewhere else on the South African map (usually not too far apart), but each polygon has the same four-digit postal code. This is visually apparent when working with GIS software, or simply Googling postal codes with this characteristic. Thus, each postal code polygon making up the map of South Africa is not unique on the corresponding four-digit postal code.

After first creating a data set with the SAL socio-economic indicators, I use the GIS software to create a second data set which combines information from both the SAL shapefile and postal code shapefile. The main purpose of this second data set is to calculate the proportion of each postal code polygon that falls within each small area layer polygon. To compare the South African postal code boundaries with the small area layer boundaries, I overlay the shapefiles. It is evident in Figure 21a that the SAL polygons do not cover the full area of South Africa, like postal codes polygons. These excluded areas are blank because they have no households for the census to sample – for example, areas covered by mountains.

Figure 21: Small Area Layer (SAL) polygons do not cover the full area of South Africa



(a) Overlap of SAL and postal code polygons

(b) Overlap of SAL and postal code polygons (magnified)

Source: Statistics South Africa (2013) and STGroup (2012).
Notes: The blue area is the postal code polygons. In the foreground, the green area is the SAL polygons and they do not cover the full area of South Africa. Areas of South Africa not covered by a SAL have no households to sample, according to the 2011 census. I assume constant socio-economic characteristics across the area of each SAL polygon. It is, therefore, plausible to weight the socio-economic information of the empty portion of a postal code as zero because there are no households or individuals living here.

As some postal codes have multiple postal code polygons, I first create a unique identifier code for each postal code polygon and calculate its area. I then use the union geoprocessing tool found in GIS software to calculate the proportion of each postal code polygon that falls within each SAL polygon. The SAL shapefile acts as the input layer and the postal code shapefile acts as the overlay layer. The union

tool cuts the input layer into smaller pieces and calculates the area of the geometric intersection between the SAL polygons and postal code polygons. The geometric intersection is added as an additional new layer to the data set. I then calculate the area of the geometric intersection as a percentage of the postal code polygon. I refer to these percentages as the *intersection weight*. These percentages sum to 100 for each postal code polygon. This data set, with the newly calculated weights, is then merged with the SAL socio-economic indicator data set on the SAL.

$$intersection\ weight = \frac{area\ geometric\ intersection}{area\ postal\ code\ polygon} \times 100 \tag{1}$$

The merged GIS and SAL socio-economic indicator data set need to be uniquely identifiable on the postal code (and not on the postal code *polygon*) to merge with the student-level institutional data. For postal codes with multiple postal code polygons, the sum of the postal code polygon areas equals the total area for the postal code. I can, therefore, calculate the area of the postal code polygon as a percentage of the total area for the postal code. This is the *postal code weight*.

$$postal\ code\ weight = \frac{area\ postal\ code\ polygon}{area\ total\ postal\ code} \times 100 \tag{2}$$

Thus to get to a final data set of socioeconomic characteristics that are unique of postal code, this process entails weighting, and then re-weighting, the mean SAL socio-economic indicators within each postal code. I first weight the SAL socio-economic indicators by the intersection weight to calculate the mean SAL socio-economic indicators within each postal code polygon. I then need to re-weight these mean postal code polygon socio-economic indicators by the postal code weight to calculate the socio-economic estimates within each postal code. These data I merge with the student-level institutional data to achieve my final desired student-level data set with socio-economic information from the 2011 census.

6.2 Analysis of where students in the sample reside

6.2.1 Defining the sample

A consequence of the method I adopt is that when I merge the postal code level data set (that contains the accompanying socio-economic indicators) with the student-level institutional data, students with an invalid postal code are dropped from the analysis. Reasons for invalid postal codes include: the student's home postal code is a foreign address, the postal code recorded does not exist in South Africa's postal code data set, or no postal code is recorded at all.

Table 8 shows the number of university students recorded in HEMIS, the number of these students with valid postal codes and the number of students with invalid postal codes for each year of data (2008 to 2019). With each year that has passed since 2010, the number of students with invalid postal codes has declined as the total number of students enrolling has increased. Hence, the share of university students dropped from the data set I construct has reduced from 15.60% of university students in 2008

to 7.66% of university students in 2019.

Given that I want to compare the socio-economic characteristics of student bodies between and within institutions over time, I explicitly restrict the sample of university students in my analysis to undergraduate, first-time entering students.[35] Of the undergraduate, first-time entering sample of university students, a lower share have invalid postal codes compared to the full sample of university students. Table 8 further shows the percent of undergraduate, first-time entering students with an invalid postal code declined from 13.07% in 2008 to 6.61% in 2019.

I do not restrict the 2019 TVET sample to first-time entering students because I do not have this information. Furthermore, I do not believe it it necessary for the TVET sample to be restricted because I only assess a single point in time (therefore, students remaining in the system longer than others is not a concern) and there are no postgraduate qualifications at TVET colleges. Nevertheless, 6.17% of the full sample of 2019 TVET students have an invalid postal code. Hence, a similar share of students are dropped from the university and TVET college data due to invalid postal codes.

Table 8: Number of students with valid and invalid postal codes

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
HEMIS	805,403	843,592	899,122	944,095	958,103	991,658	974,656	990,646	980,938	1,042,083	1,090,185	1,081,147
Valid postal codes	679,721	712,610	758,178	816,413	833,560	896,198	882,817	900,826	892,983	954,868	1,003,402	998,364
Invalid postal codes	125,682	130,982	140,944	127,682	124,543	95,460	91,839	89,820	87,955	87,215	86,783	82,783
(%)	**15.60**	**15.53**	**15.68**	**13.52**	**13.00**	**9.63**	**9.42**	**9.07**	**8.97**	**8.37**	**7.96**	**7.66**
HEMIS UGFTEN	152,261	164,507	168,413	178,831	169,645	158,493	168,631	172,149	159,690	194,184	209,493	188,631
Valid postal codes	132,353	145,528	147,417	160,996	153,117	146,038	155,437	158,911	147,052	180,620	195,729	176,161
Invalid postal codes	19,908	18,979	20,996	17,835	16,528	12,455	13,194	13,238	12,638	13,564	13,764	12,470
(%)	**13.07**	**11.54**	**12.47**	**9.97**	**9.74**	**7.86**	**7.82**	**7.69**	**7.91**	**6.99**	**6.57**	**6.61**

Source: Author's constructed data set.
Notes: This tables shows the total number of university students recorded in HEMIS and what number of these students have valid or invalid postal codes. The percent of students with invalid postal codes is also expressed. The same information is shown for undergraduate, first-time entering students (UGFTEN) recorded in HEMIS.
The number of TVET college students recorded in TVETMIS for 2019 is 588,917 students. 552,564 (93.83%) of these students have valid postal codes and 36,353 (6.17%) students have invalid postal codes. Data on first-time entering TVET college students is unavailable.

Table 9 shows the majority of university students dropped from my sample are South African rather than being students with foreign nationalities. In 2008, 54.85% of the group of students with invalid postal codes are South African. This increased to 71.92% in 2019. Although this share has increased from 2008 to 2019, this is a function of the total number of students with invalid postal codes that are South African declining at a slower rate (10,920 to 8,969 students, shown in Table 9) than the total number of all students with invalid postal codes (19,908 to 12,470 students, shown in Table 8). The South African students dropped in this group of students with invalid postal codes are most likely to have entered postal codes that do not exist in South Africa's postal code database or they have missing postal code information recorded.

Furthermore, Table 9 shows between 97% to 99% of students with valid postal codes are South African. The remaining 1% to 3% of these students are foreign. I choose to keep these students in the

[35] Table E2 shows the mean socio-economic characteristics of postgraduate students is higher than undergraduate students.

sample because, although they have foreign nationalities, they could have lived in South Africa for a reasonable number of years given that they have a valid South African postal code recorded as their home address.

Table 9: Share of university students that are South African

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Valid postal codes (%)	98.60	98.55	98.29	98.10	98.04	97.91	97.88	97.87	98.95	98.30	99.04	99.04
	130,495	143,411	144,903	157,944	150,114	142,980	152,149	155,528	145,509	177,547	193,858	174,462
Invalid postal codes (%)	54.85	64.25	63.35	56.82	58.64	65.23	65.20	63.79	65.09	69.17	73.02	71.92
	10,920	12,194	13,301	10,133	9,692	8,125	8,602	8,444	8,226	9,382	10,050	8,969

Source: Author's constructed data set.
Notes: This table shows amongst the group of university students (sample of UGFTEN students) with valid postal codes, 97% to 99% are South African; and amongst the group of university students with invalid postal codes, 54% to 73% are South African. The corresponding number of students are presented in gray.

The share of university students dropped by university type varies. Table 10 presents this breakdown. From 2008 to 2019, the largest share of students with invalid postal codes (typically between 30% to 40%) attend traditional universities. The second largest share of students with invalid postal codes (31.89% in 2008) stem from Unisa. Unisa has since accounted for a smaller share of dropped students – accounting for as low as 3.79% of dropped students in 2016. Comprehensive universities and universities of technology have each accounted for between 14% and 30% of university students with invalid postal codes. Since 2013, the second largest share of dropped students attended universities of technology.

Table 10: Share of university students with invalid postal codes by university type

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Comprehensive (%)	14.03	17.73	15.53	18.52	17.09	22.10	23.62	26.11	26.48	22.66	21.98	23.22
	2,794	3,365	3,260	3,303	2,825	2,752	3,116	3,456	3,346	3,074	3,026	2,896
Technology (%)	19.15	22.10	19.04	21.55	23.04	29.99	28.63	24.80	27.26	26.28	23.23	23.50
	3,812	4,195	3,997	3,843	3,808	3,735	3,777	3,283	3,445	3,565	3,198	2,930
Traditional (%)	34.93	39.66	43.61	31.66	32.33	39.04	39.77	39.31	42.48	38.87	36.21	37.49
	6,953	7,527	9,157	5,646	5,344	4,862	5,247	5,204	5,368	5,272	4,984	4,675
Unisa (%)	31.89	20.51	21.82	28.28	27.54	8.88	7.99	9.78	3.79	12.19	18.57	15.79
	6,349	3,892	4,582	5,043	4,551	1,106	1,054	1,295	479	1,653	2,556	1,969
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	19,908	18,979	20,996	17,835	16,528	12,455	13,194	13,238	12,638	13,564	13,764	12,470

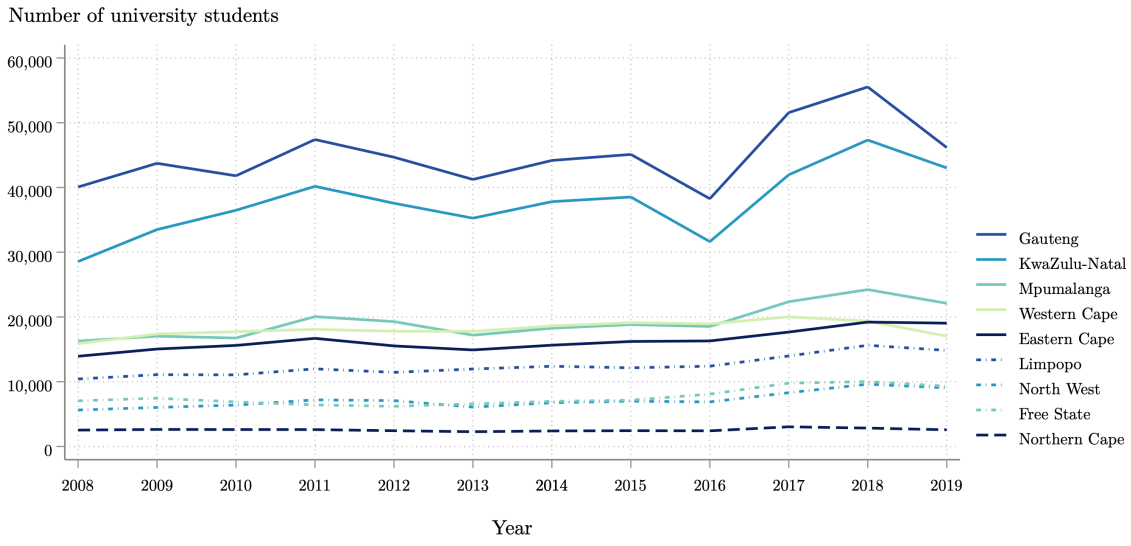
Source: Author's constructed data set.
Notes: This table shows, from 2008 to 2019, the largest share of university students (sample of UGFTEN students) with invalid postal codes attend traditional institutions. The corresponding number of students are presented in gray.

Only students with valid postal codes remain in the sample. All further analysis in the paper focuses on the sample of Undergraduate, First-Time Entering (UGFTEN) students enrolled at universities and the full sample of students enrolled at TVET colleges. Students may be South African or of foreign nationality but Table 9 showed approximately 97% to 99% of the university sample is South African.

6.2.2 Provincial representation

Assessing where students reside at the province level offers a broader and more familiar overview of where students are coming from. It serves as a first approach to checking the data. The province in which students reside, using their home postal code, can be estimated as postal codes are systematically coded according to different provincial regions. However, marginal measurements errors may arise given that postal code boundaries do not perfectly align with province boundaries. A plot of the sample of university students, according to their home province, from 2008 to 2019 is shown in Figure 22.

Figure 22: Number of university students by home province



Source: Author's constructed data set.
Notes: This figure shows Gauteng and KwaZulu-Natal are the home provinces for the largest number of university students (UGFTEN sample restriction).

Collectively, Gauteng and KwaZulu-Natal are the home provinces for the majority of students in the sample. These two provinces accounted for 68,636 students (51.86%) in 2008 and increased to a high of 102,858 students (52.55%) in 2018 before declining to 89,183 students (50.63%) in 2019. The large representation of Gauteng and KwaZulu-Natal aligns with expectations, given that these provinces held the highest number of individuals age 15 to 34 years, according to the 2019 mid-year population estimates produced by Statistics South Africa.[36]

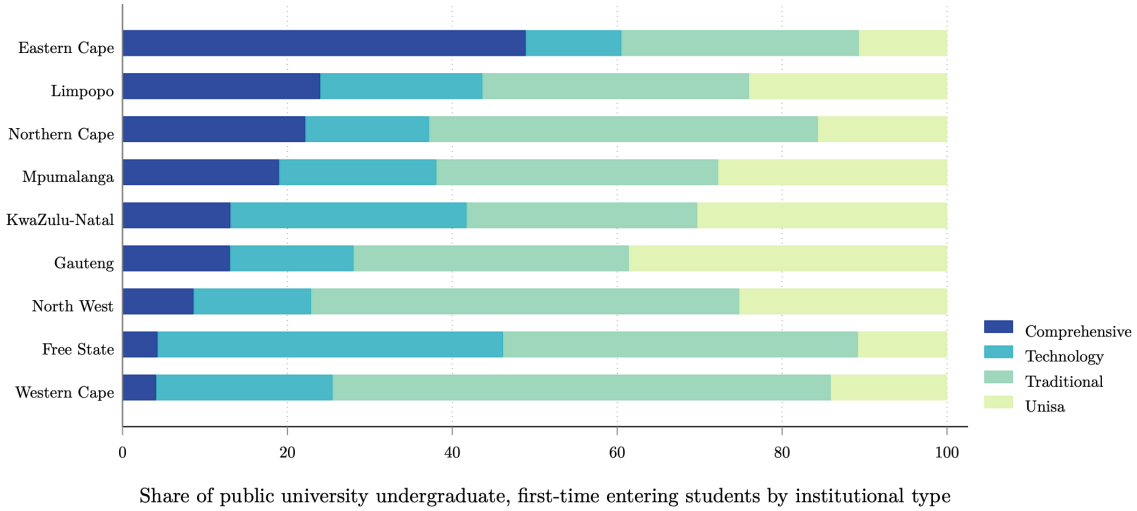
The pronounced peaks and troughs for Gauteng and KwaZulu-Natal – in 2011, 2013, 2016 and 2018 – follow the same pattern as the overall enrolment trend for Unisa (see Figure 8). Furthermore, as per Figure 23, Gauteng and KwaZulu-Natal are the home provinces with the largest share of students enrolled at Unisa (38.58% and 30.28% respectively) according to 2019 data. Hence, the trend in these provinces seems predominantly influenced by students who attend Unisa.

[36] In 2019, the mid-year population estimates for individuals age 15 to 34 years were 5,695,021 in Gauteng and 4,056,250 in KwaZulu-Natal (Statistics South Africa, 2020).

The Northern Cape accounts for the lowest number of university students (2,578 students in 2019) and it is also the province with the smallest number of individuals age 15 to 34 years, according to the mid-year population estimates (Statistics South Africa, 2020). Likewise, the North West (9,059 students in 2019) and Free State (9,293 students in 2019) account for the second and third lowest number of undergraduate, first-time entering university students; and are the provinces with the second and third smallest number of individuals age 15 to 34 years, according to the mid-year population estimates. This pattern of aligning with mid-year population estimates is evident across the provinces, except for Mpumalanga. In Figure 22, Mpumalanga is the province with the third highest number of students entering university but ranks sixth (out of nine provinces) according to the number of individuals age 15 to 34 years mid-year population estimates. It is possible that the unexpected large number of students identified as residing in Mpumalanga (relative to the mid-year population estimates) could be driven by students residing in postal codes that fall near the province boundary and may be distorting the true number of university students whose home province is Mpumalanga.

Figure 23 shows the share of students attending each university type within each province associated with students' home postal code.[37] Almost half (48.92%) of the students in the sample, whose home province is the Eastern Cape, attend a comprehensive university. No other province attracts this high a share of students to comprehensive universities. An influential factor here is 10 of the 21 comprehensive university campuses are located in the Eastern Cape.[38]

Figure 23: Distribution of university type enrolment by students' home province in 2019



Source: Author's constructed data set.
Notes: This figure shows the share of students attending each university type within each province associated with students' home postal code.

[37] See Table C3 for the number and share of students by university type within each province in 2019.

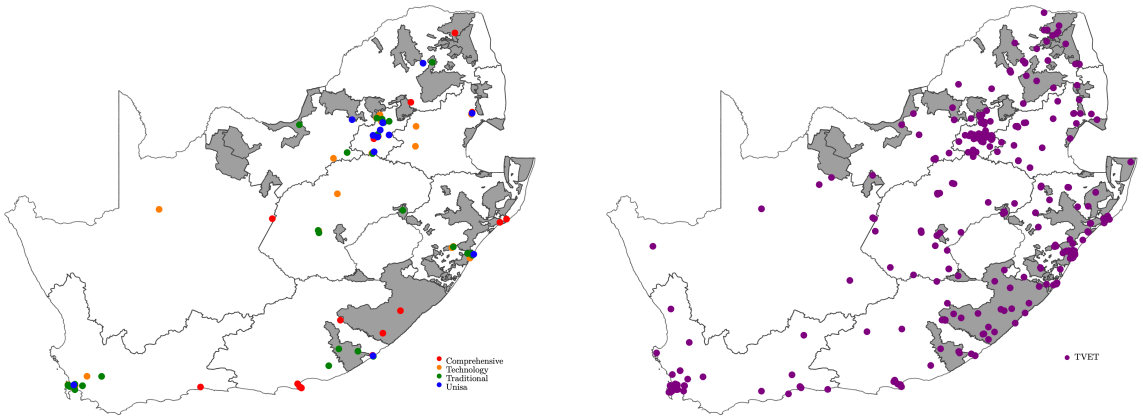
[38] See Figure 24a, and Table C2.

The largest share of students who grew up in the Western Cape (60.39%), North West (51.94%), Northern Cape (47.17%), Free State (43.03%), Mpumalanga (34.19%) and Limpopo (32.34%) enrol at a traditional university, where 12 of the 37 traditional university campuses are situated in the Western Cape. The Free State has an equally high share of students (41.92%) who enrol at a university of technology. However, the Free State only holds 2 university of technology campuses, whereas Gauteng, KwaZulu-Natal and the Western Cape all have 8 university of technology campuses. Lastly, as previously mentioned, the largest share of students in Gauteng (38.58%) and KwaZulu-Natal (30.28%) enrol at Unisa.

Hence, Figure 23 depicts some university types attract more students than others within the province where students grew up. A number of factors could influence this. One aspect to explore would be the number of institutional types within each province, as a students' choice of institution may be influenced by the types of institutions they have seen located in their community, the types of institutions attended by friends and family in the region, or proximity to an institution if this is a constraint for students.

Figure 24a plots the university campuses by university type and Figure 24b plots the TVET college campuses to support the above discussion.[39] Even though there are 26 university institutions and 50 TVET college institutions, each institution can have more than one campus. Figure 24a depicts 21 comprehensive university campuses, 31 university of technology campuses, 37 traditional university campuses, and 14 Unisa campuses; and Figure 24b depicts 323 TVET college campuses distributed across South Africa.

Figure 24: Where are the university and TVET college campuses located?



(a) University campuses (b) TVET college campuses

Source: Youth Explorer website (2021).
Notes: The former homeland areas are shaded in gray. Only public university and TVET college campuses are depicted. In Figure 24a, there are 21 comprehensive university campuses, 31 university of technology campuses, 37 traditional university campuses, and 14 Unisa campuses. In Figure 24b, there are 323 TVET college campuses distributed across South Africa.

[39] Table C2 presents the underlying data for the maps depicted in Figure 24, showing the number of university campuses per province by university type.

Students do not necessarily attend an institution in the same province as the home in which they grew up. Table 11 shows the number of students by the province associated with home postal code (home province) and the number of students attending an institution within that province in 2019. The latter takes into account the number of enrolments at the institutions within each province. The ratio of the two numbers is also presented. The ratio serves as a crude estimate of whether institutions in that province are attracting students from outside provinces (a ratio less than 1); or whether students in that province are attracted to an institution within a province different to their home province (a ratio greater than 1). When the ratio is greater than 1, it may suggest that there is a higher demand for enrolments relative to the capacity of institutions within these provinces. Alternatively, beyond the realms of capacity constraints, students may be attracted to institutions in a province outside their home province based on institutional characteristics. Students attending Unisa have been excluded from this table because Unisa has campuses present in multiple provinces and is a predominantly distance learning institution. These data are presented in such a format where it is not possible to tease out the number of Unisa enrolments per campus.

Table 11: Number of students by home province and number of students attending an institution within each province

	University students[a]			TVET college students		
	Home province	Institution province	Ratio	Home province	Institution province	Ratio
North West	6,779	11,536	**0.59**	31,588	27,220	**1.16**
Gauteng	23,456	38,573	**0.61**	107,518	132,246	**0.81**
Free State	8,291	11,708	**0.71**	38,952	39,456	**0.99**
Western Cape	14,635	19,257	**0.76**	73,650	78,295	**0.94**
Eastern Cape	17,012	16,410	**1.04**	63,063	58,698	**1.07**
KwaZulu-Natal	29,993	22,586	**1.33**	99,800	102,149	**0.98**
Limpopo	11,280	7,419	**1.52**	59,463	64,618	**0.92**
Northern Cape	2,175	685	**3.18**	16,903	13,876	**1.22**
Mpumalanga	15,974	1,421	**11.24**	61,744	36,123	**1.71**
Total	129,595	129,595		552,681	552,681	

Source: Author's constructed data set.
Notes: Home province is an estimated calculation based on the home postal code. Postal code boundaries do not perfectly align with province boundaries, hence it is an estimated calculation.
[a] University students are restricted to UGFTEN. Students attending Unisa are excluded from this table because Unisa has a presence in multiple provinces.

The North West, Gauteng, Free State and Western Cape have a ratio less than one; and the Eastern Cape, KwaZulu-Natal, Limpopo, Northern Cape and Mpumalanga have a ratio greater than one for the sample of university students. The ratios for TVET college students are closer to 1 than for university students across all provinces. The North West, Eastern Cape, Northern Cape and Mpumalanga are the only provinces for TVET college students with a ratio greater than one.

Further calculations estimate that 78,265 students attend a university in the same province as their home postal code out of a total 129,595 UGFTEN university students in 2019 (excluding those enrolled Unisa). By comparison, 469,295 students attend a TVET college in the same province as their home postal code out of a total 552,681 TVET college students in 2019. Hence, 60.39% of university students in the sample attended a university in the same province as their home; whereas 84.91% of TVET college students attended a TVET college in the same province as their home.

6.2.3 Postal code representation

In assessing whether student bodies at different types of institutions are from similar socio-economic backgrounds, I am essentially measuring how the postal code representation differs amongst these various student bodies. Understanding where students reside, the volume of students in different postal codes, and how this information has changed between 2008 to 2019 is imperative to interpreting whether student bodies at different types of institutions are from similar (or different) socio-economic backgrounds. Given that the socio-economic measures assigned to each student in the data set are at the postal code level, all students from a particular postal code are assigned the same socio-economic measures. Furthermore, the socio-economic information is from one point in time (2011 census). Therefore, all variation is from representation of postal codes and their distribution within and between institutions.

For university students, I assess 12 years of data from 2008 to 2019. Table 12 shows the number of years a postal code is the home postal codes of at least one student. Of the total 2,833 postal codes in the country, 192 are not the home postal code of any student between 2008 and 2019.[40] There are 131 postal codes that appear once over the 12 years of data. Overall, 1,628 (57.47%) postal codes are represented throughout the 12 years of data.

Table 12: Number of years a postal code is the home postal code of at least one university student

Years of data	0	1	2	3	4	5	6	7	8	9	10	11	12	**Total**
Number of postal codes	192	131	100	82	59	54	58	69	88	85	111	176	1,628	**2,833**
Percentage of postal codes	6.78	4.62	3.53	2.89	2.08	1.91	2.05	2.44	3.11	3.00	3.92	6.21	57.47	**100.01**

Source: Author's constructed data set.
Notes: This table shows 57.47% of postal codes are the home postal code of at least one university student (UGFTEN restriction) throughout the 12 years of data. Only 6.78% of all postal codes are not identified as the postal code of a student registered at a university from 2008 to 2019.

Table 13 presents the number and percent of postal codes represented in each year of the data. Between 2008 to 2019, the number of unique postal codes represented amongst the sample of university enrolees stayed relatively stable with the lowest number of postal codes represented in 2016 (2,135 postal codes or 75.74% of postal codes) and the highest number of postal codes represented in 2010 (2,190 postal codes or 77.69%). From Table 12, I know that 1,628 of these postal codes are consistently represented in the data set throughout all 12 years of data.

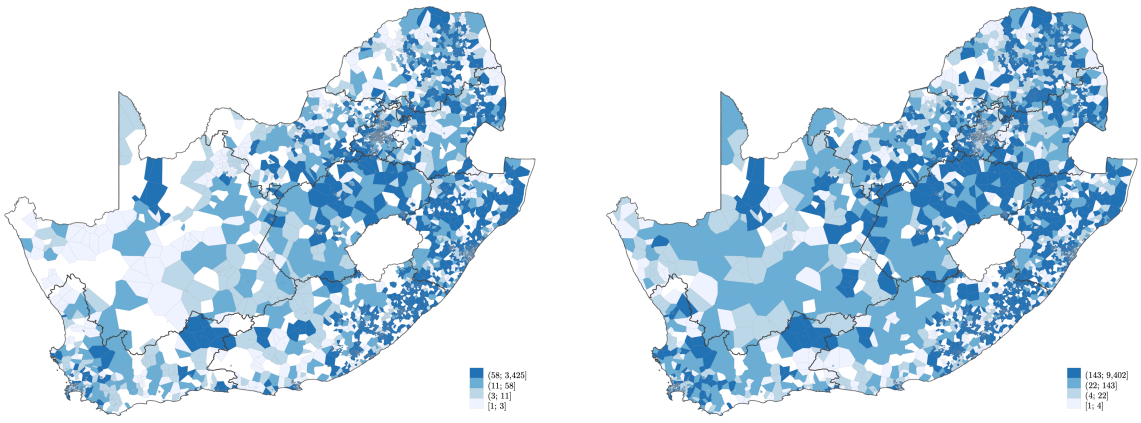
[40] There are 14 postal codes where no SAL polygon falls within it. Therefore, these 14 postal codes are automatically dropped from the analysis and are included in the 192 postal codes with no UGFTEN university student between 2008 and 2019.

Table 13: Number of students per postal code

	Postal codes		Students	Distribution of students per postal code							
	n	%		p10	p25	p50	p75	p90	Max	Mean	Std.
University											
2008	2,145	76.09	132,353	1	4	14	53	167	2,616	61.70	139.10
2009	2,178	77.26	145,528	1	4	14	56	185	2,967	66.82	151.91
2010	2,190	77.69	147,417	1	4	14	57	187	3,032	67.31	158.56
2011	2,169	76.94	160,996	2	5	15	61	208	2,056	74.23	166.87
2012	2,177	77.23	153,117	1	4	14	56	190	2,815	70.33	166.56
2013	2,168	76.91	146,038	1	3	12	49	188	3,221	67.36	166.38
2014	2,161	76.66	155,437	1	4	12	53	200	3,689	71.93	180.22
2015	2,137	75.81	158,911	1	3	11	54	208	3,574	74.36	188.05
2016	2,135	75.74	147,052	1	3	11	51	197	3,453	68.88	165.30
2017	2,164	76.76	180,620	1	3	12	58	242	3,719	83.47	211.68
2018	2,173	77.08	195,729	1	3	12	63	257	3,774	90.07	234.44
2019	2,168	76.91	176,161	1	3	11	58	236	3,425	81.26	203.26
TVET											
2019	2,443	86.66	552,564	2	4	22	144	600	9,402	226.18	597.14

Source: Author's constructed data set.
Notes: This table presents the number and percent of postal codes represented in each year of the data, and the distribution of the number of students across postal codes in each year. University students are restricted to UGFTEN.

Figure 25: Distribution of university UGFTEN and TVET college students' home postal code in 2019



(a) University UGFTEN students (b) TVET college students

Source: Author's constructed data set.
Notes: This figure visually represents the 2019 data presented in Table 13. On a quintile spectrum of light to dark blue, the lightest postal codes are the home postal codes with the lowest number of students (bottom 25th percentile). The darkest postal codes are the home postal codes with the highest number of students (top 25th percentile).

For the sample of TVET college students, 2,443 unique postal codes (or 86.66% of postal codes) are represented. Therefore, Table 13 shows there is a difference of 275 more postal codes represented by the sample of TVET college students relative to the sample of university students in 2019.

Table 13 provides further information on the distribution of the number of students across postal codes in each year. To complement Table 13, the distribution for the university sample and TVET college sample in 2019 is visually represented in Figure 25a and Figure 25b. There is a wide range in the number of students residing within a single postal code – from a minimum of 1 student to a maximum of between 2,616 to 3,774 university students, or a maximum of 9,402 TVET college students.

Between 2008 and 2019, the number of university students within a postal code at the 10th, 25th, 50th and 75th percentile declined but the 90th percentile, maximum, and mean increased. Furthermore, the deviation from the mean number of university students per postal code is increasing over time given that the standard deviation has increased. Not only is the data skewed to the right but it is getting more skewed as the gap between the mean and the median diverge further. As the share of postal codes represented stays fairly constant, growth in the number of university students entering the system is coming from postal codes that already have a large number of students and are at the upper end of the distribution (for example, postal codes that fall in the 90th percentile).

Table 14 shows the number of unique postal codes that are represented within each university type for undergraduate first-time entering students from 2008 to 2019. Traditional universities have the greatest number of postal codes represented for this sub-group of students amongst all the university types. This is somewhat expected given that traditional universities account for the greatest share of undergraduate, first-time entering students. Furthermore, the number of postal codes represented amongst traditional universities increased over time – from 1,722 postal codes in 2008 to 1,867 postal codes in 2019. Comprehensive universities have also increased their postal codes representation from 1,292 postal codes in 2008 to 1,390 postal codes in 2019. Hence, as the number of students in the sample increased from 2008 to 2019, so too did the representation of postal codes in these university types.

However, while these two trends are likely to be related, an increase (decrease) in the number of students does not necessarily imply an increase (decrease) in the representation of postal codes. Universities of technology and Unisa experienced a decline in the number of postal codes represented by 145 and 499 postal codes, respectively. Yet, universities of technology experienced an increase in the number of undergraduate students entering the institution; whereas Unisa experienced a decline in the number of students. For Unisa, this reduction in postal code representation coincided with a reduction in the number of undergraduate students entering the institution.

All university types experienced an increase in their respective ratio of the number of students to the number of postal codes represented because the number of students are increasing at a greater rate than the change in the number of postal codes represented. Hence, there is a greater concentration of students per postal code overtime.

Table 14: Number and share of postal codes represented by the number of students at university types

	Comprehensive				Technology			
	UGFTEN	Postal codes		Ratio	UGFTEN	Postal codes		Ratio
		n	%			n	%	
2008	21,680	1,292	45.83	16.78	32,631	1,668	59.17	19.56
2009	27,815	1,376	48.81	20.21	36,999	1,685	59.77	21.96
2010	24,562	1,337	47.43	18.37	35,422	1,657	58.78	21.38
2011	26,110	1,361	48.28	19.18	31,044	1,625	57.64	19.10
2012	23,581	1,373	48.71	17.17	32,629	1,630	57.82	20.02
2013	25,865	1,354	48.03	19.10	34,860	1,643	58.28	21.22
2014	28,450	1,423	50.48	19.99	35,788	1,633	57.93	21.92
2015	27,508	1,390	49.31	19.79	34,789	1,596	56.62	21.80
2016	29,532	1,376	48.81	21.46	37,895	1,628	57.75	23.28
2017	26,747	1,332	47.25	20.08	39,713	1,602	56.83	24.79
2018	28,360	1,369	48.56	20.72	38,447	1,583	56.15	24.29
2019	29,819	1,390	49.31	21.45	37,328	1,523	54.03	24.51

	Traditional				Unisa			
	UGFTEN	Postal codes		Ratio	UGFTEN	Postal codes		Ratio
		n	%			n	%	
2008	38,699	1,722	61.09	22.47	39,343	1,695	60.13	23.21
2009	44,098	1,776	63.00	24.83	36,616	1,693	60.06	21.63
2010	45,205	1,798	63.78	25.14	42,228	1,729	61.33	24.42
2011	48,389	1,829	64.88	26.46	55,453	1,792	63.57	30.94
2012	49,510	1,845	65.45	26.83	47,397	1,691	59.99	28.03
2013	52,733	1,858	65.91	28.38	32,580	1,449	51.40	22.48
2014	57,390	1,866	66.19	30.76	33,809	1,418	50.30	23.84
2015	54,748	1,824	64.70	30.02	41,866	1,345	47.71	31.13
2016	60,410	1,872	66.41	32.27	19,215	856	30.37	22.45
2017	61,054	1,872	66.41	32.61	53,106	1,159	41.11	45.82
2018	63,763	1,876	66.55	33.99	65,159	1,247	44.24	52.25
2019	62,448	1,867	66.23	33.45	46,566	1,196	42.43	38.93

Source: Author's constructed data set.
Notes: This table shows the number of university students in the sample for each university type (UGFTEN restriction), along with the number and share of unique postal codes each respective group of students represents. The ratio is the number of UGFTEN students to the number of postal codes.

Table 15 shows the share of unique postal codes represented within each individual university (anonymised), organised by university type. The share of postal codes represented amongst institutions within each university type varies. For example, in 2019, the share of postal codes represented ranged from: 6.46% to 33.42% at comprehensive universities; 10.54% to 33.20% at universities of technology; 8.55% to 40.05% at traditional universities; and 42.43% at Unisa. Hence, some individual universities are attracting students from a wider range of postal codes from South Africa but there is no clear pattern that a particular type of university has a high or low share of postal codes represented across all of its individual universities.

Table 15: Share of postal codes represented by university from 2008 to 2019

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Comprehensive universities												
Comprehensive 1	34.02	35.90	33.13	35.54	33.98	33.10	36.29	33.88	33.38	31.96	32.03	33.42
Comprehensive 2	15.61	18.20	18.34	21.11	20.61	20.61	22.49	21.25	22.56	23.09	25.90	24.94
Comprehensive 3	6.88	7.09	7.13	7.88	7.95	9.54	9.44	9.12	10.00	8.41	9.22	10.04
Comprehensive 4	8.97	10.43	10.36	10.22	10.15	10.64	10.18	13.37	12.45	12.17	12.52	12.77
Comprehensive 5	9.79	9.37	10.96	9.54	10.39	9.86	9.29	9.33	8.02	7.73	8.02	8.83
Comprehensive 6							1.53	2.20	3.76	4.08	5.43	6.46
Comprehensive 7							1.70	2.91	4.61	5.00	6.88	10.00
Universities of technology												
Technology 1	23.38	22.38	22.88	23.52	23.24	24.12	22.84	22.21	22.10	21.32	24.23	25.82
Technology 2	12.17	13.73	12.45	12.13	13.02	15.50	15.36	15.79	17.63	19.44	14.79	15.29
Technology 3	15.29	16.42	17.56	16.11	16.64	17.77	16.46	16.99	15.86	17.81	17.35	17.35
Technology 4	37.67	38.45	36.01	33.52	35.93	37.32	37.85	35.93	36.32	35.72	35.15	33.20
Technology 5	23.24	24.51	25.68	24.16	23.34	21.46	21.43	19.26	23.16	21.18	22.53	20.75
Technology 6	10.11	9.05	9.51	9.76	10.07	10.89	10.50	10.64	11.71	12.03	10.50	10.54
Traditional universities												
Traditional 1	19.23	21.60	19.62	18.16	19.87	18.66	18.73	20.18	20.43	20.75	20.65	20.82
Traditional 2	7.45	8.30	7.88	7.48	8.30	8.16	9.58	9.76	9.68	10.43	11.53	11.14
Traditional 3	21.04	20.65	26.39	25.93	26.46	28.45	29.73	26.46	32.88	31.43	33.06	31.04
Traditional 4	14.19	16.39	16.50	15.93	16.25	16.46	17.35	16.46	17.13	17.99	17.38	16.67
Traditional 5	15.57	15.29	18.34	19.05	18.06	17.52	17.99	13.52	14.93	13.76	14.51	15.57
Traditional 6	22.84	26.18	27.67	30.05	29.27	35.79	30.61	35.30	36.61	39.98	40.19	40.05
Traditional 7	8.80	7.45	5.71	25.97	26.46	29.02	30.90	29.76	28.10	28.31	31.93	32.07
Traditional 8	14.90	16.21	14.76	12.59	13.55	15.15	15.11	16.60	14.08	14.01	14.51	15.36
Traditional 9	19.44	16.85	16.46	16.78	5.11	5.29	6.63	6.63	6.67	6.63	7.41	8.55
Traditional 10	14.33	14.37	16.25	16.18	17.77	16.74	17.56	18.80	20.43	20.08	20.57	22.63
Traditional 11	22.84	27.24	25.12	24.87	26.25	26.71	27.88	26.64	29.83	29.27	28.70	28.31
Traditional 12								11.07	11.42	11.64	11.49	11.39
Unisa												
Unisa 1	60.13	60.06	61.33	63.57	59.99	51.40	50.30	47.71	30.37	41.11	44.24	42.43

Source: Author's constructed data set.
Notes: This table shows there is variation in the share of postal codes represented within each individual university over time, and between individual universities within each university type. The sample of university students is restricted to undergraduate, first-time entering (UGFTEN). University names have been anonymised.

Individual institutions within each university type also display variation in the share of postal codes represented over time, as displayed in Table 15. For example, narrowing in on traditional universities: most institutions increased their postal code representation over time but the Traditional 9 university reduced their postal code representation from 19.44% in 2008 to 8.55% in 2019. The Traditional 1 university and Traditional 8 university held a fairly static representation of postal codes over time. The Traditional 6 university, on the other hand, doubled their postal code representation from 22.84% to 40.05%; and the Traditional 7 university increased their postal code representation 3-fold from 8.80% to 32.07%.

Individual comprehensive universities increased their share of postal codes over time with one exception – the Comprehensive 5 university, which experienced a marginal decline from 9.79% in 2008 to 8.83% in 2019. Amongst the universities of technology, half of the individual universities increased their postal code representation over time (Technology 1 university, Technology 2 university, and Technology 3 university), two individual universities experienced a decline in postal code representation over time (Technology 4 university and Technology 5 university), and one university (Technology 6 university) held a steady share.

In this section, I define my analysis sample and investigate the representation of postal codes in the student-level institutional data. Understanding which postal codes are represented, and how this has changed over time for the sample of university students will drive observed changes in the socio-economic measures discussed in Section 7. I find a higher share of postal codes are represented amongst the sample of TVET colleges students (86.66%) relative to the sample of university students (76.91%) in 2019. The composition of postal codes represented by different student bodies (for example, university students compared to TVET college students; or traditional university students compared to comprehensive university students) directly influences the socio-economic profile assessed. Furthermore, I find as the number of students enrolled at universities grows over time, the bulk of "new" students tend to fall in postal codes that already have a high number of university students. The socio-economic measures of these postal codes will, therefore, have a higher weight over time as there are more students stemming from these postal codes. Thus, influencing the changes observed in the socio-economic profile of university student bodies from 2008 to 2019.

6.3 Caveats and considerations of the data set

There are three key caveats and considerations of the data set I acknowledge. First, I only use socio-economic information from a single point in time (2011 census) so changes in socio-economic circumstances within postal codes are not observed. Second, I cannot observe the socio-economic distribution within each postal code. Therefore, according to the data set, all students from the same postal code are identical in terms of their socio-economic background. Third, the construction of the data set and, in turn, who is kept in the sample of students depends crucially on accurate reporting of postal codes within institutional data sets and the precision of the compiled postal code shapefile by STGroup. I elaborate on these three considerations below.

When comparing the socio-economic profile of student bodies over time, the socio-economic

information of each postal code is held constant because I only use the 2011 census community profiles. Hence, I do not consider to what extent the socio-economic profile of each respective postal code has improved or declined. By holding the postal code socio-economic estimates constant, changes in the student body mean socio-economic indicators over time are caused by: (1) a change in the number of postal code represented, and (2) a change in the number of students residing in a particular type of postal code.

In the data set, all students from the same postal code are assigned the same socio-economic measures. However, in reality there is a socio-economic distribution within each postal code that I cannot observe. Hence, I do not know where a student falls within the postal code's socio-economic distribution. For example, I do not know if the set of post-school students within a particular postal code are bunched at the upper end of the socio-economic distribution or dispersed throughout the socio-economic distribution. Thus, the postal code socio-economic measures – although representative of the mean household in the postal code – may not be representative of post-school students' households in the postal code.

Furthermore, the method used to construct the data implicitly assumes that selection into institutions within a postal code is random but this may not be the case. I cannot observe whether students who attend different universities – but are from the same postal code – look different in terms of their socio-economic status. In other words, I do not know if these students fall at drastically different points in the postal code's socio-economic distribution. Hence, do students from the upper end of the postal code's socio-economic distribution, for example, select into universities rather than TVET colleges? Or, do students from the upper end of the postal code's socio-economic distribution select into traditional universities over universities of technology? This is something I am unable to answer with my data set because students from the same postal code will be equal in terms of socio-economic status.

Lastly, given that connecting the institutional data to the socio-economic data relies on a match between the postal code in these respective data sets, accuracy of postal code information is essential. It is acknowledged that the postal code field in the institutional data is not audited and the accuracy of this information may vary between institutions and/or over time. Furthermore, the postal code boundaries compiled by STGroup are from 2012. Therefore, students from newly developed areas may not be accurately represented.

7 Results and discussion

In this section, I showcase the usefulness of the data set I have constructed. With it, I descriptively analyse the socio-economic profile of student bodies at different institutional types, as well as the distribution of their socio-economic characteristics between universities and colleges, between institutions of a similar type, and within particular institutions – something no other research (that I know of) has explicitly set out to assess. The data set also enables me to compare these socio-economic profile findings to that of the population and location of campuses, as points of reference. In addition, I explore the changes in the socio-economic profile and its distribution over time (from 2008 to 2019), both for the university types and the individual universities making up these types.

To assess the socio-economic profile of various student bodies, I compare the mean statistic and generalised entropy index of a range of individual and household socio-economic indicators for these students' home postal code. The mean statistic captures, on average, the socio-economic profile of student bodies; whereas the generalised entropy index draws on data from the full distribution of the socio-economic indicators. In this way, the generalised entropy index provides an estimation of how homogenous (or heterogeneous) student bodies are based on the socio-economic indicators under review.

The underlying socio-economic indicators I select encompass a broad set of socio-economic information relative to the commonly captured information on students' gender and population group in institutional data sets. I select data on language, education, employment, geographical area, household infrastructure, household asset ownership, and annual household income (in 2011 South African Rands) as indicators from the 2011 census. These social and economic factors are collectively used to describe the type of community environment in which students at different types of institutions find themselves as they are representative of students' home postal code rather than students' actual household.

The student bodies of interest are all TVET college students in 2019 and all undergraduate, first-time entering university students in 2008 to 2019. The undergraduate, first-time entering university students can be grouped by university type (comprehensive universities, universities of technology, traditional universities, and Unisa) and by each of the 26 universities (anonymised). I use these groupings to compare the relative between and within contributions to the generalised entropy index.

In presenting these findings, I reflect on the significance of the mean statistics and generalised entropy indices in relation to the changes that took place in the post-school education system since the end of apartheid.

7.1 Socio-economic profile of student bodies at different institutions

Table 16 presents the 2019 mean socio-economic characteristics of student bodies at TVET colleges and university students from my analysis sample. I benchmark these mean characteristics against the mean socio-economic characteristics of all postal codes in the country, weighted by the estimated number of people living in each postal code. The table also presents the mean annual household income level of postal codes where each respective institutional type is located.

Table 16: Mean socio-economic characteristics of student bodies' home postal code by institutional type in 2019

			Undergraduate, first-time entering university students				
	Population	TVET	All	Comprehensive	Technology	Traditional	Unisa
Number of students	-	552,564	176,161	29,819	37,328	62,448	46,566
Number of postal codes	2,820	2,443	2,168	1,390	1,523	1,867	1,196
Individual characteristics (MIS)[a]							
Gender (%)							
Female	48.12	59.44	60.46	54.92	53.50	59.45	70.95
Population group (%)							
Black	75.40	91.87	80.92	91.30	90.89	64.56	88.24
Coloured	8.87	7.10	5.47	3.68	5.27	8.09	3.27
Indian	2.26	0.23	2.92	1.28	1.61	5.00	2.24
White	12.85	0.03	9.57	3.74	1.97	19.59	5.94
Other	0.62	0.78	1.11	0.00	0.26	2.76	0.30
Individual characteristics (postal code)							
Mean number of individuals	18,341.14						
Language (%)							
Afrikaans	15.86	16.22	15.30	10.44	13.40	20.73	12.65
English	10.10	8.70	12.31	9.75	10.53	14.52	12.43
isiXhosa	14.29	14.64	12.43	26.62	9.73	11.03	7.38
isiZulu	21.28	19.82	22.38	19.30	28.45	17.08	26.56
Education (%)							
No school	7.91	7.86	7.49	8.24	8.03	7.22	6.95
Grade 12	16.61	16.42	17.30	15.34	16.75	17.17	19.16
Post-school qualification (all)	7.98	7.60	9.73	7.88	7.80	11.65	9.88
Certificates	1.97	2.05	2.25	1.89	2.02	2.39	2.46
Diplomas, Bachelors, Postgraduate	6.01	5.55	7.48	5.99	5.77	9.26	7.42
Mean years of education	6.73	6.65	6.92	6.49	6.67	7.10	7.15
Employment (%)							
Employed	45.93	46.16	48.03	42.39	46.14	51.19	48.94
Unemployed	12.63	12.68	11.63	11.99	12.08	10.35	12.73
Discouraged	5.06	5.01	4.72	5.76	5.27	4.17	4.37
Not economically active	36.37	36.15	35.62	39.86	36.51	34.29	33.96
Type of geographical area (%)							
Urban	41.42	39.86	42.39	31.54	38.27	44.41	49.95
Household characteristics (postal code)							
Mean number of households	5,118.84						
Infrastructure (%)							
Electricity (light)	74.60	73.29	74.75	68.87	72.76	76.86	77.30
Piped water	44.99	42.92	45.80	35.80	42.05	51.16	48.04
Asset ownership (%)							
Car	30.81	28.49	32.58	26.79	28.80	37.37	32.89
Cellphone	87.33	86.89	88.11	86.44	87.36	88.42	89.38
Computer	21.55	19.10	22.99	17.92	19.33	27.27	23.45
DVD player	50.97	49.02	50.79	45.39	48.61	53.03	53.00
Internet	9.02	7.38	9.94	7.43	7.98	12.59	9.57
Satellite television	23.80	21.56	24.79	20.60	21.95	27.80	25.71
Radio	66.15	65.23	66.90	63.45	66.30	68.76	67.10
Washing machine	30.63	28.03	31.04	23.88	27.08	36.74	31.17
Annual household income (in 2011 South African Rands)							
Mean income	R111 997.20	R96 271.78	R124 497.30	R99 150.41	R101 729.10	R150 785.60	R123 738.40
Standard deviation	R110 842.80	R76 550.63	R112 099.60	R97 774.75	R86 421.36	R131 230.40	R102 867.50
Mean income of campuses	-	R124 645.10	R181 794.80	R133 674.20	R194 415.40	R196 166.30	R188 047.80
Standard deviation	-	R97 900.16	R112 682.30	R86 727.10	R98 750.54	R137 663.70	R91 206.63

Source: Author's constructed data set.
Notes: This table includes the mean socio-economic characteristics of the South African population, defined as the mean characteristics of all 2,820 postal codes weighted by the number of individuals or households within each postal code. The socio-economic data is representative of 2011 information. The TVET student body sample is not restricted to first-time entering students.
[a] Management Information System (MIS) data is not presented for the population estimates. Instead, the population estimates here reveal the share of individuals that are female and the share of individuals that belong to each respective population group in the mean postal code for the population.
[b] Some institutions have more than one campus and these campuses can fall in different postal codes. Hence, these estimates represent the mean income of postal codes in which the campuses of these institutional types fall.

The socio-economic characteristics in Table 16 are presented in three categories: (1) individual-level information collected in the Management Information System (MIS) (for example, HEMIS 2019 and TVETMIS 2019); (2) socio-economic information based on the share of individuals in students' home postal code; and (3) socio-economic information based on the share of households in students' home postal code. Higher mean estimates typically reveals that students enrolled at that institutional type are from more socio-economically advantaged postal codes; except for the individual-level MIS measures and postal code indicators such as language, no school, and unemployment.

Familiar demographic information on the share of female students and the share of students by population group are first presented for my analysis sample. Here it is shown that, in 2019, approximately 60% of university and TVET college students are female. TVET colleges have a higher share of Black students (91.87%) than the share of Black students at universities (80.92%). Although the aggregate body of university students in my sample looks approximately similar in terms of gender to the TVET college student body, there is almost a 20 percentage point deviation across the university types. The vast majority of the new cohort of students at Unisa identify as female (70.95%), whereas just more than half (53.50%) of the new cohort of students at universities of technology identify as female.

Approximately 90% of undergraduate, first-time entering students enrolled at comprehensive universities, universities of technology and Unisa identify from the Black population group. This aligns with the share of Black students at TVET colleges. However, of this remaining 10%, only 0.03% of TVET college students are from the White population group, whereas the university types have higher shares of White students. Particularly traditional universities, which saw 19.59% of its student body in the sample identify from the White population group in 2019. This is substantially higher than the share of White students at comprehensive universities (3.74%), universities of technology (1.97%) and Unisa (5.94%). Traditional universities also have the largest share of students identify from the Other population group category at 2.76%, whereas the other university types have less than 0.30% of their students identify from the Other population group category.

For the socio-economic postal code indicators presented in Table 16, the estimates suggest university students are more socio-economically advantaged than TVET college students. The 2019 cohort of university students in my analysis sample reside in postal codes where a higher share of individuals hold a post-school qualification, are employed, and live in urban areas. In addition, these postal codes also have a higher share of households that have access to electricity for light, piped water, and asset ownership. There is an approximate R30 000.00 difference (in 2011 South African Rands) between the mean annual household income of the home postal codes of TVET college students (R96 271.78 with a standard deviation of R76 550.63) and university students (R124 497.30 with a standard deviation of R112 099.60). This comparison supports the findings by (Branson and Kahn, 2018), who found mean socio-economic differences between grade 12 students who enrol at a public university versus a TVET college the subsequent year.

Benchmarking these mean socio-economic postal code indicators for university and TVET college student bodies to that of the postal code for the mean person in the population reveals that not only do TVET college students reside in postal codes that have lower socio-economic indicators than that of

university students, but they are also often lower than that for the population. For example, in the home postal codes of TVET college students, on average, a lower share of individuals attain grade 12 as their highest level of education, the mean number of years of education per individual is lower, and there is a lower share of individuals residing in urban areas relative to the population. On the other hand, there is a marginally higher share of individuals with certificates as their highest level of education or employed in in the home postal codes of TVET college students compared to the population.

What is compelling here is the mean socio-economic postal code indicators for students enrolled at comprehensive universities and universities of technology are also lower than that of the population, despite the full sample of university students having higher mean socio-economic estimates. For example, the home postal codes of students enrolled at comprehensive universities and universities of technology have a higher share of individuals with no schooling (8.24% and 8.03% respectively) relative to the population (7.91%). Furthermore, comprehensive universities have lower socio-economic estimates for all of the household characteristics relative to the population and, in most scenarios, relative to TVET college students too. Students at universities of technology also reside in postal codes that are, on average, below the population on almost all household characteristics except for access to a cellphone and access to a radio. This finding is further supported by the differences in mean annual household income in students home postal codes. The home postal codes of students enrolled at TVET colleges (R96 271.78, standard deviation of R76 550.63), comprehensive universities (R99 150.41, standard deviation of R97 774.75), and universities of technology (R101 729.10, standard deviation of R86 421.36) have lower mean annual household income levels than the postal code of the mean person in the population (R111 997.20, standard deviation of R110 842.80).

By comparison, the home postal codes of students enrolled at traditional universities and Unisa have higher shares of individuals employed and residing in urban areas; and a higher share of households with access to electricity, piped water, and asset ownership compared to the population. Mean annual household income levels in the postal codes of these university types are also higher relative to the population. As the majority of university students in my analysis sample stem from traditional universities or Unisa, this is pulling the mean socio-economic estimates of all university students up to suggest that students enrolled at universities reside in postal codes with higher socio-economic characteristics than TVET students.

Of the university types, traditional university student bodies are, on average, from postal codes with the highest share of home language Afrikaans-speakers (20.73%) or English-speakers (14.52%). Within comprehensive universities, isiXhosa is the most spoken home language of students' home postal code (26.62%); and within universities of technology and Unisa, isiZulu is the most spoken home language (28.45% and 26.56%, respectively). Given the low share of individuals speaking Afrikaans or English (which are the predominant mediums of instruction at institutions) in the home postal code of students from comprehensive universities, universities of technology and Unisa, a proportion of students at these institutions are more likely to be learning in a language that is different to that of their home language than at traditional universities, as an example.

Students' level of educational aspiration may be influenced by the educational attainment levels of

their community. Table 16 shows, based on the 2019 cohort, the home postal codes of students at traditional universities have, on average, a higher share of individuals who have attained a post-school qualification (11.65%), relative to students at other university types or students at TVET colleges. These post-school qualifications can be categorised as certificates; and diplomas, bachelors degrees, and postgraduate degrees. This is noticeably higher than the mean share of individuals with a post-school qualification in the home postal codes of students at Unisa (9.88%), comprehensive universities (7.88%), universities of technology (7.80%), and TVET colleges (7.60%). Of the 11.65% of individuals who hold a post-school qualification in the home postal codes of traditional university students, 2.39% hold a certificate and 9.26% hold a diploma, bachelors degree, or postgraduate degree as their highest level of educational attainment. With these higher levels of educational attainment, the home postal codes of traditional university students also see the highest share of individuals employed (51.19%) relative to the home postal codes of students at comprehensive universities (42.39%), universities of technology (46.14%), Unisa (48.94%), or TVET colleges (46.16%).

Turning to the type of geographical area and infrastructure where students stay, basic resources – such as electricity and connectivity to the internet – are more often than not required to learn from "home" (or from a distance to the physical location of a post-school institution). According to Unisa's 2019 students in my analysis sample, 49.95% of individuals in their home postal codes reside in urban areas. These areas are more likely to be developed and, as such, have better access to electricity for light and piped water. 77.30% of households, on average, have access to electricity for light and 48.04% of households, on average, have access to piped water in the home postal codes of Unisa students. All of Unisa's undergraduate, first-time entering students are distance learners, according to the Power Higher Education Data Analyser. Hence, these students do not necessarily have the luxury of using campus facilities to carry out their studies. They are reliant on the accessible resources in their home postal code.

The mean share of households with access to electricity in students' home postal codes is relatively high (on average, between 68.87% and 77.30%) across all institutional types. The household access to electricity in the home postal codes of comprehensive university students fall on the lower end of this spectrum and are also found to have the lowest share of households with piped water (35.80%), on average. Furthermore, household asset ownership is consistently the lowest for all indicators measured – namely ownership of a car, cellphone, computer, DVD player, internet, satellite television, radio and washing machine. In comparison to the students at Unisa – or even students at traditional universities who come from postal codes where a higher share of households have access to basic infrastructure and asset ownership – students at comprehensive universities may be more reliant on campus resources that they do not necessarily have at home.

Socio-economic differences between university types are also evident along income. The mean annual household income of comprehensive university students' home postal code is approximately R50 000.00 lower than that of traditional university students' home postal code. Here, in terms of income, the home postal codes of traditional university students appear as the most socio-economically advantaged (R150 785.50 mean annual household income, standard deviation of R131 230.40) and the home postal

codes of comprehensive university students rank as the least socio-economically advantaged (R99 150.41 mean annual household income, standard deviation of R97 774.74) amongst the university types. Hence, on average, the mean annual income of the home postal codes of comprehensive university students is two-thirds that of the home postal codes of traditional university students. The mean annual income of university of technology students' home postal code is R 101 729.10 (only marginally higher than the home postal code of comprehensive university students and has a standard deviation of R86 421.36) and R123 738.40 (standard deviation of R123 738.40) for Unisa students.

The composition of the home postal codes of TVET college students have a similar share of households with asset ownership as that of the university of technology students, and a higher share of households with asset ownership compared to comprehensive university students. Yet, when it comes to mean annual household income, the home postal codes of TVET college students (R96 271.78, standard deviation R76 550.63) are more closely aligned to the home postal codes of comprehensive university students (R99 150.41, standard deviation of R97 774.75). Moreover, the mean annual income of campus location postal codes for TVET colleges and comprehensive universities are more similar than that of universities of technology, traditional universities and Unisa.

Table 16 shows the mean annual household income in the postal codes of campus locations is higher for all institutional types relative to the mean annual household income of their respective student bodies' home postal codes. While the student bodies at comprehensive universities and universities of technology had similar mean annual household incomes, the campuses of these institutions are, on average, located in postal codes with meaningful differences in mean annual household income. Comprehensive university campuses are situated in postal codes with a mean annual household income of R133 674.20 (standard deviation of R86 727.10); and universities of technology campuses can be found in postal codes where the mean annual household income is R194 415.40 (standard deviation of R98 750.54). Hence, there is a greater than R60 000.00 difference in the mean annual household income in the postal codes of these institutional types' campus locations. On this measure, universities of technology (R194 415.40, standard deviation of R98 750.54) appear more similar to traditional universities (R196 166.30, standard deviation of R137 663.70) and Unisa (R188 047.80, standard deviation of R91 206.63); whereas on the characteristics of their student body, universities of technology appear more socio-economically similar to comprehensive universities and even TVET colleges.

Contextualising these findings of the socio-economic differences between university types from my data set in relation to the institutional characteristics and student characteristics profiled in Section 4, I infer the following three findings. First, most undergraduate, first-time entering university students are enrolled at a traditional university or Unisa (see Figure 8). Hence, a greater volume of university students are enrolled at these institutions where the student body is, on average, characterised as being more socio-economically advantaged. Second, amongst the types of undergraduate qualification programmes, degrees are the most popular at traditional universities. Hence, students enrolling for degree programmes at traditional universities are more likely, on average, to come from postal codes that are more socio-economically advantaged than the other university types. Another example would be that the majority of undergraduate students at universities of technology enrol for diplomas and Table 16 shows that these

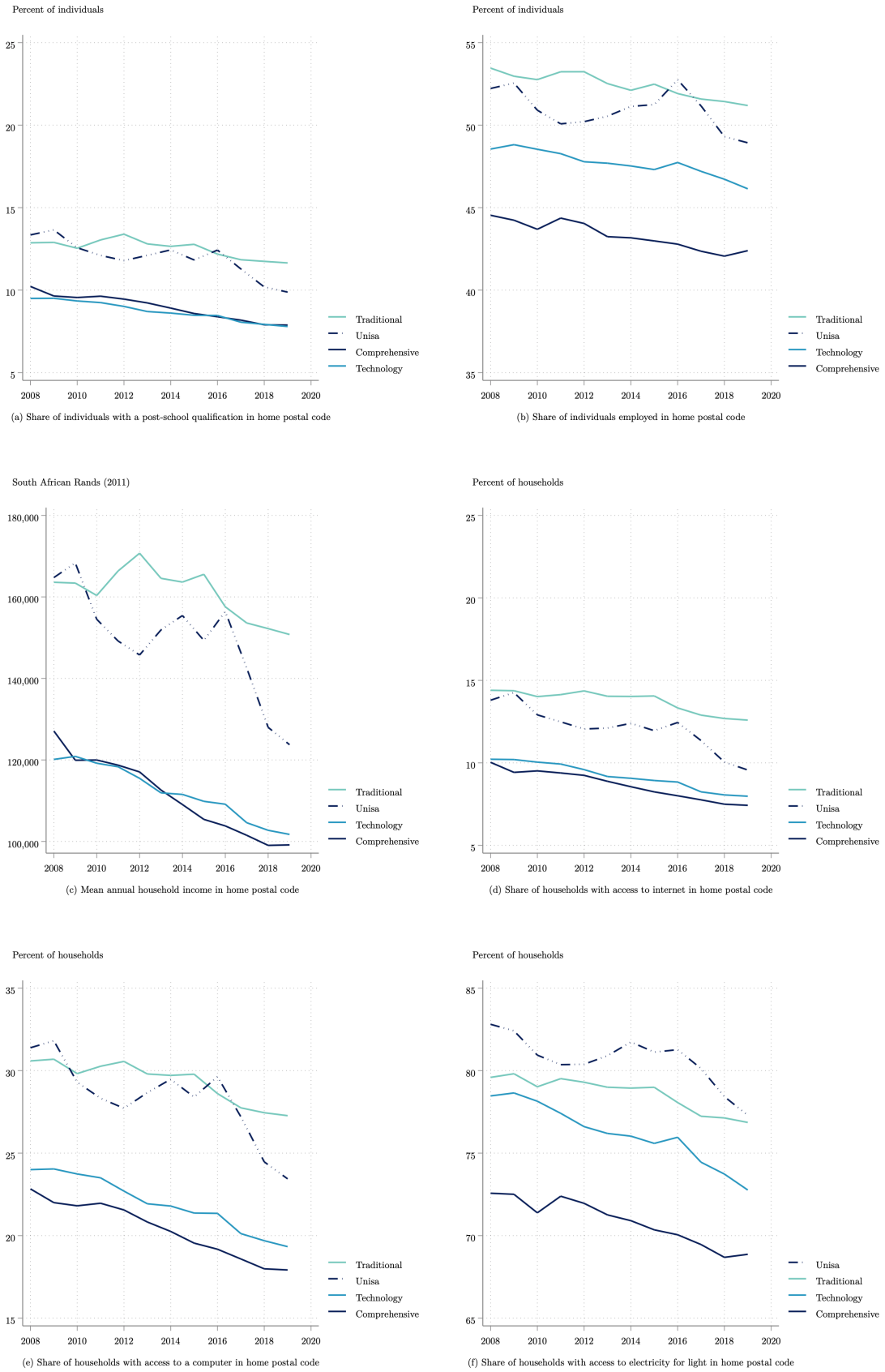
students tend to come from postal codes with lower socio-economic characteristics compared to students at traditional universities or Unisa. Third, the composition of students at universities of technology and comprehensive universities is characterised as having a falling female representation but a strong representation of Black students (approximately, 91% of students in the analysis sample). However, these students are shown to come from home postal codes with lower socio-economic characteristics than that of traditional universities and Unisa where there is a growing representation of female students but a lower share of Black students (64.56% and 88.24%, respectively).

Although the mean socio-economic postal code measures presented in Table 16 do not represent students' actual household, they epitomise the circumstances of the community in which they have grown up. From these mean measures in Table 16 alone, differences in students' socio-economic background between different institutions are evident. While previous research suggests students who enrol at universities are more socio-economically advantaged than students who enrol at TVET colleges (Branson and Kahn, 2018), the findings presented in Table 16 suggest that it is not all university types with enrolees who are more socio-economically advantaged than TVET college enrolees. On some socio-economic characteristics, students enrolled at comprehensive universities are from postal codes with lower socio-economic circumstances (and to some extent students enrolled at universities of technology too) than students enrolled at TVET college students. These findings also convey that TVET colleges, comprehensive universities and universities of technology do not seem to be attracting students from postal codes with higher socio-economic characteristics to the same extent as traditional universities or Unisa.

Cooper (2015) argues, as previously discussed, that some of the socio-economic differences amongst the universities are already inherited based on which institutions did and did not merge. Of the 12 traditional universities, half were historically White institutions and were not involved in the merging of institutions. By comparison, among universities of technology only 2 (out of 6 total institutions) were historically White. There are no historically White institutions that are now classified as a comprehensive university (besides any merged institution). On the other hand, Unisa is the result of the University of South Africa and Technikon South Africa – both of which are historically White institutions.[41] Following Cooper's argument, it is then expected that the mean socio-economic measures of traditional universities and Unisa are likely to be higher than comprehensive universities and universities of technology given historically White institutions where better resourced and located in wealthier, White neighbourhoods during apartheid. This expectation is evident in Figure 26.

[41] See Appendix C2.

Figure 26: Mean socio-economic characteristics of student bodies' home postal code by university type from 2008 to 2019



(a) Share of individuals with a post-school qualification in home postal code

(b) Share of individuals employed in home postal code

(c) Mean annual household income in home postal code

(d) Share of households with access to internet in home postal code

(e) Share of households with access to a computer in home postal code

(f) Share of households with access to electricity for light in home postal code

Source: Author's constructed data set.
Notes: The socio-economic data is representative of 2011 information. The sample is restricted to undergraduate, first-time entering university students.

An additional use of the data set I have constructed is to compare socio-economic differences between institutional types over time. Figure 26 plots six selected socio-economic measures from 2008 to 2019 for student bodies by university type using my constructed data set. The selected socio-economic measures are: the share of individuals with a post-school qualification (Figure 26a), the share of individuals employed (Figure 26b), the mean annual household income (Figure 26c), the share of households with access to the internet (Figure 26d), the share of households with access to a computer (Figure 26e), and the share of households with access to electricity for light (Figure 26f). From these six selected measures, there are two consistent observations. First, the mean socio-economic measures of students at traditional universities and students at Unisa are similar over time, and, likewise, the mean socio-economic measures of students at universities of technology and comprehensive universities are closely plotted over time. Second, the mean socio-economic measures have declined from 2008 to 2019 for all university types.

Even though traditional universities and Unisa have students from postal codes with higher socio-economic mean indictors than comprehensive universities and universities of technology, the gap between means vary across indicators.[42] For indicators such as the share of individuals with a post-school qualification (Figure 26a) and the share of households with access to the internet (Figure 26d), the mean trends for the university types are within range by approximately 5 percentage points. By comparison, the differences along socio-economic means are more stark for the share of individuals employed (Figure 26b), the share of households with access to a computer (Figure 26e), and the share of households with access to electricity for light (Figure 26f) where the mean trends are within range by approximately 10 percentage points. These student bodies in my analysis sample are arguably more similar in terms the share of individuals with a post-school qualification and the share of households with access to the internet in their home postal codes; and more different in terms of the share of individuals employed, the share of households with access to a computer and the share of households with access to electricity for light.

Reasons for the socio-economic trends declining over time are driven by: (1) a change in the number of postal code represented, and (2) a change in the number of students residing in a particular type of postal code. When descriptively analysing the constructed data set in Table 13, it was shown that growth in the number of students is coming from postal codes which already have a large number of students. The socio-economic characteristics of these postal codes with a large number of students (towards the upper end of the distribution in terms of number of students) are lower than the respective 2008 mean value. Hence, this pulls down the overall socio-economic mean of a university type, resulting in the declining trend witnessed from 2008 to 2019 in Figure 26.

Expanding on the differences between institutional types over time, I am further able to assess the differences in particular institutions (within the university types) over time with my constructed data set. In Table 17, I present the mean annual household income (and its standard deviation) for students' home postal code by university for 2008, 2012, 2016 and 2019. This shows the range of mean income values within the university types for each year of data and for which individual universities the mean income value has increased or decreased.

[42] The scale of the y-axes are kept constant for Figure 26 (except for Figure 26c where the y-axis is in South African Rands).

The mean annual household income estimates of comprehensive university students' home postal code ranged from R136 696.00 to R56 189.96 (approximate difference of R80 500.00). This range has declined over time noticeably driven by the Comprehensive 1 university, which is the institution with the highest mean income estimate of all the comprehensive universities. The mean income estimate of its student body was R187 084.90 in 2008 and this estimate reduced to R136 696.00 by 2019. The other comprehensive universities kept a relatively stable trend in the mean income estimates of its student body. This suggests that students are coming from postal codes with fairly similar income levels over time.

Universities of technology have the tightest range – from R92 824.26 to R129 692.50 in 2019 (approximate difference of R37 000.00). However, this range used to be wider and has converged with time. The comprehensive universities with student bodies from postal codes with higher mean annual incomes (particularly the Technology 1 university, Technology 3 university, and Technology 4 university) experienced a decline in their student body's mean annual household income, whereas the comprehensive university with the lowest mean annual household income in 2008 (Technology 6 university) experienced a rise in its student body's postal code mean annual household income levels. Thus, condensing the range of mean income values between comprehensive universities.

Traditional universities, on the other hand, have witnessed a rise in the range of mean income values at its universities over time. In 2008, this range was R69 312.39 to R248 748.30; and in 2019 the expanded range was R74 889.05 to R309 953.70 (approximate difference of R235 000.00.) The group of universities making up traditional universities arguably show the most diversity in terms of income levels. The expansion in the range of income levels is particularly driven by the rise in students' postal code mean annual household income at the Traditional 9 university, which increase by approximately R100 000.00 from 2008 to 2019. Hence, this institution is attracting students from postal codes with higher income levels over time. By comparison, Unisa 1 university is attracting students from postal codes with lower income levels over time.

While Table 16 and Figure 26 show differences in the mean socio-economic characteristics for university types, Table 17 shows how much socio-economic variation is possible within university types. Yet, even though variation in mean income levels is evident within the university types, I would argue that it's skewed. For example: 4 out of the 7 comprehensive universities have mean income levels less than R90 000.00; 5 out of the 6 universities of technology have mean income levels greater than R90 000.00 but less than R150 000.00; and 6 out of the 12 traditional universities have mean income levels greater than R150 000.00 in 2019. There is not a single comprehensive university or university of technology that had a mean income level above R150 000.00 in 2019; although Comprehensive 1 university and Technology 1 university did reach mean income levels above R150 000.00 in 2008 and 2012. It would seem that comprehensive universities and universities of technology (and TVET colleges) are not the first institutional type of choice for students from postal codes with high mean annual household income levels. Hence, there are low, middle and high clusters of income levels between the institutional types.

Table 17: Mean annual household income of student bodies' home postal code by university

	2008	2012	2016	2019
Comprehensive universities				
Comprehensive 1	R187 084.90	R162 247.20	R144 857.70	R136 696.00
	R164 736.10	R145 891.30	R130 230.80	R126 392.20
Comprehensive 2	R125 092.90	R125 359.40	R126 515.00	R117 400.00
	R87 330.95	R90 508.02	R93 954.81	R90 095.82
Comprehensive 3	R59 387.84	R54 149.72	R53 814.97	R56 189.96
	R60 781.74	R58 712.32	R52 510.35	R58 804.00
Comprehensive 4	R64 920.94	R67 737.02	R63 460.92	R64 634.27
	R61 627.34	R65 395.83	R62 666.64	R67 617.58
Comprehensive 5	R73 826.47	R79 439.07	R71 267.00	R76 664.14
	R55 329.27	R68 250.90	R57 598.24	R63 521.65
Comprehensive 6			R109 251.30	R107 038.20
			R52 053.57	R61 978.94
Comprehensive 7			R80 193.54	R87 890.68
			R64 487.04	R74 925.58
Universities of technology				
Technology 1	R160 948.60	R152 180.00	R140 860.70	R129 692.50
	R109 333.00	R107 310.00	R98 374.65	R97 522.99
Technology 2	R91 705.84	R92 401.06	R89 702.00	R81 997.98
	R53 689.81	R56 403.78	R56 734.64	R46 479.23
Technology 3	R122 799.30	R113 774.60	R115 855.20	R102 118.50
	R97 628.59	R96 509.75	R93 883.62	R86 825.17
Technology 4	R117 344.80	R108 665.00	R100 793.40	R98 165.86
	R108 000.80	R100 044.30	R92 515.97	R89 105.99
Technology 5	R98 701.01	R104 862.50	R98 670.97	R95 952.59
	R85 001.53	R92 336.91	R86 473.27	R83 518.20
Technology 6	R85 406.92	R88 298.36	R92 776.16	R92 824.26
	R76 669.22	R81 187.67	R83 051.14	R84 767.23
Traditional universities				
Traditional 1	R248 748.30	R259 045.00	R249 526.40	R232 055.60
	R161 926.20	R167 151.40	R163 683.80	R160 850.80
Traditional 2	R96 101.56	R89 050.76	R85 752.06	R81 178.34
	R81 392.91	R78 820.91	R76 069.47	R76 568.50
Traditional 3	R108 004.10	R161 966.20	R105 684.40	R93 416.31
	R83 815.79	R149 256.00	R89 380.58	R72 451.90
Traditional 4	R144 489.00	R129 921.70	R120 837.20	R109 074.60
	R110 951.90	R107 088.90	R102 547.30	R94 558.07
Traditional 5	R69 312.39	R78 026.30	R73 619.45	R74 889.05
	R69 321.67	R79 674.29	R73 515.02	R75 208.39
Traditional 6	R117 639.40	R108 026.30	R109 080.30	R110 811.40
	R92 692.30	R80 877.19	R85 160.35	R88 705.65
Traditional 7	R219 360.90	R215 990.70	R214 059.80	R205 159.80
	R99 690.52	R143 626.40	R153 158.60	R154 879.50
Traditional 8	R215 177.80	R220 222.10	R192 245.70	R155 174.60
	R173 052.20	R176 766.00	R161 654.00	R135 354.50
Traditional 9	R211 533.60	R316 550.60	R310 373.20	R309 953.70
	R130 545.50	R69 400.20	R78 084.71	R84 287.42
Traditional 10	R156 719.70	R164 930.50	R160 742.30	R159 533.30
	R100 911.00	R107 815.00	R102 276.60	R105 634.30
Traditional 11	R235 384.90	R218 746.30	R196 904.80	R203 073.90
	R184 265.90	R178 875.70	R167 181.30	R172 084.20
Traditional 12			R99 092.52	R93 752.35
			R92 831.29	R83 389.14
Unisa				
Unisa 1	R164 755.70	R145 752.40	R156 424.20	R123 729.80
	R132 078.10	R121 631.10	R112 831.10	R102 867.50

Source: Author's constructed data set.
Notes: The sample is restricted to undergraduate, first-time entering university students. University names have been anonymised.

7.2 Heterogeneity estimates of students' socio-economic backgrounds within and between institutions

The comparison of the mean socio-economic postal code measures between institutional types, offers a sense of whether these students are, on average, from similar socio-economic backgrounds. However, the mean statistics presented in Table 16, Figure 26 and Table 17 fail to recognise whether all students within an institutional type are concentrated around the mean (a homogeneous student body) or widely dispersed between the lower and upper end of the distribution (a heterogeneous student body).

To gauge how heterogeneous (or homogenous) different student bodies are, I use the generalised entropy index. Equation 3 presents the formula for the generalised entropy index of order α for any distribution x. The generalised entropy index ranges between 0 (perfect homogeneity) and infinity – the higher the measure, the higher the extent of heterogeneity because it is further way from a perfect uniform distribution.[43] The parameter, α, represents the weight given to different parts of the distribution (World Bank, 2014). The three most common values of α are: 0, 1, and 2 (Haughton and Khandker, 2009). At $\alpha = 2$, the generalised entropy measure is transfer neutral like the Gini coefficient but for $\alpha < 2$, the transfer sensitivity axiom is still satisfied. For values of $\alpha < 1$, the generalised entropy index is more sensitive to transfers at the lower end of the distribution. For values of $\alpha > 1$, the generalised entropy measures are more sensitive to transfers at the upper end of the distribution (Foster et al., 2013).

$$GE(x,\alpha) = \begin{cases} \frac{1}{N\alpha(\alpha-1)} \sum_{i=1}^{N} [(\frac{x_i}{\mu}^{\alpha} - 1] & ,\alpha \neq 0,1 \\ \frac{1}{N} \sum_{i=1}^{N} \frac{x_i}{\mu} ln \frac{x_i}{\mu} & ,\alpha = 1 \\ \frac{1}{N} \sum_{i=1}^{N} ln \frac{\mu}{x_i} & ,\alpha = 0 \end{cases} \tag{3}$$

where, μ is the mean socio-economic indicator:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{4}$$

A desirable attribute of the class of generalised entropy indices is that it satisfies the sub-group consistency axiom and it is the only inequality measure that satisfies the traditional form of the additive decomposability axiom with no residual term like the Gini coefficient (Shorrocks, 1980). The sub-group consistency axiom is imperative as I want to descriptively estimate the extent of socio-economic heterogeneity between student bodies at universities and TVET colleges, between student bodies at different university types, and within each university. This means I am able to decompose the overall generalised entropy measure into the part that is driven by heterogeneity within each respective student

[43] Under GE(1), the maximum value is ln(N).

body (e.g. within a particular type of institution) and the part that is driven by heterogeneity between different student bodies (e.g. between institutional types). Usually when decomposing generalised entropy measures, the literature suggests heterogeneity within-groups accounts for at least 75% of the overall population heterogeneity, and heterogeneity between-groups accounts for the remaining 25% (Haughton and Khandker, 2009). I choose to adopt $\alpha = 1$ when calculating the generalised entropy indices of the postal code socio-economic indicators to estimate the heterogeneity of university and TVET college student bodies.[44]

Table 18 presents the GE(1) estimates of socio-economic postal code indicators for the 2019 cohort of students from my analysis sample at TVET colleges and universities. The indicators with GE(1) estimates that are closest to zero (signalling homogeneity within each respective institutional type) are the share of individuals in students' home postal code with grade 12 as their highest level of education; and the share of individuals who are employed. For the average number of years of individuals' highest level of educational attainment in students' home postal code, the GE(1) estimate is also close to zero. What this means is that the home postal codes of these students (within each respective institutional type) have a similar share of individuals with grade 12 as their highest level of education, who are employed, or the average number of years of educational attainment is similar across the home postal codes of students within an institution. Likewise, the share of households in students' home postal code with access to electricity for light, a cellphone, or a radio is similar. Hence, these indicators have GE(1) estimates that are close to zero too.

On the other hand, the indicators with relatively high GE(1) estimates (signalling higher levels of heterogeneity) are the share of individuals in students' home postal code who speak Afrikaans, English, isiXhosa, or isiZulu as their home language; the share of individuals with a post-school qualification; the share of individuals residing in urban areas; the share of households with access to a computer; the share of households with access to the internet; the share of households with access to a washing machine; and the mean annual household income of students' home postal code. In terms of these socio-economic indicators, students' home postal codes are more different.

The GE(1) estimates for students at TVET colleges tend to be marginally lower relative to that for the analysis sample of university students, as displayed in Table 18. The exceptions are for the share of individuals in students' home postal code whose home language is isiZulu and the share of individuals in students' home postal code who reside in urban geography areas, which have higher GE(1) estimates for students at TVET colleges. Yet, the GE(1) estimates for the infrastructure indicators (the share of households in students' home postal code with access to electricity for light or piped water) is identical for the TVET college and university students bodies in 2019.[45]

[44] See Appendix F for a discussion of the inequality axioms and theoretical motivation as to why I chose $\alpha = 1$.

[45] For the collective 2019 cohort of TVET college students and all undergraduate, first-time entering university students, the GE(1) estimate for mean annual household income in students' home postal code is 0.273. The relative within contribution is 0.976 and the relative between contribution is 0.024.

Table 18: Generalised Entropy (GE(1)) estimates for the socio-economic indicators of students' home postal code by institutional type in 2019

		Undergraduate, first-time entering university students				
	TVET	All	Comprehensive	Technology	Traditional	Unisa
Individual characteristics						
Language (%)						
Afrikaans	0.76	0.77	0.90	0.84	0.63	0.77
English	0.73	0.71	0.73	0.79	0.66	0.68
isiXhosa	1.16	1.22	0.91	1.28	1.19	1.16
isiZulu	1.00	0.88	1.06	0.78	1.06	0.67
Education (%)						
No school	0.22	0.22	0.19	0.20	0.25	0.23
Grade 12	0.09	0.09	0.13	0.09	0.08	0.06
Post-school qualification (all)	0.28	0.35	0.38	0.32	0.34	0.30
Certificates	0.26	0.25	0.32	0.27	0.25	0.19
Diplomas, Bachelors, Postgraduate	0.34	0.42	0.44	0.39	0.40	0.37
Years	0.02	0.02	0.03	0.02	0.03	0.02
Employment (%)						
Employed	0.08	0.08	0.13	0.09	0.07	0.07
Unemployed	0.15	0.16	0.14	0.15	0.17	0.14
Discouraged	0.27	0.32	0.26	0.29	0.38	0.32
Not economically active	0.07	0.07	0.08	0.06	0.07	0.06
Type of geographical area (%)						
Urban	0.67	0.60	0.84	0.70	0.55	0.47
Household characteristics						
Infrastructure (%)						
Electricity (light)	0.05	0.05	0.06	0.05	0.04	0.04
Piped water	0.23	0.23	0.35	0.24	0.21	0.18
Asset ownership (%)						
Car	0.17	0.19	0.25	0.18	0.18	0.16
Cellphone	0.00	0.00	0.00	0.00	0.00	0.00
Computer	0.28	0.32	0.41	0.32	0.30	0.26
DVD player	0.07	0.07	0.09	0.07	0.07	0.06
Internet	0.43	0.49	0.58	0.48	0.45	0.43
Satellite television	0.19	0.23	0.30	0.23	0.21	0.19
Radio	0.01	0.01	0.02	0.01	0.01	0.01
Washing machine	0.29	0.31	0.43	0.33	0.27	0.25
Annual household income (in 2011 South African Rands)						
Mean income	0.24	0.32	0.35	0.28	0.32	0.28

Source: Author's constructed data set.
Notes: The TVET student body is not restricted to first-time entering students. The GE(1) (also known as Theil-T) estimates range from 0 (perfect homogeneity) to ln(N). The higher the GE(1) estimate, the higher the extent of heterogeneity.

When comparing the mean socio-economic estimates in Table 16, the estimates for TVET colleges are most similar to the comprehensive university estimates. Here, in Table 18, the GE(1) estimates for TVET colleges are most similar to the GE(1) estimates for Unisa (particularly for the household characteristics). Relative to the other university types, the GE(1) estimates for Unisa tend to be the lowest. Even though Unisa had one of the highest mean estimates for the socio-economic indicators, the relatively low GE(1) estimates signal that the undergraduate, first-time entering students at Unisa are more likely to come from home postal codes of a similar socio-economic profile compared to the student bodies at other university types.

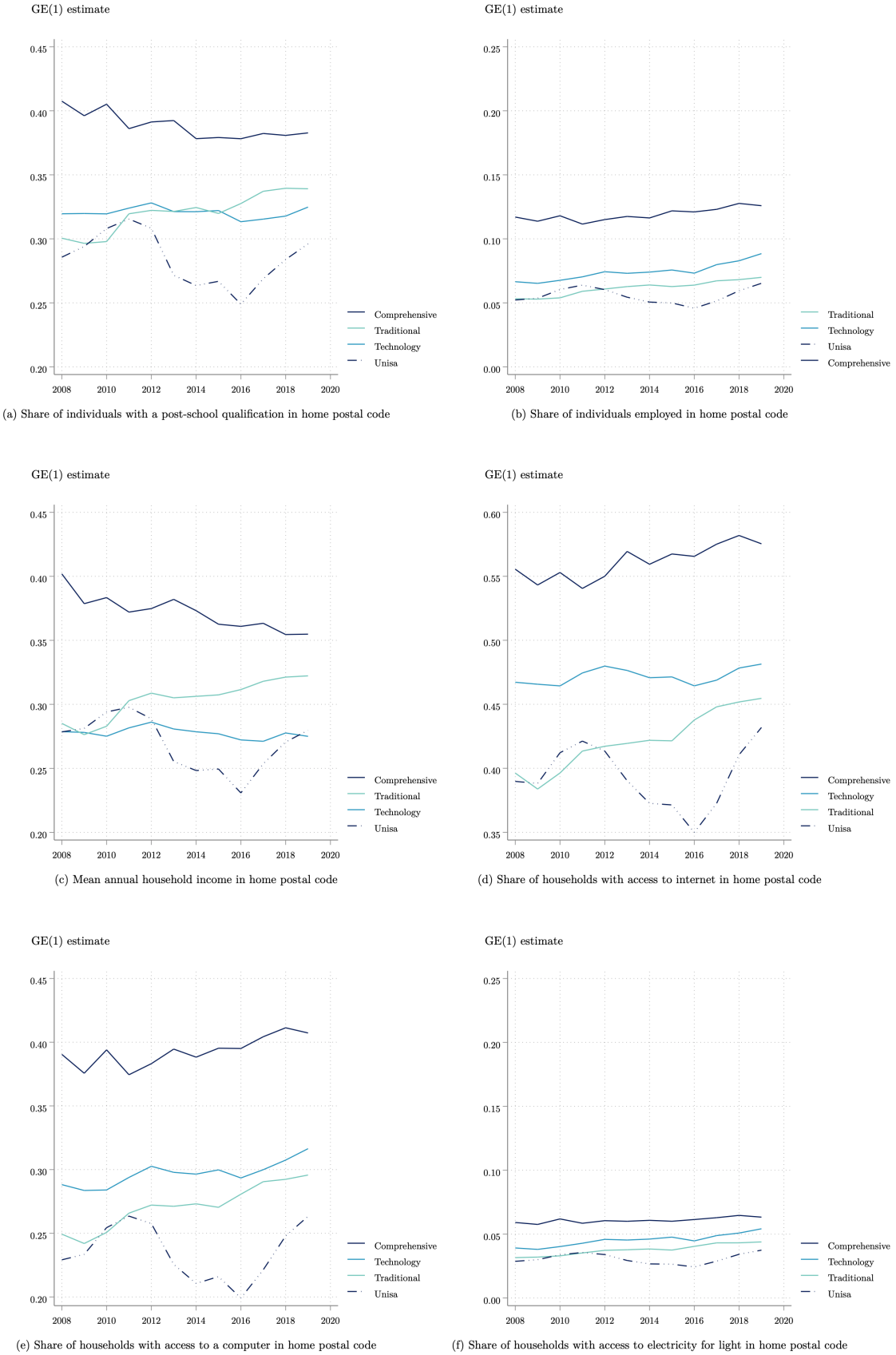
To assess the GE(1) estimate trends between university types over time, Figure 27 plots each university type's GE(1) estimates from 2008 to 2019 for: the share of individuals with a post-school qualification (Figure 27a), the share of individuals employed (Figure 27b), the mean annual household income (Figure 27c), the share of households with access to the internet (Figure 27d), the share of households with access to a computer (Figure 27e), and the share of households with access to electricity for light (Figure 27f).[46] The key findings from these figures are: (1) university types that had the highest mean socio-economic indicators (traditional universities and Unisa) tend to have the lowest GE(1) estimates between 2008 and 2019; and (2) there is no uniform trend in whether heterogeneity is increasing or decreasing over time across the socio-economic indicators presented and across the university types, compared to the clear and consistent downward trend that was shown in the mean socio-economic indicators across all university types.

Comprehensive universities, which had the lowest mean estimates in Table 16, show the highest level of socio-economic heterogeneity amongst their undergraduate, first-time entering student bodies across all the socio-economic characteristics presented in Figure 27. Hence, students at comprehensive universities are more likely to be drawn from postal codes that differ more extensively in their socio-economic characteristics than that of traditional universities, universities of technology and Unisa. On the other hand, Unisa tends to show the lowest level of socio-economic heterogeneity along with having high mean socio-economic estimate relative to the other institutional types. This suggests the analysis sample of students at Unisa are likely to come from home postal codes that are socio-economically similar and found towards the upper end of the socio-economic distribution.

Heterogeneity levels of employment (Figure 27b) and household access to the internet (Figure 27d), computer (Figure 27e) or electricity (Figure 27f) increased across all university types from 2008 to 2019. The extent of heterogeneity amongst traditional universities, universities of technology and Unisa are closer to each other relative to comprehensive universities, which shows much higher levels of heterogeneity. Although, Unisa shows the greatest volatility over time in its GE(1) estimates so it does not always have similar heterogeneity levels to these university types. This is predominantly due to the fluctuation in the number of enrolled undergraduate, first-time entering students and their respective postal code representation. For both the share of individuals with a post-school qualification (Figure 27a) and the mean annual household income in students' home postal code, comprehensive universities show a reduction in the levels of heterogeneity, traditional universities show an increase in

[46] See Table F1, Table F2, Table F3, Table F4, Table F5, and Table F6 for the GE(1) estimates presented in Figure 27.

Figure 27: Generalised Entropy (GE(1)) estimates for the socio-economic indicators of students' home postal code by university type from 2008 to 2019



(a) Share of individuals with a post-school qualification in home postal code

(b) Share of individuals employed in home postal code

(c) Mean annual household income in home postal code

(d) Share of households with access to internet in home postal code

(e) Share of households with access to a computer in home postal code

(f) Share of households with access to electricity for light in home postal code

Source: Author's student level data set with socio-economic information from the 2011 census and own calculations.
Notes: Sample is restricted to first-time entering undergraduate students. The GE(1) (also known as Theil-T) estimates range from 0 (perfect homogeneity) to ln(N). The higher the GE(1) estimate, the higher the extent of heterogeneity. See Appendix F.3 for the underlying values presented in these figures.

the levels of heterogeneity, and universities of technology and Unisa show heterogeneity levels have stayed approximately the same from 2008 to 2019. Therefore, besides the post-school and income indicator for comprehensive universities, the majority of the trends presented in Figure 27 show an increase in heterogeneity.

The relative within contribution, by institutional type, to the population GE(1) estimates (defined as all undergraduate, first-time entering university students) displayed in Figure 27 is exceptionally high (greater than 95% for all socio-economic indicators presented). This leaves the relative between contribution to be very small (less than 5%).[47] In other words, the vast majority of socio-economic heterogeneity amongst the full sample of undergraduate, first-time entering university students is driven by heterogeneity within the institutional types (comprehensive universities, universities of technology, traditional universities and Unisa) and marginally so between the institutional types. This result is driven by the fact that between institutional types there is an overlap in the postal codes represented. For example, in 2019, 76.91% of postal codes were represented amongst the full sample of undergraduate, first-time entering university students across all university types (see Table 13); and the representation of postal codes at the university types were 49.31% at comprehensive universities, 54.03% at universities of technology, 66.23% at traditional universities, and 42.43% at Unisa (see Table 14). Hence, of the 76.91% of postal codes represented by the full sample, 86.11% of these postal codes[48] have at least one student that attended a traditional university in 2019.

Taking this analysis one step further, Table 19 details the GE(1) estimates by each university for the mean annual household income in students' home postal code. Hence, this is a more in-depth analysis of the trends presented in Figure 27c. The institutions are grouped by their institutional type, with the institutional type GE(1) estimates from Figure 27c presented too. With the data grouped in this way, it is apparent that there is no strong pattern in terms of the extent of heterogeneity between individual universities within an institutional type group. For example, in 2019, the extent of heterogeneity of the mean annual household income in students' home postal code ranged from 0.152 to 0.354 between comprehensive universities, 0.124 to 0.302 between universities of technology, and 0.045 to 0.340 between traditional universities.

Universities where the heterogeneity of the mean annual household income in students' home postal code is particularly low, in 2019, include: the Traditional 9 university (0.045), Technology 2 university (0.124), and Comprehensive 6 university (0.152). This suggests that the group of students within each respective institution are from postal codes with similar mean annual household income levels. Comprehensive 4 university (0.354), Traditional 2 university (0.340), Comprehensive 1 university (0.338), Traditional 11 university (0.333), and Comprehensive 3 university (0.327) have relatively high levels of heterogeneity for this indicator suggesting there is a wider variation in the mean annual household income levels amongst these students' home postal code. Notably, these individual universities with the highest levels of income heterogeneity are comprehensive and traditional university types.

The relative between contribution, by university, ranges from 13.7% to 23.1%. This is substantially

[47] See Table F1, Table F2, Table F3, Table F4, Table F5, and Table F6 for the GE(1) estimates presented in Figure 27.

[48] (66.23/76.91) x 100 = 86.11%

higher than the relative between contribution by institutional type (which was less than 5%), and by universities and TVET colleges (2.44% in 2019). These higher relative between contributions suggest that horizontal stratification between particular universities (regardless of their institutional type) is more evident than between university types, or even between universities and TVET colleges. This finding further supports the argument by Cooper (2015) that there is still implicit socio-economic segregation between universities, which is less evident when assessing institutions based on their merged institutional type.

In an explicit sense, higher levels of socio-economic heterogeneity in student bodies are not necessarily a "good" or "bad" thing. Increasing levels of heterogeneity over time signal that student bodies are coming from a more diverse group of postal codes and, in turn, more diverse socio-economic backgrounds.[49] This then suggests that students of a broader socio-economic distribution are participating in post-school education, unlike during the country's apartheid years where student bodies of homogenous characteristics were enforced at institutions. It is hoped that this increased participation, via further success in the labour market, will translate into higher rates of educational and economic mobility. These outcomes will then help individuals escape inequality traps and facilitate reductions in overall levels of poverty. The counter argument here is that high levels of socio-economic heterogeneity may make policy decisions (where students' socio-economic circumstances play a role) more challenging if responses want to take a system-wide approach, as per the vision of DHET's single, national post-school education and training system.

It is encouraged to foster this type of heterogeneity so as to support students of all socio-economic backgrounds to enrol and graduate through post-school education. With this understanding in mind, it comes with the responsibility of recognizing that students have access to different resources and live under different circumstances. If policy responses are not tailored to students of different backgrounds, it runs the risk of students who are more socio-economically disadvantaged dropping out of the system and not completing their journey through post-school education; thus defeating the point of them accessing it in the first place.

More broadly, the socio-economic profile of students on average varies across institutions. In addition to their differing student bodies, institutions still vary in their levels of assets and capabilities. If institutions choose to go against acting collectively as a single-system and instead opt to look out for their own interests, there runs the risk of further diverging the socio-economic gap between institutions as those who are still trying to catch-up may not be able to operate at the same capacity of well-resourced institutions. This would disadvantage the students attending institutions with less resource capabilities and make it more likely for them to fall behind or dropout.

On the other hand, low levels of socio-economic heterogeneity in student bodies suggest that the majority of students are coming from postal codes that are socio-economically similar. This could be a socio-economically advantaged homogenous student body or socio-economically disadvantaged homogenous student body. One can then imagine that it may be an isolating experience for the group of

[49] From a national perspective, the GE(1) estimate of all postal codes in South Africa for mean annual household income is 0.392, which suggests there are high levels of income heterogeneity between postal codes in the country.

students who are in the socio-economic minority within an institution. In some ways this draws parallels to the homogenous institutional culture built under apartheid and the one the country is trying to rebuild in the post-apartheid era.

It is noted that these findings are data dependent. Given the caveats and considerations of the data set, these findings should be interpreted with caution. With this data set, I cannot observe the distribution of socio-economic characteristics within each postal code which would undoubtedly influence the overall findings. It is hoped that the data set will be become more refined with access to complementary data; which will improve the accuracy of what conclusions can – and what cannot – be drawn.

Table 19: Decomposition of the Generalised Entropy (GE(1)) estimates for the mean annual household income in students' home postal code by university from 2008 to 2019

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Comprehensive universities	0.402	0.379	0.383	0.372	0.375	0.382	0.373	0.363	0.361	0.363	0.355	0.355
Comprehensive 1	0.339	0.343	0.336	0.346	0.337	0.330	0.341	0.328	0.325	0.333	0.329	0.338
Comprehensive 2	0.223	0.226	0.220	0.239	0.233	0.229	0.238	0.238	0.242	0.260	0.253	0.253
Comprehensive 3	0.335	0.304	0.311	0.280	0.343	0.306	0.284	0.261	0.293	0.263	0.322	0.327
Comprehensive 4	0.314	0.306	0.318	0.330	0.331	0.321	0.344	0.313	0.325	0.325	0.332	0.354
Comprehensive 5	0.198	0.219	0.214	0.224	0.250	0.251	0.253	0.235	0.208	0.232	0.240	0.225
Comprehensive 6							0.129	0.104	0.107	0.125	0.132	0.152
Comprehensive 7							0.302	0.221	0.241	0.228	0.241	0.268
Universities of technology	0.279	0.278	0.275	0.282	0.286	0.281	0.279	0.277	0.272	0.271	0.278	0.275
Technology 1	0.211	0.210	0.211	0.211	0.226	0.220	0.226	0.225	0.219	0.227	0.249	0.250
Technology 2	0.129	0.138	0.131	0.152	0.141	0.137	0.138	0.147	0.145	0.142	0.129	0.124
Technology 3	0.262	0.267	0.264	0.270	0.285	0.272	0.264	0.268	0.266	0.266	0.279	0.277
Technology 4	0.326	0.328	0.322	0.329	0.321	0.322	0.320	0.318	0.310	0.304	0.307	0.302
Technology 5	0.260	0.256	0.263	0.272	0.281	0.277	0.289	0.274	0.279	0.286	0.275	0.275
Technology 6	0.274	0.269	0.264	0.276	0.293	0.263	0.280	0.296	0.292	0.303	0.295	0.300
Traditional universities	0.285	0.276	0.283	0.303	0.309	0.305	0.306	0.307	0.312	0.318	0.321	0.322
Traditional 1	0.211	0.227	0.216	0.223	0.209	0.203	0.207	0.202	0.212	0.223	0.225	0.232
Traditional 2	0.296	0.303	0.276	0.301	0.309	0.326	0.309	0.296	0.318	0.328	0.326	0.340
Traditional 3	0.221	0.233	0.237	0.343	0.369	0.313	0.280	0.265	0.254	0.220	0.223	0.211
Traditional 4	0.249	0.252	0.259	0.267	0.278	0.267	0.266	0.275	0.290	0.283	0.293	0.289
Traditional 5	0.305	0.331	0.332	0.325	0.331	0.324	0.323	0.333	0.312	0.301	0.297	0.320
Traditional 6	0.229	0.226	0.212	0.223	0.206	0.220	0.222	0.224	0.228	0.244	0.237	0.239
Traditional 7	0.089	0.081	0.067	0.204	0.212	0.230	0.241	0.245	0.240	0.247	0.265	0.264
Traditional 8	0.300	0.317	0.301	0.293	0.300	0.321	0.317	0.321	0.316	0.319	0.323	0.319
Traditional 9	0.173	0.162	0.123	0.104	0.029	0.034	0.035	0.039	0.038	0.035	0.036	0.045
Traditional 10	0.192	0.192	0.198	0.195	0.199	0.187	0.190	0.208	0.191	0.203	0.208	0.206
Traditional 11	0.299	0.326	0.316	0.323	0.319	0.323	0.328	0.327	0.331	0.332	0.339	0.333
Traditional 12								0.343	0.324	0.324	0.328	0.298
Unisa												
Unisa 1	0.278	0.281	0.294	0.298	0.289	0.255	0.248	0.250	0.231	0.254	0.271	0.280
All universities	**0.308**	**0.306**	**0.308**	**0.317**	**0.320**	**0.314**	**0.313**	**0.310**	**0.317**	**0.309**	**0.315**	**0.322**
Relative contribution (out of 1.00)												
Within public universities	0.838	0.851	0.842	0.863	0.832	0.793	0.794	0.794	0.769	0.803	0.816	0.804
Between public universities	0.162	0.149	0.158	0.137	0.168	0.207	0.206	0.206	0.231	0.197	0.184	0.196

Source: Author's constructed data set.
Notes: Sample is restricted to undergraduate, first-time entering students. The GE(1) (also known as Theil-T) estimates range from 0 (perfect homogeneity) to ln(N). The higher the GE(1) estimate, the higher the extent of heterogeneity. University names have been anonymised.

8 Concluding thoughts and reflections

8.1 Summary

As South Africa transitioned to democracy, racially homogenous student bodies could no longer be enforced, meaning students now had access to institutions they were previously excluded from. However, in the late 1990s, overall enrolments at higher education institutions fell below expectations and the gap between enrolments at historically Black and historically White institutions widened. The FET/TVET college sector was also deemed fragmented and inefficient given the high failure and repetition rates experienced by its programmes. Both institutional types also faced financial and human resource constraints. Although the system now offered greater student institution choice (as students were no longer restricted on where they could enrol), it lacked the necessary systemic changes to operate as a nationally coherent system.

This prompted policy discussions, as well as policy development, in a series of draft papers and acts for select institutions to merge within the higher education sector, and within the FET/TVET college sector. This process consolidated the number of institutions and gave institutions a fresh opportunity to build new institutional identities and cultures – ones that were, in theory, no longer defined by population group (Black or White) or language (English- or Afrikaans-speaking). The reformed post-school system also brought with it a new way of categorising institutions based on the orientation of learning. It is known that the university institutional types (comprehensive, technology, traditional, and Unisa) and TVET colleges differ in their characteristics, offering different types of qualifications and fields of specialisation. Although the reformed post-school system is differentiated by design to attract students of all education levels, it is not meant to be segregated along socio-economic lines.

Prior to this work, the gradient of student body socio-economic differences between universities and colleges, between institutions of a similar type, and within particular institutions was not well documented. Part of the reason for this is because potential data sets that could be used to answer this question fall short on dimensions needed to fully explore the extent of socio-economic differences amongst student bodies by institutional types. For example, I show to what extent census data (10% sample) alone can be used to reveal socio-economic differences between students enrolled at universities versus FET/TVET colleges. However, this data set lacks information on which specific institutions students attend, it is only for a single time period of enrolled students, and it is not clear whether the sample of enrolled students is nationally representative. Alternatively, prior research that works around this dilemma map students in institutional data sets to the GPS coordinates of their secondary school Van Broekhuizen et al. (2016) or home address Kerr et al. (2016) to proxy for students' socio-economic characteristics, and Branson and Kahn (2018) used longitudinal data to assess the socio-economic characteristics of grade 12 students prior to their decision to enrol at a post-school institution. However, these researchers were not estimating students' socio-economic information for the purpose of directly examining the distribution of student socio-economic characteristics across institutional types or across time for different student cohorts.

Guided by their methods, and acknowledging that institutional data holds the best student-level data

and census data captures nationally representative socio-economic information, I adopt a complementary method to Van Broekhuizen et al. (2016) and Kerr et al. (2016). This approach involves calculating SAL socio-economic indicators (from the 2011 census community profile data) and weighting the information by the area share of each postal code that it falls within. This leaves me with a postal code level data set with each postal code's socio-economic estimates. I then merge this data with the institutional data using students' home postal code. The result is a student-level data set that holds detailed institutional information and a broad set of socio-economic measures on each student's home postal code.

The value of this data set is that it holds sufficient information to profile the socio-economic circumstances of student bodies at different institutional types. I can compare the socio-economic profile of student bodies at universities and TVET colleges, at institutions of a similar type (for example, traditional and comprehensive universities), and within individual institutions. As I do not assess to what extent these socio-economic measures have improved or declined, a further use of the data set I construct allows the researcher to observe trends in student body socio-economic circumstances over time, relative to a fixed set of socio-economic data. I am also able to draw comparisons between the characteristics of student bodies and the South African population, or even between the characteristics of campus locations.

With this data set, I show where students reside, the socio-economic profile of student bodies by institutional type, and whether institutions' student bodies have become more heterogenous in terms of their socio-economic profile since 2008. My sample of interest is students with valid postal codes. Of these students, I keep the full sample of FET/TVET college students and further restrict the sample of university students to undergraduate, first-time entering. I am also interested in a time period analysis (from 2008 to 2019) for the sample of university students.

Where students reside has not shifted significantly over time. The share of South African postal codes represented amongst the analysis sample of university students has remained fairly constant — approximately 76%. A higher share of postal codes are represented amongst the sample of FET/TVET college students — approximately 87%. As the number of university students enrolled expands over time, I find that the growth in the number of students per postal code is coming from postal codes that already have a high number of students, rather than introducing new postal codes. There are also differences in postal code representation between university types. For example, in 2019, the share of postal codes represented was 49% at comprehensive universities, 54% at universities of technology, 66% at traditional universities, and 42% at Unisa. It is important to note the composition of postal codes represented amongst student bodies at different institutional types (and at specific institutions) because the changes in this composition ultimately influence the changes in the socio-economic profile of university institutions assessed over time.

The socio-economic profile of universities, based on the mean estimates of the home postal codes of their student bodies, is more advantageous than the socio-economic profile of FET/TVET colleges. However, by comparing the socio-economic profile of different university types it becomes apparent that not all of them have student bodies that are, on average, from a higher socio-economic background relative to FET/TVET college student bodies. Student bodies at comprehensive universities and universities of

technology appear more socio-economically similar to that of student bodies at FET/TVET colleges. On some characteristics (such as the share of households with access to electricity, piped water or a computer) comprehensive universities even show lower mean estimates relative to FET/TVET colleges. I also find the socio-economic characteristics of students at FET/TVET colleges, comprehensive universities and (to some extent) universities of technology tend to be lower than the socio-economic characteristics of the population. The other two university types, traditional universities and Unisa, attract students from postal codes with higher socio-economic characteristics. These university types also show higher socio-economic trends over time relative to comprehensive universities and universities of technology.

A further finding on the socio-economic profile of university types was that the mean socio-economic trend declined from 2008 to 2019 across all university types. This suggests that as enrolment increased over time, a greater share of students were coming from postal codes with lower socio-economic characteristics, thus altering the mean socio-economic profile at these institutions. However, the socio-economic profile ranking between institutional types stayed approximately the same. In other words, traditional universities and Unisa maintained a higher socio-economic profile than comprehensive universities and universities of technology, despite all the university types following a downward trend over time.

Yet within these university types, there is further variation in mean socio-economic estimates for individual universities. I provide an overview in the differences in students' postal code mean annual household income by university for 2008, 2012, 2016 and 2019. The range in mean income values between universities in 2019 is greatest for traditional universities (an approximate range of R235 000.00); and tightest for comprehensive universities (an approximate range of R37 000.00). Despite the variation amongst universities within university types, the majority of mean income levels for comprehensive universities are below R90 000.00, and above R150 000.00 for traditional universities, with the majority of universities of technology sitting in the middle of these ranges. Hence, there are arguably low, middle and high clusters of income levels between the institutional types.

There was no clear trend identified as to whether heterogeneity is increasing or decreasing over time across the socio-economic indicators presented and across the university types, compared to the clear and consistent downward trend that was shown in the mean socio-economic indicators across all university types. However, the university types which showed the highest mean socio-economic indicators (traditional universities and Unisa) tended to have the lowest GE(1) estimates between 2008 and 2019. For example, comprehensive universities displayed the lowest mean socio-economic estimate but the highest GE(1) estimates — thus contributing to increased levels of heterogeneity. While the mean socio-economic estimates for FET/TVET colleges were most similar to the comprehensive university estimates, here the GE(1) estimates for TVET colleges were most similar to the GE(1) estimates for Unisa (particularly for the household characteristics).

For the sample of university students, I show the relative between contribution for the GE(1) estimate for annual household income by university ranges from 13.7% to 23.1% over the period analysed. This is greater than the relative between contribution by university type (which is less than 5%) or the relative between contribution between universities and FET/TVET college (which is less than 3%).

These findings suggest there is more evidence for horizontal stratification between particular universities (regardless of their institutional type) in the reformed post-school era rather than between university types, or between universities and TVET colleges.

8.2 Implications for policy

Organisations, such as Universities South Africa (USAf) and Siyaphumelela, have called attention to the need to better understand the country's post-school students in order to put student success at the centre of policy discussions and decisions. By better understanding students, the environment can be designed to match their needs (Universities South Africa (USAf), 2018). Yet, when it comes to policy planning and responses, the DHET faces this challenge between moving forward as a coherent system versus institutions carrying out their own individual responses. For example, at the start of the Covid-19 pandemic there was tension as to when and how institutions should proceed with learning.[50] While some institutions had the capacity and infrastructure to rollout online learning platforms, others did not. It was a contentious debate as to whether institutions who could swiftly transition to online learning should at the expense of other institutions playing catch-up; or whether to delay the return of academic learning but proceed as a collective group of institutions in a more equitable fashion. As this research shows, students stem from a diverse set of socio-economic backgrounds both between and within institutional types. While in the long term this hopefully improves educational and economic mobility, but the prevalent high levels of inequality that are still imbedded in the country today make it challenging to equitably respond to situations in the short term as a system.

However, it is hoped that in initiating this in-depth descriptive analysis of the heterogeneity in South Africa's post-school education system, it can be used to guide further research and responses to planning in post-school education in areas such as student application decisions, admissions policies, affirmative action, financial aid structures, tuition fees and the distribution of government funding across institutions. In this way, the system can be designed to help students succeed by ensuring equitable access to all institutions and improving student completion rates, regardless of their socio-economic circumstances. For example, if an institution decides to transition to remote learning during a pandemic but its student body is from a diverse range of backgrounds (some with low access to a computer or internet services), then it helps to be able to identify who these students are. From there, the institution can assist with providing resources or develop an alternative strategy of learning for these students.

It is particularly students from lower socio-economic backgrounds that need to succeed in the post-school system (in other words, graduate) for post-school education to be an effective policy tool in tackling the country's broader challenges — such as alleviating high levels of poverty, inequality and unemployment. If a disproportionate share of students from higher socio-economic backgrounds are graduating from post-school institutions and entering the labour market, then post-school education is to some extent perpetuating the socio-economic class standings in the country, which is the opposite of policy intensions.

[50] See South African Union of Students' brief on recommendations in response to the Covid-19 pandemic, and Public Universities with a Public Conscience: A Proposed Plan for a Social Pedagogy Alternative in the Time of Pandemic.

To assist with funding constraints for low-income students, the National Student Financial Aid Scheme announced in 2018 that they would fund all eligible post-school students whose household income was R350 000 or less. In addition, there exists a group of students classified as the "missing middle" – students' whose household income falls above the R350 000 threshold but it is not enough to cover the cost of tuition. At the time of compiling this study, the DHET was undergoing the development of a new bursary system; one that aimed to account for the measurements of annual household income and the subsequent qualification (or not) of students to be bursary recipients. I did not have access to valid NSFAS data while conducting this analysis. Hence, it was not possible to interrogate to what extent these new bursary schemes changed the socio-economic composition of students at post-school institutions. I leave this analysis for future research when data availability allows for its undertaking.

Tracking the socio-economic profile of enrolees (and graduates) from the post-school system over time, and whether and where they are entering the labour market, allows policy makers to quantify the profile of new labour market participants. The higher the share of students from lower socio-economic backgrounds and bursary recipients succeeding, the more effective the country's investment in post-school education, from the perspective of aiding individuals to break inequality cycles.

8.3 Implications for future research

This research has offered two main contributions to future research. The first is a postal code level data set with a broad range of socio-economic measures. The second is a well documented comparison of the socio-economic profile of post-school student bodies by institutional type.

The construction of a postal code level data set with its respective socio-economic indicators (due to access to a postal code shapefile) holds value because it can be re-used multiple times. Each year a new institutional MIS data set is published, this postal code data set can be merged with it to continue to observe the changes in the composition of student bodies. Alternatively, rather than using this data set to assess the socio-economic distribution of students, these socio-economic measures of students home postal codes can be used in various regression models and analyses in supplementary research projects. The data set also holds potential to be used in collaborative partnerships for research outside of post-school education as it can be merged with other administrative data sets that hold postal code information.

Furthermore, the methodology I adopt to construct this data set — linking information from the census to student-level data — opens the door to draw on a wider range of indicators in the 2011 census which have not been incorporated in the postal code level data set or explored to their full extent in this paper. The construction of a single socio-economic index, or multiple indices for various socio-economic attributes (such as an asset ownership index, household infrastructure index, or wealth index) could prove useful in future research too.

How well these indicators resemble students' true socio-economic values is rooted in the weighting of the geospatial areas with the census characteristics. Hence, ongoing research continues to refine the data set via findings from other data sources to better understand the variance of socio-economic characteristics within a postal code and where post-school households fall in this distribution. Furthermore, an updated version of the census is being conducted (2022), meaning that the

socio-economic indicators can be updated to reflect more current socio-economic measures once the data is made available.

The findings in this paper are the first of its kind and serve as an initial step to understanding the socio-economic profile of post-school student bodies by institutional type. It prompts further research questions and discussions on the socio-economic profile of post-school students such as: Why are students from postal codes with higher socio-economic backgrounds less likely to enrol at a comprehensive university, university of technology or FET/TVET college? To what extent does cost, distance to the institution or academic eligibly play a role in this? What does the socio-economic profile of student bodies at other post-school institutions look like, such as those in community colleges or private institutions?

Another way of using these data sets could be to assess the share of students from postal codes with lower socio-economic circumstances and the types of qualifications they enrol for, as well as the share that go on to graduate. Thus, to what extent is post-school education helping individuals from lower socio-economic backgrounds improve their quality of life? Does investment in post-school education move the needle on issues such as unemployment, income inequality, and educational and economic mobility in the country as much as policy documents, such as the National Development Plan (NDP), believe it has the potential to? If not, why not? Are there other mechanisms that can be put in place to improve these outcomes? These questions on understanding the role post-school education plays in helping South Africa over come some its biggest socio-economic challenges are left for future research.

References

Black, S. E. and Devereux, P. J. (2011). Recent Developments in Intergenerational Mobility. volume 4 of *Handbook of Labor Economics*, pages 1487 – 1541. Elsevier.

Branson, N. (2018). An Analysis of Out of School Youth Who Have Not Completed Matric: What Can Available Data Tell Us? Working Paper 232, Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town (UCT).

Branson, N., Culligan, S., and Ingle, K. (2020). Developing Siyaphambili: A Stronger South African Nation Website. Moving Towards a Unified Goal to Combat Inequality and Unemployment.

Branson, N. and Kahn, A. (2018). The Post Matriculation Enrolment Decision: Do Public TVET Colleges Provide Students with a Viable Alternative? In Rogan, M., editor, *Post-school Education and the Labour Market in South Africa*.

Brophy, T. S., Daniels, R., and Musundwa, S. (2015). gpsbound: A Command for Importing and Verifying Geographical Information from a User-Provided Shapefile. *The Stata Journal*, 15(2):523–536.

Bunting, I. (2006a). Students. In *Transformation in Higher Education*, pages 147–170. Springer.

Bunting, I. (2006b). The Higher Education Landscape Under Apartheid. In *Transformation in Higher Education*, pages 35–52. Springer.

Chapman, D. W. (1981). A Model of Student College Choice. *The Journal of Higher Education*, 52(5):490–505.

Chisholm, L. (1992). *South African Technical Colleges: Policy Options*. Education Policy Unit, University of the Witwatersrand.

Cloete, N. (2006). New South African Realities. In *Transformation in Higher Education*, pages 414–446. Springer.

Cloete, N. and Bunting, I. (2000). *Higher Education Transformation: Assessing Performance in South Africa*, volume 3. Pretoria: CHET.

Cooper, D. (2015). Social Justice and South African University Student Enrolment Data by 'Race', 1998–2012: From 'Skewed Revolution'to 'Stalled Revolution'. *Higher Education Quarterly*, 69(3):237–262.

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, 27(3):79–102.

Council on Higher Education (2000). Towards a New Higher Education Landscape. Meeting the Equity, Quality, and Social Development Imperatives of South Africa in the 21st Century.

Cowell, F. A. (1985). Measures of Distributional Change: An Axiomatic Approach. *The Review of Economic Studies*, 52(1):135–151.

Department of Education (1996). Green Paper on Higher Education Transformation.

Department of Education (1997). Education White Paper 3: A Programme for the Transformation of Higher Education.

Department of Education (1998). Green Paper on Further Education and Training: Preparing for the 21st Century through Education, Training, and Work.

Department of Education (2001). National Plan for Higher Education.

Department of Education (2003). Education Statistics in South Africa at a Glance in 2001.

Department of Education (2005a). Education Statistics in South Africa at a Glance in 2003.

Department of Education (2005b). Education Statistics in South Africa at a Glance in 2004.

Department of Education (2008). National Plan for Further Education and Trianing Colleges in South Africa.

Department of Education (2020). Revised Strategic Plan 2020/2025.

Department of Higher Education and Training (2013). White Paper for Post-school Education and Training: Building an Expanded, Effective and Integrated Post-school System.

Department of Higher Education and Training (2021). Statitics on Post-School Education and Training in South Africa: 2019.

Essop, A. (2020). The Changing Size and Shape of the Higher Education System in South Africa, 2005–2017. *Johannesburg: University of Johannesburg*.

Foster, J., Seth, S., and Lokshin, M. (2013). A Unified Approach to Measuring Poverty and Inequality.

Gibbon, T. (2004). Creating Comprehensive Universities in South Africa: A Concept Document.

Gibbon, T. and Kabaki, J. (2006). Staff. In *Transformation in Higher Education*, pages 123–152. Springer.

Haughton, J. and Khandker, S. (2009). *Handbook on Poverty and Inequality*. World Bank Publications.

Hossler, D., Schmit, J., and Vesper, N. (1999). *Going to College: How Social, Economic, and Educational Factors Influence the Decisions Students Make*. JHU Press.

Kerr, A., Piraino, P., and Ranchhod, V. (2016). Estimating the Size and Impact of Affirmative Action at the University of Cape Town. Working Paper 172, Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town (UCT).

Khalfani, A. K. and Zuberi, T. (2001). Racial Classification and the Modern Census in South Africa, 1911–1996. *Race and Society*, 4(2):161–176.

Lam, D., Finn, A., and Leibbrandt, M. (2015). Schooling Inequality, Returns to Schooling, and Earnings Inequality: Evidence from Brazil and South Africa. Technical report, WIDER Working Paper.

Lapping, B. (1987). *Apartheid: A History*. Paladin.

Leibbrandt, M., Ranchhod, V., and Green, P. (2018). Taking Stock of South African Income Inequality.

Lombaard, M. (2009). South African Postcode Geography.

Lucas, S. R. (2001). Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects. *American Journal of Sociology*, 106(6):1642–1690.

Lustig, N. (2020). The "Missing Rich" in Household Surveys: Causes and Correction Approaches. Technical report, ECINEQ, Society for the Study of Economic Inequality.

Maassen, P. and Cloete, N. (2006). Global Reform Trends in Higher Education. In *Transformation in Higher Education*, pages 13–57. Springer.

National Commission on Higher Education (1996). A Framework for Transformation.

National Planning Commission of South Africa (2012). National Development Plan 2030: Our Future-make it work.

Pellicer, M. and Ranchhod, V. (2012). Inequality Traps and Human Capital Accumulation in South Africa. SALDRU Working Papers 86, Southern Africa Labour and Development Research Unit,

University of Cape Town.

Republic of South Africa (1996). South African Schools Act No. 84 of 1995, Government Gazette 377(17579).

Republic of South Africa (1997). Higher Education Act No. 101 of 1997, Government Gazette 390(19281).

Republic of South Africa (1998a). Education White Paper 4: A Programme for the Transformation of Further Education and Training, Government Gazette 399(19281).

Republic of South Africa (1998b). Further Education and Training Act No. 98 of 1998, Government Gazette 410(19421).

Republic of South Africa (2002). Transformation and restructuring: A new institutional landscape for higher education. Government Gazette, 444(23549).

Republic of South Africa (2006). Further Education and Training Act No. 16 of 2006, Government Gazette 497(29469.

Shifa, M. and Ranchhod, V. (2019). Handbook on Inequality Measurement for Country Studies. Technical report, African Centre of Excellence for Inequality Research.

Shorrocks, A. F. (1980). The Class of Additively Decomposable Inequality Measures. *Econometrica: Journal of the Econometric Society*, pages 613–625.

Simkins, C. (2004). What Happened to the Distribution of Income in South Africa Between 1995 and 2001.

Solon, G. (1999). Intergenerational Mobility in the Labor Market. In *Handbook of labor economics*, volume 3, pages 1761–1800. Elsevier.

Statistics South Africa (2020). Mid-year population estimates 2019.

Torche, F. (2011). Is a College Degree Still the Great Equalizer? Intergenerational Mobility Across Levels of Schooling in the United States. *American Journal of Sociology*, 117(3):763–807.

Universities South Africa (USAf) (2018). What We Are Learning About Our Students.

Universities South Africa (USAf) (2020). FAQs - What is a matriculation exemption?

Van Broekhuizen, H. and Spaull, N. The "Martha Effect": The Compounding Female Advantage in South African Higher Education. Working Paper 14. Stellenbosch: Stellenbosch University.

Van Broekhuizen, H., Van der Berg, S., and Hofmeyr, H. (2016). Higher Education Access and Outcomes for the 2008 National Matric Cohort.

van der Berg, S., Gustafsson, M., and Malindi, K. (2020). Education and Skills for the Economy and Links to Labour Markets in South Africa. Technical report, Resep, Stellenbosch University.

Vrontis, D., Thrassou, A., and Melanthiou, Y. (2007). A Contemporary Higher Education Student-Choice Model for Developed Countries. *Journal of Business Research*, 60(9):979–989. Marketing of Higher Education, 1st Symposium of the Academy of Marketing Special Interest Group, Cyprus 2006.

Wildschut, A., Mncwango, B., Rogan, M., Rust, J., Fongwa, S., and Meiring, L. (2018). The National Student Financial Aid Scheme (NSFAS) and its Impact: Exploring the Absorption into Employment of NSFAS-funded Graduates.

World Bank (2014). Introduction to Poverty Analysis. Chapter 6: Inequality Measures. Technical report, Washington, DC: World Bank Group.

Data

Department of Basic Education. (2009-2019). Technical Report - National Senior Certificate Examination Results, 2009-2019.

Department of Education. (1999-2009). Education Statistics in South Africa at a Glance: Higher Education Management Information System, 2000-2009.

Department of Higher Education and Training. (2010-2019). Post-school Education and Training Statistics reports: Higher Education Management Information System, 2010-2019; Technical and Vocational Education and Training Management Information System, 2016-2019; TVET College Annual Survey, 2014-2015; Annual Survey of Public FET Colleges, 2013.

Department of Higher Education and Training. Higher Education Management Information System, 2008 to 2019 [data set]. Pretoria: Department of Higher Education and Training [producer]. [privately distributed].

Department of Higher Education and Training. Technical and Vocational Education and Training Management Information System, 2019 [data set]. Pretoria: Department of Higher Education and Training [producer]. [privately distributed].

Power Higher Education Data Analyser (PowerHEDA). (2000-2019). Higher Education Management Information System.

Statistics South Africa. Quarterly Labour Force Survey 2020: Q1 [data set]. Version 1. Pretoria: Statistics South Africa [producer], 2020. Cape Town: DataFirst [distributor], 2020.
DOI: https://doi.org/10.25828/vkhb-2j69.

Statistics South Africa. South African Census Community Profiles 2011 [data set]. Version 1.Pretoria: Statistics SA [producer], 2014.Cape Town: DataFirst [distributor], 2015.
DOI: https://doi.org/10.25828/6n0m-7m52.

Statistics South Africa. South African Census 2011, 10% sample [data set]. Version 2. Pretoria: Statistics South Africa [producer], 2015. Cape Town: DataFirst [distributor], 2015.
DOI: https://doi.org/10.25828/vjy1-tz66.

STGroup Spatial Asset Management. South African postal code shapefile [data set]. Johannesburg: STGroup Spatial Asset Management [producer], 2012. Johannesburg: STGroup Spatial Asset Management [distributor], 2012. [privately distributed].

Youth Explorer website. Explore, visualise and interact with youth-centered data: Higher Education, 2021 [data set]. Cape Town: Youth Project, Southern Africa Labour and Development Research Unit [producer]. Accessed: https://www.youthexplorer.org.za.

Appendices

A Terminology of education institutions in South Africa

Basic education runs from grade R to grade 12 - primary schooling refers to grade R to 7, and secondary schooling refers to grade 8 to 12. According to the South African Schools Act (1996), compulsory education consists of schooling from grade R to grade 9 or until the learner is 15 years of age - whichever occurs first. Basic education is designed for students to complete grade 12 at age 18 but students legally have the option to leave the basic education stream after completing grade 9 (or turning 15 years old). The provision of post-school education and training, therefore, is envisaged to serve those who have - and have not - completed primary and secondary schooling, and those who have never attended school (Department of Higher Education and Training, 2013).

The three main branches of institution types within South Africa's public post-school education system are: (1) higher education institutions (universities and technikons), (2) Further Education and Training (FET)/Technical and Vocational Education and Training (TVET) colleges, and (3) Adult Basic Education and Training (ABET)/Community Education and Training (CET) colleges. Of these three main types of post-school institutions, this paper chooses to assess only public higher education institutions and FET/TVET colleges. ABET/CET institutions, as well as private universities and private colleges, are not assessed in this paper.

Universities and technikons collectively made up what was known as higher education institutions in the years prior to 2005. From 2005 onwards, higher education became synonymous with universities only. On the other hand, TVET colleges were previously known as FET colleges prior to 2013. For simplicity, I use the term FET/TVET colleges to describe what is commonly referred to today as TVET colleges when discussing the period prior to 2013 or when discussing a period when these naming conventions overlap (for example, 2005 to 2019). Given that I only use 2019 TVET data in my analysis, I use the term TVET colleges as it is most appropriate.

B Reform of higher education institutions in South Africa

Table B1 lists the total 36 higher educations that existed prior to the merging of institutions. They are categorised by their historical type classification, which was driven by apartheid legislation. Table B2 lists which institutions merged (along with their historical type classification) and the respective post-merge institution that formed as a result of the merge.

Table B1: List of total 36 higher education institutions (prior merges) by historical type

Historically Advantaged Institutions	Historically Disadvantaged Institutions
Historically White (Afrikaans) Universities (RSA)	**Historically Black Universities (RSA)**
University of the Orange Free State	University of Durban-Westville
University of Port Elizabeth	Medunsa University
University of Pretoria	University of the North
Potchefstroom University	Vista University
Rand Afrikaans University	University of the Western Cape
University of Stellenbosch	University of Zululand
Historically White (English) Universities (RSA)	**Historically Black Universities (TBVC)**
University of Cape Town	University of Fort Hare
University of Natal	North West University
Rhodes University	University of Transkei
University of the Witwatersrand	Venda University
Historically White Technikons (RSA)	**Historically Black Technikons (RSA)**
Cape Technikon	ML Sultan Technikon
Free State Technikon	Mangosuthu Technikon
Natal Technikon	Technikon Northern Transvaal
Port Elizabeth Technikon	Peninsula Technikon
Pretoria Technikon	
Vaal Triangle Technikon	**Historically Black Technikons (TBVC)**
Technikon Witwatersrand	Border Technikon
	Eastern Cape Technikon
Distance University (RSA)	North West Technikon
University of South Africa (Unisa)	
Distance Technikon (RSA)	
Technikon South Africa (TSA)	

Source: Bunting (2006b).
Notes: This table lists the total 36 public higher education institutions under apartheid by historical type. The responsible authority for all 19 of the historically advantaged institutions (for White individuals) was the Republic of South Africa (RSA). The independent republics (Transkei, Bophuthatswana, Venda, and Ciskei) were responsible for 7 of the historically disadvantaged institutions and the RSA was responsible for the other 10 historically disadvantaged institutions (for Coloured, Indian, and Black individuals).

Table B2: Higher education institution merges between 2002 and 2005

	Historical type	Pre-merge institutions	Post-merge institution
2002			
	White	Technikon Natal	Durban University of Technology
	Black	ML Sultan Technikon	
2004			
	White	University of Potchefstroom	North West University
	Black	University of the North West	
	White	University of Natal	University of KwaZulu-Natal
	Black	University of Durban Westville	
	White	University of South Africa	University of South Africa
	White	Technikon South Africa	
	White	Pretoria Technikon	Tshwane University of Technology
	Black	Technikon Northern Gauteng	
	Black	Technikon North West	
2005			
	White	University of Port Elizabeth	Nelson Mandela University
	White	Port Elizabeth Technikon	
	White	Rand Afrikaans University	University of Johannesburg
	White	Technikon Witwatersrand	
	Black	University of Transkei	Walter Sisulu University
	Black	Border Technikon	
	White	Cape Technikon	
	Black	Eastern Cape Technikon	Cape Peninsula University of Technology
	Black	Peninsula Technikon	
	Black	University of the North	University of Limpopo
	Black	Medunsa University	

Source: Department of Education (2005b).
Notes: This table shows which public higher education institutions merged between 2002 and 2005. In addition, Vista University closed and the Vista Campuses were incorporated into the University of Pretoria, Rand Afrikaans University, North West University, University of Port Elizabeth, University of the Free State, Central University of Technology, and Unisa. The East London campus of Rhodes University was incorporated into the University of Fort Hare; and The Dentistry faculty at the University of Stellenbosch was incorporated into the University of the Western Cape. The name changes that occurred include: Vaal Triangle Technikon is now called Vaal University of Technology; and Technikon Free State is now called Central University of Technology.

C Provincial representation

Using postal code data, I can estimate the respective province according to the postal code ranges for each province provided in Table C4. Using these ranges, I present the estimated number of students by home province in Table C1, the number of campuses per province by institutional type in Table C2, and the distribution of university type enrolment by students' home province for 2019 in Table C3.

Table C1: Estimated number of students by home province

	GP	KZ	MP	WC	EC	LP	NW	FS	NC	**Total**
University										
2008	40,072	28,564	16,278	15,892	13,933	10,422	5,634	7,037	2,529	132,353
2009	43,737	33,500	17,037	17,361	15,048	11,093	6,026	7,453	2,630	145,528
2010	41,804	36,481	16,748	17,730	15,606	11,060	6,395	6,885	2,611	147,417
2011	47,409	40,180	20,052	18,091	16,702	11,988	7,200	6,419	2,604	160,996
2012	44,676	37,567	19,289	17,798	15,531	11,445	7,080	6,205	2,431	153,117
2013	41,243	35,257	17,186	17,768	14,908	11,965	6,078	6,584	2,282	146,038
2014	44,174	37,809	18,264	18,616	15,643	12,394	6,757	6,964	2,396	155,437
2015	45,105	38,535	18,826	19,114	16,218	12,148	6,997	7,143	2,434	158,911
2016	38,264	31,636	18,529	18,937	16,300	12,405	6,878	8,080	2,411	147,052
2017	51,563	41,960	22,357	20,019	17,672	13,981	8,301	9,778	3,037	180,620
2018	55,530	47,328	24,229	19,361	19,209	15,622	9,622	10,024	2,835	195,729
2019	46,162	43,021	22,107	17,037	19,043	14,841	9,059	9,293	2,578	176,161
TVET										
2019	99,800	61,744	73,650	63,063	59,463	38,952	31,588	16,903	107,518	552,681

Source: Author's constructed data set.
Notes: This table shows the estimated number of students by students' home province, as depicted in Figure 22. The provinces are Gauteng (GP), KwaZulu-Natal (KZ), Mpumalanga (MP), Western Cape (WC), Eastern Cape (EC), Limpopo (LP), North West (NW), Free State (FS), and Northern Cape (NC). The sample of university students are restricted to undergraduate, first-time entering.

Table C2: Number of campuses per province by institutional type

	TVET	**University**				
		Comprehensive	Technology	Traditional	Unisa	**Total**
Eastern Cape	46	10	0	4	1	15
Free State	29	0	2	3	0	5
Gauteng	40	1	8	7	7	23
KwaZulu-Natal	67	2	8	5	1	16
Limpopo	22	1	0	1	1	3
Mpumalanga	44	4	3	3	2	12
Northern Cape	11	1	1	0	0	2
North West	17	1	1	2	1	5
Western Cape	47	1	8	12	1	22
Total	**323**	**21**	**31**	**37**	**14**	**103**

Source: Youth Explorer website (2021). Author's own calculations.
Notes: This table shows the number of campuses per province by institutional type, as depicted in Figure 24a and Figure 24b.

Table C3: Distribution of university type enrolment by students' home province in 2019

	Comprehensive	Technology	Traditional	Unisa	**Total**
Eastern Cape	48.92	11.59	28.82	10.67	100
	9,315	2,208	5,489	2,031	19,043
Free State	4.26	41.92	43.03	10.78	100
	396	3,896	3,999	1,002	9,293
Gauteng	13.06	14.98	33.39	38.58	100
	6,027	6,915	15,412	17,808	46,162
KwaZulu-Natal	13.10	28.66	27.96	30.28	100
	5,637	12,328	12,028	13,028	43,021
Limpopo	23.99	19.68	32.34	23.99	100
	3,560	2,921	4,799	3,561	14,841
Mpumalanga	19.00	19.07	34.19	27.74	100
	4,201	4,215	7,558	6,133	22,107
Northern Cape	22.19	15.01	47.17	15.63	100
	572	387	1,216	403	2,578
North West	8.63	14.26	51.94	25.17	100
	782	1,292	4,705	2,280	9,059
Western Cape	4.08	21.44	60.39	14.10	100
	695	3,652	10,288	2,402	17,037

Source: Author's constructed data set.
Notes: This table shows the share and number of undergraduate, first-time entering university students who attend each university type students' estimated home province. This data is depicted in Figure 23. It reveals some university types attract more students than others within the province where students grew up.

Table C4: Postal code range for each province

Province	Postal code range		
Gauteng	0001	to	0299
	1400	to	2199
North West	0300	to	0499
	2500	to	2899
Mpumalanga	1000	to	1399
	2000	to	2499
Limpopo	0500	to	0999
KwaZulu-Natal	2900	to	4730
Eastern Cape	4731	to	6499
Western Cape	6500	to	8099
Northern Cape	8100	to	8999
Free State	9300	to	9999

Source: Lombaard (2009).
Notes: This table shows the postal codes office sorting lines. These ranges are used to estimate the associated province from postal code data.

D Constructing the socio-economic indicators

Each variable of interest is represented as a categorical variable in a distinct data file for the 2011 census community profile data. The data is at the SAL-level and provides the number of individuals (or households) that have responded to each category. To construct the socio-economic indicators at the SAL-level I, therefore, calculate the number of individuals that have responded to my category of interest as a share of the total number of responses.

For example, one of the community profile data sets I use captures access to the internet at the household level. It is expressed as the number of households that mainly access the internet from: (1) home, (2) cellphone, (3) work, (4) elsewhere, or (5) no access to the internet. As the question asks what is the main form of access, respondents can only select one category. I divide the first category, the number of households that mainly access the internet from home, by the total number of household responses to calculate an estimate for the share of households with household access to the internet within each small area layer. As a hypothetical example, I may calculate that 15% of households within a particular small area layer have household access to the internet. This then becomes one of my SAL characteristic estimates. However, indicators such as the mean number of years of education per individual or the mean annual household income in a postal code are constructed differently to the other socio-economic indicators.

D.1 Indicator for number of years of education

Table D1 lists the highest level of education categories provided in the 2011 census community profile database. I use this data set to estimate the mean number of years of education per individual per SAL.

The data set specifies how many individuals in each SAL fall within each education category. I then associate an estimated number of years of education typically required to reach each level of education, as specified in Table D1. The number of years associated with each level of education ranges from 0 to 18 years, where valid years of education only start from the completion of grade 1. Individuals who responded with grade 0, other, no schooling, unspecified or not applicable as their highest level of education were assigned 0 years of education.

I, therefore, multiply the number of individuals in each category by its estimated number of years. Summing these values up per SAL leaves me with the total number of years of education represented amongst the residents of each SAL. I then divide this total years of education by the total number of individuals in the SAL to represent the mean number of years of education per individual per SAL.

Table D1: Highest level of education categories in 2011 census community profile data set and assigned number of years of education

Highest level of education	Number of years of education
Basic education	
Grade 0	0
Grade 1	1
Grade 2	2
Grade 3	3
Grade 4	4
Grade 5	5
Grade 6	6
Grade 7	7
Grade 8	8
Grade 9	9
Grade 10	10
Grade 11	11
Grade 12	12
NC(V)s, NATED 191	
NTC I/N1/NCV level 2	10
NTC II/N2/NCV level 3	11
NTC III/N3/NCV level 4	12
N4/NTC 4	13
N5/NTC 5	13
N6/NTC 6	13
Certificates	
Certificate with less than Grade 12	13
Certificate with Grade 12	13
Diplomas	
Diploma with less than Grade 12	14
Diploma with Grade 12	14
Higher diplomas and degrees	
Higher Diploma	15
Post Higher Diploma Masters and Doctoral Diploma	16
Bachelors Degree	15
Bachelors Degree and Postgraduate Diploma	16
Honours Degree	16
Higher Degree Masters/PhD	18

Source: Statistics South Africa. South African Census Community Profiles 2011.
Notes: This table shows the highest level of education categories used by the 2011 census questionnaire and the corresponding number of years I associate with this level of education.

D.2 Indicator for annual household income

Each category in the annual household income data set is an income band. Hence, the data tells the user how many households fall within each income band per SAL. Rather than working with a share variable for income (like the other socio-economic indicators I construct), I estimate the mean annual household income per SAL.

I calculate this by multiplying the number of households in each income band by the respective imputed mid point value except for the lowest income band (no income) and highest income band (R2 457 601.00 or more). For the no income band I impute R1.00, what is regarded as a sufficiently small income value. For the highest income band, I double R2 457 601.00 (Simkins, 2004). There are 675 households that provide unspecified annual income figures. I choose to drop these households from my estimation because I did not have any further information about these households to impute a reasonable estimate for them.

Once multiplying the number of households by these imputed values, I sum the values up to represent the total annual household income in each SAL. I also sum the total number of household responses (every household except the households that provided unspecified answers). I then divide the estimated total annual household income by the number of household responses in each SAL to calculate the mean annual household income indictor that is then reweighed to be representative of the postal codes.

Table D2: Annual household income categories in the 2011 census community profile data set and imputed midpoint values

	Income band	Imputed midpoint
1	No income	R1.00
2	R1.00 to R4 800.00	R2 400.50
3	R4 801.00 to R9 600.00	R7 200.50
4	R9 601.00 to R19 600.00	R14 600.50
5	R19 601.00 to R38 200.00	R28 900.50
6	R38 201.00 to R76 400.00	R57 300.50
7	R76 401.00 to R153 800.00	R115 100.50
8	R153 801.00 to R307 600.00	R230 700.50
9	R307 601.00 to R614 400.00	R461 000.50
10	R614 401.00 to R1 228 800.00	R921 600.50
11	R1 228 801.00 to R2 457 600.00	R1 290 240.50
12	R2 457 601.00 or more	R4 915 202.00
13	Unspecified	Dropped

Source: Statistics South Africa. South African Census Community Profiles 2011.
Notes: This table shows the annual household income categories used by the 2011 census questionnaire and the associated imputed mid-points I use.

E Comparison of socio-economic measures

Table E1 presents a comparison of the mean and GE(1) socio-economic estimates using different geospatial weights. Regardless of which geospatial weight is chosen, the socio-economic estimates prove to be relatively similar.

Table E1: Comparison of socio-economic estimates using different geospatial weights for the analysis sample of university students in 2019

	Mean			GE(1)		
	SAL	Sub Place	Main Place	SAL	Sub Place	Main Place
Individual characteristics						
Language (%)						
Afrikaans	6.648	12.443	15.760	0.766	0.696	0.654
English	14.796	12.187	12.326	0.711	0.649	0.609
isiXhosa	15.149	7.596	12.821	1.217	1.135	1.153
isiZulu	26.193	26.336	22.634	0.885	0.675	0.853
Education (%)						
No school	7.492	6.772	7.095	0.224	0.205	0.204
Grade 12	17.298	19.017	17.529	0.086	0.061	0.082
Post-school qualification (all)	9.729	9.517	9.779	0.345	0.267	0.318
Certificates	2.249	2.396	2.223	0.252	0.223	0.201
Diplomas, Bachelors, Postgraduate	7.480	7.121	7.556	0.415	0.329	0.381
Years	6.920	7.114	6.986	0.024	0.019	0.024
Employment (%)						
Employed	48.034	50.046	48.541	0.083	0.050	0.058
Unemployed	11.625	12.683	11.997	0.156	0.126	0.119
Discouraged	4.724	4.078	4.544	0.325	0.258	0.246
Not economically active	35.616	33.192	34.919	0.067	0.049	0.050
Type of geographical area (%)						
Urban	42.391	49.830	44.137	0.602	0.470	0.558
Household characteristics						
Infrastructure (%)						
Electricity (light)	74.755	76.560	74.271	0.048	0.032	0.042
Piped water	45.803	47.133	45.875	0.233	0.157	0.197
Asset ownership (%)						
Car	32.577	31.194	31.710	0.193	0.157	0.174
Cellphone	88.110	89.139	88.048	0.003	0.002	0.003
Computer	22.994	22.611	23.071	0.319	0.24	0.290
DVD player	50.792	51.989	50.729	0.071	0.059	0.070
Internet	9.942	9.235	10.057	0.491	0.374	0.410
Satellite television	24.790	24.818	24.638	0.231	0.184	0.212
Radio	66.901	65.454	65.532	0.014	0.015	0.013
Washing machine	31.042	29.888	30.701	0.312	0.245	0.287
Annual household income (2011 ZAR)						
Mean income	R124 497.30	R116 401.30	R121 636.70	0.322	0.238	0.268

Source: Author's constructed data set using different geospatial weights.
Notes: This table presents the mean and GE(1) socio-economic estimates for the 2019 cohort of undergraduate, first-time entering university students using different geospatial weights. The Small Area Layer (SAL) estimates are used in the paper.

Table E2 presents a comparison of the mean socio-economic estimates for different university sample restrictions. The table shows the distinction in the socio-economic estimates between undergraduate and postgraduate students, with postgraduate students showing higher socio-economic characteristics.

Table E2: Mean socio-economic characteristics by type of student body at universities in 2019

	All	Undergraduate		Postgraduate	
		All UG	UGFTEN	All PG	PGFTEN
Individual characteristics (postal code)					
Language (%)					
Afrikaans	7.97	7.52	6.65	10.41	11.38
English	18.25	16.65	14.80	26.85	26.61
isiXhosa	14.01	14.47	15.15	11.98	13.49
isiZulu	22.76	23.90	26.19	16.72	16.73
Education (%)					
No school	6.91	7.09	7.49	5.97	6.29
Grade 12	17.98	17.83	17.30	18.70	18.18
Post-school qualification (all)	11.11	10.50	9.73	14.26	13.20
Certificates	8.64	8.11	7.48	11.40	10.52
Diplomas, Bachelors, Postgraduate	2.47	2.39	2.25	2.85	2.69
Years	7.12	7.05	6.92	7.48	7.36
Employment (%)					
Employed	49.97	49.26	48.03	53.50	52.49
Unemployed	11.23	11.47	11.63	9.94	10.24
Discouraged	4.22	4.38	4.72	3.36	3.63
Not economically active	34.59	34.88	35.62	33.19	33.65
Type of geographical area (%)					
Urban	47.93	46.36	42.39	55.90	52.14
Household characteristics (postal code)					
Infrastructure (%)					
Electricity (light)	77.10	76.31	74.75	81.13	79.99
Piped water	50.08	48.61	45.80	57.64	55.45
Asset ownership (%)					
Car	35.85	34.60	32.58	42.33	40.38
Cellphone	88.83	88.58	88.11	90.10	89.45
Computer	26.09	24.85	22.99	32.50	30.57
DVD player	53.16	52.38	50.79	57.14	55.87
Internet	11.46	10.77	9.94	15.06	14.14
Satellite television	27.32	26.40	24.79	32.09	30.53
Radio	68.04	67.63	66.90	70.17	69.45
Washing machine	34.70	33.35	31.04	41.70	39.93
Annual household income (2011 ZAR)					
Mean income	R141 044.00	R133 672.90	R124 497.30	R179 318.40	R166 671.70

Source: Author's constructed data set.
Notes: This table shows how the mean socio-economic estimates of university students' home postal code compare for different sample group restrictions. The sample restrictions presented in this table are for all university students (with valid postal codes), all undergraduate (all UG), undergraduate first-time entering (UGFTEN), all postgraduate (all PG), and postgraduate first-time entering (PGFTEN).

F Choice of the generalised entropy index to measure socio-economic heterogeneity of student bodies

I use the generalised entropy index, specifically the Theil-T index, to descriptively estimate the extent of socio-economic heterogeneity between student bodies at universities and TVET colleges, between student bodies of different university types, and within each university. The Theil-T index is predominantly used to measure economic inequality. However, there are a number of alternative measures also used in the economic inequality literature. The choice of inequality measures depend on the inequality axiom properties the inequality measure satisfies (Cowell, 1985). Further considerations depend on the unit of data and the type of question being answered. In this section, I provide the theoretical motivation for my choice of the generalised entropy index to measure socio-economic heterogeneity of student bodies.

F.1 Inequality axioms

There are four key axioms an inequality measure should satisfy: anonymity (symmetry), population invariance, scale invariance, and the transfer principle (Foster et al., 2013). Anonymity states if two people within a society switched measures, the extent of inequality will remain the same. Hence, the identity of the individual associated with the measure is not relevant. Population invariance states that the measure of inequality is independent of the size of the society under investigation if the vector of units is multiplied by some factor. Scale invariance states that the measure of inequality will not change if each unit within the population increased by the same proportion. Hence, the size of the units are irrelevant as inequality is a relative concept (Shifa and Ranchhod, 2019). The transfer axiom states that a regressive transfer between two people in a society should increase inequality and a progressive transfer between two people should reduce inequality. However, the transfer axiom does not specify where in the distribution the transfer must take place. Transfer sensitivity, the sixth axiom, states that the inequality measure is more sensitive to transfers at the lower end of the distribution.

The sixth axiom is normalisation which states that the inequality index is zero when the income distribution is egalitarian. An additional property, axiom seven, is decomposability whereby inequality may be broken down by groups. This property includes the additive decomposability axiom and sub-group consistency axiom. Additive decomposability means that the overall inequality measure is equal to a weighted sum of the sub-group inequalities. Sub-group consistency means that if inequality rises within a particular sub-group and none of the other sub-groups experience a decline in inequality, the overall inequality measure will rise (Shifa and Ranchhod, 2019).

The Theil indices satisfy all seven axioms (Haughton and Khandker, 2009), thus making it an ideal measure for my analysis.

F.2 Why GE(1) is best suited for my analysis

A special case of the generalised entropy index is at $\alpha = 1$ (also known as the Theil-T index) or $\alpha = 0$ (also known as the Theil-L index or the mean logarithmic deviation). A useful feature of the Theil

indices ($\alpha = 1$ or $\alpha = 0$) is that under the decomposition formula, the weights of the respective sub-groups sum to one; thus giving the measure a between-group and within-group inequality interpretation (Foster et al., 2013). For values of $\alpha \neq 1$ or $\alpha \neq 0$, the weights may not always sum to one. A further distinction here is that when $\alpha = 1$, the sub-groups are weighted by the proportion of the socio-economic indicator accruing to the j^{th} group ($\frac{X_j}{x}$); and when $\alpha = 0$, the sub-groups are weighted by the share of individuals accruing to the j^{th} group ($\frac{N_j}{N}$).

Between 2008 and 2019, the total number of undergraduate, first-time entering university students in the data set I create increases from 132,353 students to 176,161 students (see Table 13). Furthermore, growth in the number of these students varies across the types of universities (see Table 14) and across each particular university. Hence, this volatility in the sub-group weights for $\alpha = 0$ is not ideal because it will distort the interpretation of the socio-economic heterogeneity estimates over time. As I hold the socio-economic estimates of each postal code constant using 2011 census data, changes in the proportion of the socio-economic indicator accruing to the j^{th} group (namely, when $\alpha = 1$) are the result of changes in the composition of students' postal codes within the j^{th} group. This is the analogy I am most interested in. Hence, I adopt $\alpha = 1$ when calculating the generalised entropy indices of the postal code socio-economic indicators to estimate the heterogeneity of university and TVET college student bodies.

The decomposition of Theil-T ($\alpha = 1$) and Theil-L ($\alpha = 0$) is as follows:

$$\begin{aligned} GE(\alpha = 1) &= \frac{1}{N}\sum_{i=1}^{N}\frac{x_i}{\mu}ln\frac{x_i}{\mu} \\ &= \sum_{i=1}^{N}\frac{x_i}{N\mu}ln(\frac{x_iN}{N\mu}) \\ &= \sum_{i=1}^{N}\frac{x_i}{X}ln(\frac{x_iN}{X}) \\ &= \sum_{j}T_j + \sum_{j}\frac{X_j}{X}ln(\frac{X_j/X}{N_j/N}) \end{aligned} \tag{5}$$

$$\begin{aligned} GE(\alpha = 0) &= \frac{1}{N}\sum_{i=1}^{N}ln\frac{\mu}{x_i} \\ &= \sum_{i-1}^{N}\frac{1}{N}ln(\frac{\mu}{X_i}) \\ &= \sum_{j}(\frac{N_j}{N})L_j + \sum_{j}\frac{N_j}{N}ln(\frac{\mu}{\mu_j}) \end{aligned} \tag{6}$$

F.3 Decomposition of GE(1) estimates for socio-economic indicators by university type

The GE(1) estimates presented in Table F1 to Table F6 are discussed in Section 7.2.

Table F1: GE(1) estimates for the share of individuals with a post-school qualification by university type

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Comprehensive universities	0.408	0.396	0.405	0.386	0.391	0.392	0.378	0.379	0.378	0.382	0.381	0.383
Universities of technology	0.320	0.320	0.320	0.324	0.328	0.321	0.321	0.322	0.313	0.315	0.318	0.325
Traditional universities	0.301	0.296	0.298	0.320	0.322	0.321	0.324	0.320	0.328	0.337	0.340	0.339
Unisa	0.286	0.294	0.308	0.315	0.308	0.272	0.264	0.267	0.249	0.269	0.284	0.296
All universities	**0.325**	**0.329**	**0.330**	**0.337**	**0.340**	**0.333**	**0.332**	**0.329**	**0.337**	**0.331**	**0.335**	**0.345**
Relative contribution (out of 1.00)												
Within university types	0.969	0.961	0.971	0.972	0.964	0.959	0.954	0.952	0.951	0.958	0.962	0.959
Between university types	0.031	0.039	0.029	0.028	0.036	0.041	0.046	0.048	0.049	0.042	0.038	0.041

Source: Author's constructed data set.
Notes: $\alpha = 1$. See Figure 27 for a visual representation of the data presented in this table.

Table F2: GE(1) estimates for the share of individuals employed by university type

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Comprehensive universities	0.117	0.114	0.118	0.112	0.115	0.118	0.116	0.122	0.121	0.123	0.128	0.126
Universities of technology	0.067	0.065	0.068	0.070	0.074	0.073	0.074	0.076	0.073	0.080	0.083	0.088
Traditional universities	0.053	0.053	0.054	0.059	0.061	0.063	0.064	0.063	0.064	0.067	0.068	0.070
Unisa	0.052	0.054	0.060	0.064	0.060	0.054	0.051	0.050	0.046	0.051	0.059	0.065
All universities	**0.067**	**0.069**	**0.070**	**0.072**	**0.073**	**0.074**	**0.074**	**0.073**	**0.077**	**0.075**	**0.078**	**0.083**
Relative contribution (out of 1.00)												
Within university types	0.970	0.968	0.972	0.975	0.972	0.968	0.968	0.965	0.963	0.969	0.973	0.974
Between university types	0.030	0.032	0.028	0.025	0.028	0.032	0.032	0.035	0.037	0.031	0.027	0.026

Source: Author's constructed data set.
Notes: $\alpha = 1$. See Figure 27 for a visual representation of the data presented in this table.

Table F3: GE(1) estimates for the mean annual household income by university type

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Comprehensive universities	0.402	0.379	0.383	0.372	0.375	0.382	0.373	0.363	0.361	0.363	0.355	0.355
Universities of technology	0.279	0.278	0.275	0.282	0.286	0.281	0.279	0.277	0.272	0.271	0.278	0.275
Traditional universities	0.285	0.276	0.283	0.303	0.309	0.305	0.306	0.307	0.312	0.318	0.321	0.322
Unisa	0.278	0.281	0.294	0.298	0.289	0.255	0.248	0.250	0.231	0.254	0.271	0.280
All universities	**0.308**	**0.306**	**0.308**	**0.317**	**0.320**	**0.314**	**0.313**	**0.310**	**0.317**	**0.309**	**0.315**	**0.322**
Relative contribution (out of 1.00)												
Within university types	0.968	0.960	0.971	0.970	0.960	0.953	0.948	0.944	0.943	0.953	0.956	0.954
Between university types	0.032	0.040	0.029	0.030	0.040	0.047	0.052	0.056	0.057	0.047	0.044	0.046

Source: Author's constructed data set.
Notes: $\alpha = 1$. See Figure 27 for a visual representation of the data presented in this table.

Table F4: GE(1) estimates for the share of households with access to the internet by university type

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Comprehensive universities	0.556	0.543	0.553	0.541	0.550	0.569	0.559	0.567	0.566	0.575	0.582	0.575
Universities of technology	0.467	0.466	0.464	0.474	0.480	0.476	0.471	0.471	0.464	0.469	0.478	0.481
Traditional universities	0.396	0.384	0.396	0.413	0.417	0.419	0.422	0.421	0.438	0.448	0.452	0.455
Unisa	0.390	0.388	0.412	0.421	0.413	0.390	0.373	0.371	0.350	0.373	0.411	0.432
All universities	**0.442**	**0.442**	**0.448**	**0.454**	**0.458**	**0.463**	**0.459**	**0.456**	**0.472**	**0.461**	**0.476**	**0.491**
Relative contribution (out of 1.00)												
Within university types	0.972	0.962	0.973	0.974	0.967	0.960	0.954	0.952	0.951	0.957	0.959	0.955
Between university types	0.028	0.038	0.027	0.026	0.033	0.040	0.046	0.048	0.049	0.043	0.041	0.045

Source: Author's constructed data set.
Notes: $\alpha = 1$. See Figure 27 for a visual representation of the data presented in this table.

Table F5: GE(1) estimates for the share of households with access to a computer by university type

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Comprehensive universities	0.390	0.376	0.394	0.374	0.383	0.395	0.388	0.395	0.395	0.404	0.411	0.407
Universities of technology	0.288	0.284	0.284	0.294	0.303	0.298	0.296	0.300	0.293	0.300	0.307	0.316
Traditional universities	0.249	0.242	0.251	0.266	0.272	0.271	0.273	0.270	0.281	0.291	0.292	0.296
Unisa	0.229	0.233	0.254	0.263	0.258	0.226	0.211	0.216	0.199	0.221	0.248	0.263
All universities	**0.279**	**0.280**	**0.286**	**0.291**	**0.296**	**0.294**	**0.292**	**0.291**	**0.304**	**0.295**	**0.304**	**0.319**
Relative contribution (out of 1.00)												
Within university types	0.968	0.961	0.973	0.975	0.969	0.961	0.954	0.951	0.950	0.956	0.962	0.959
Between university types	0.032	0.039	0.027	0.025	0.031	0.039	0.046	0.049	0.050	0.044	0.038	0.041

Source: Author's constructed data set.
Notes: $\alpha = 1$. See Figure 27 for a visual representation of the data presented in this table.

Table F6: GE(1) estimates for the share of households with access to electricity by university type

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Comprehensive universities	0.059	0.058	0.062	0.058	0.061	0.060	0.061	0.060	0.061	0.063	0.065	0.063
Universities of technology	0.039	0.038	0.040	0.043	0.046	0.045	0.046	0.048	0.045	0.049	0.051	0.054
Traditional universities	0.032	0.032	0.033	0.035	0.037	0.038	0.038	0.038	0.040	0.043	0.043	0.044
Unisa	0.029	0.030	0.034	0.036	0.034	0.029	0.027	0.027	0.024	0.029	0.034	0.037
All universities	**0.038**	**0.038**	**0.040**	**0.041**	**0.042**	**0.042**	**0.042**	**0.041**	**0.044**	**0.044**	**0.045**	**0.048**
Relative contribution (out of 1.00)												
Within university types	0.976	0.976	0.979	0.984	0.983	0.979	0.974	0.972	0.976	0.976	0.978	0.981
Between university types	0.024	0.024	0.021	0.016	0.017	0.021	0.026	0.028	0.024	0.024	0.022	0.019

Source: Author's constructed data set..
Notes: $\alpha = 1$. See Figure 27 for a visual representation of the data presented in this table.

F.4 Alternative inequality measures

Table F7 presents a range of alternative inequality measures for annual household income – such as decile ratios (99:1, 90:10, 90:50, 50:10), the Gini coefficient, and the Generalised Entropy index with $\alpha = 0$ and $\alpha = 1$. The purpose of calculating these alternative inequality measures is to ensure that the heterogeneity outcomes discussed in the paper are not solely drive by the choice of the GE(1) measure.

For the postal code income distribution of students enrolled at comprehensive universities, those at the 99th percentile are from postal codes where the mean annual household income is 31.761 times higher than students at at the 1st percentile. This is the highest disparity between students at the 99th versus 1st percentile within their institution and motivates why comprehensive universities tend to show the highest level of heterogeneity discussed in Section 7.2. For the 90:10, 90:50 and 50:10 decile ratios, student bodies at traditional universities show the largest differences, while student bodies at TVET colleges produce the lowest differences amongst the institutional types.

For the Gini coefficient and GE(1) estimate, the comprehensive university student body displays the highest level of heterogeneity and the TVET college student body displays the lowest level of heterogeneity. Traditional universities display the highest GE(0) estimate but this is only marginally higher than the GE(0) estimate for comprehensive universities.

Table F7: Inequality measures for students' postal code mean annual household income in 2019

		Undergraduate, first-time entering university students				
	TVET	All	Comprehensive	Technology	Traditional	Unisa
Decile ratio						
99:1	19.942	24.567	31.761	18.750	25.658	19.042
90:10	5.919	8.745	7.139	6.499	9.516	7.355
90:50	2.632	3.637	3.269	3.038	3.594	3.403
50:10	2.248	2.404	2.184	2.139	2.648	2.162
Gini coefficient	0.376	0.439	0.452	0.400	0.444	0.410
Generalised entropy						
GE(0)	0.232	0.321	0.337	0.259	0.341	0.277
GE(1)	0.243	0.322	0.355	0.275	0.322	0.280

Source: Author's constructed data set.
Notes: This table presents a range of inequality measures to ensure the heterogeneity outcomes discussed in the paper are not solely driven by the choice of the GE(1) measure.

G Ethics approval and plagiarism declaration

This research forms part of the Siyaphambili project within the Southern Africa Labour and Development Research Unit (SALDRU). This project has been granted access to use select variables from the raw HEMIS (2008 to 2019) and TVETMIS (2019) databases.

Faculty of Commerce
Private Bag X3, Rondebosch, 7701
2.26 Leslie Commerce Building, Upper Campus
Tel: +27 (0) 21 650 4375/ 5748 Fax: +27 (0) 21 650 4369
E-mail: com-faculty@uct.ac.za
Internet: www.uct.ac.za
@Commerce UCT **UCT Commerce Faculty Office**

09th November 2021

Kim Ingle
School of Economics
University of Cape Town

Dear Kim Ingle

REF: REC 2019/10/045

INEQUALITIES IN ACCESS AND SUCCESS IN HIGHER EDUCATION

We are pleased to inform you that your ethics extension has been approved. Unless otherwise specified this ethical extension is valid until 31 December 2022 and may be renewed upon application.

Please be aware that you need to notify the Ethics Committee immediately should any aspect of your study regarding the engagement with participants as approved in this application, change. This may include aspects such as changes to the research design, questionnaires, or choice of participants.

The ongoing ethical conduct throughout the duration of the study remains the responsibility of the principal investigator.

We wish you well for your research.

Shandre Swain
Administrative Assistant
University of Cape Town
Commerce Faculty Office
Room 2.26 | Leslie Commerce Building

Office Telephone: +27 (0)21 650 2695 / 4375
Office Fax: +27 (0)21 650 4369
E-mail: sl.swain@uct.ac.za
Website: www.commerce.uct.ac.za<http://www.commerce.uct.ac.za/

UNIVERSITY OF CAPE TOWN
FACULTY OF COMMERCE
Igniting Knowledge and Opportunity

Plagiarism Declaration

COMPULSORY DECLARATION:

1. This dissertation has been submitted to Turnitin (or equivalent similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.
2. I certify that I have received Ethics approval (if applicable) from the Commerce Ethics Committee.
3. This work has not been previously submitted in whole, or in part, for the award of any degree in this or any other university. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works of other people has been attributed, and has been cited and referenced.

Student number	CLLSAM004
Student name	Samantha Culligan
Signature of Student	Signed by candidate
Date:	12 February 2022