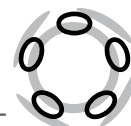# Assessing the Usability of the Western Cape Graduate Destination Survey for the Analysis of Labour Market Outcomes

Nicola Branson and Murray Leibbrandt

Assessing the Usability of the Western Cape Graduate Destination Survey for the Analysis of Labour Market Outcomes

Nicola Branson and Murray Leibbrandt

LABOUR MARKET
INTELLIGENCE PARTNERSHIP

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

ABET	Adult Basic Education and Training
CPUT	Cape Peninsula University of Technology
DHET	Department of Higher Education and Training
EMIS	Educational Management Information System
GHS	General Household Survey
HEMIS	Higher Education Management Information System
HSRC	Human Sciences Research Council
LMIP	Labour Market Intelligence Partnership
NCV	National Certificate (Vocational)
NIDS	National Income Dynamics Study
NGDS	National Graduate Destination Survey
NSCE	National Senior Certificate Examinations
NSFAS	National Student Financial Aid Scheme
SU	Stellenbosch University
TVET	Technical and Vocational Education and Training
UCT	University of Cape Town
UIF	Unemployment Insurance Fund
UNISA	University of South Africa
UWC	University of the Western Cape
WCGDS	Western Cape Graduate Destination Survey

PREFACE

In 2009 the South African government administration, informed by a results-focused philosophy, identified 12 priority outcomes for the country. Outcome 5 refers to 'a skilled and capable workforce to support an inclusive growth path', and the delivery of this outcome is led by the Minister of Higher Education and Training. Delivery Agreement 5 consists of three parts, with Output 5.1 committing the Department of Higher Education and Training (DHET) to establish a credible mechanism for skills planning, in collaboration with 20 national and provincial ministries. The DHET commissioned the Human Sciences Research Council (HSRC) to support the DHET in establishing a credible institutional mechanism for skills planning (Memorandum of Agreement between the DHET and the HSRC, February 2012). Thus the Labour Market Intelligence Partnership (LMIP) project, with six themes of research, was established.

The objective of one of the research themes is to obtain a better understanding of the pathways and transitions undertaken by young people through the education and training system into the workplace. The key question underpinning this work is: What are the dynamics of access, progression, graduation and labour market destinations along various education, training and labour market trajectories, and how can this knowledge inform skills planning in South Africa? The research therefore collected and analysed data which then provides crucial information on the following:

- Understanding the extent to which **access** is conditioned by socio-economic factors, the quality of primary and secondary schooling, as well as spatial and demographic characteristics. In particular, it is important to know which barriers affect young people who successfully finish their schooling.
- **Pathways** or trajectories through the secondary school and post-school sector refer to the choices that students make in terms of institutions, subjects, degrees and specialisations.
- **Transitions** from and through education and training into the labour market are the final step in the progression sequence. Given the large investments (at both the household and government levels) made in training and higher education, the successful matching of available skills to the demands of the labour market is of significant interest in South Africa.

The post-school education and training landscape in South Africa consists of a diverse range of sectors and institutions. These include: Adult Basic Education and Training (ABET) centres; Technical and Vocational Education and Training (TVET) colleges; workplace training programmes (learnerships and apprenticeships); as well as traditional, comprehensive and universities of technology. All of these components of the post-schooling system are of vital importance to the supply of skills to the labour market and the broader

South African economy, and understanding the issues of access, pathways and transitions will provide valuable information for skills planning.

A number of research studies were conducted within this theme of research. The key questions that each of the studies attempted to answer is reflected in the following topics:

1. What is the progression, graduation and destination of secondary school students?
2. How matric results influence university access, field of study and progression through to university.
3. What are the school-to-work transitions in the National Income Dynamic Study?
4. What are the university graduate destination outcomes: The Eastern Cape study on transitions to the labour market
5. Assessing the usability of graduate destination surveys for the analysis of labour market outcomes.
6. Scoping for a tracer study of the education and training and labour market outcomes of workplace training programmes.
7. What are the pathways of TVET college learners through the TVET colleges and beyond?
8. Who accesses adult education programmes and where do they progress to: An exploratory tracer study on community education and training centres.

ABSTRACT

Graduate destination studies have the potential to provide detailed information about graduate transitions to work that cannot easily be collected in household surveys. However, response rates are typically very low and raise the concern that the non-response is not random and that inferences using data on those who respond will be inaccurate. This study examines response rates in the Western Cape Graduate Destination Survey where 22% of all 2010 university graduates from the four Western Cape universities were successfully interviewed in 2012. We examine differences in observable baseline characteristics, assess the extent of non-response bias for a labour market participation analysis, compare rates of continued study with those in the Higher Education Management Information System (HEMIS) database, and implement a selection-correction methodology that uses type of email address as an exclusion restriction. We find that those who successfully responded to the survey were more likely to be studying in 2012 and have some systematically different baseline information that signals that response is not random. Our selection-correction methodology, however, finds limited impact for an equation of employment. This study provides important input into plans for a National Destination Survey. We recommend that the focus be directed at preparing and standardising the sampling frame, and that detailed records of the survey process be kept. In addition, we illustrate the potential benefits of linking graduate destination study data with administrative resources to assess bias and supplement the survey information obtained.

1. INTRODUCTION

Graduate destination studies are used internationally to aid higher education planning. In South Africa, limited information is available on the pathways graduates take once leaving university and, in particular, their success in the labour market. Given the resources involved in educating graduates and the concerns around skills shortages, even small levels of graduate unemployment raise concerns.

The current data available to investigate these questions come either from national household surveys, where detail is limited and data are aggregated, or from infrequent graduate destination studies (see CHEC 2013 for a summary of graduate destination studies in South Africa) that are plagued by low response rates. These data limitations have contributed to a heated public debate around the level of graduate unemployment. The blunt qualification question included in most household surveys[1] has resulted in some studies reporting high levels of 'graduate unemployment' when those with diplomas/certificates who have not completed matric are included in the definition of a graduate. This has resulted in erroneously high rates of graduate unemployment being found in the survey data (Van den Berg & Van Broekhuizen 2012; Bhorat 2004). Some of the differences in findings are attributed to the inability of the household surveys to distinguish effectively between different graduate types (HESA 2014), with the average not being representative of those in non-degree programmes and with variation across institutions being left unexplored.

Another potential contributor to these differences stems from the low response rates that plague graduate destination studies. The design of a graduate destination study makes it inherently vulnerable to selection bias. Graduate destination studies typically use administrative records to construct a sampling frame and provide baseline individual information. While this is advantageous, as it provides a complete listing of graduates, many graduates get new contact details on leaving university and starting a new phase in their lives, and baseline contact information quickly becomes out of date. Thus, attempting to contact a cohort of university leavers – even within a short period of graduation – presents a significant challenge (Du Toit et al. 2014).

The Western Cape Graduate Destination Survey (WCGDS), run by the Cape Higher Education Consortium (CHEC), attempted to contact all 2010 graduates from the four Western Cape institutions in 2012, two years after graduating. The survey relied on voluntary responses, and Table 1 shows that a response rate of 22.5% was achieved, with rates varying marginally across institutions. These responses were linked to the Higher Education Management Information System (HEMIS) database of all 2010 graduates from these institutions in order to statistically weight the successfully contacted graduates to the actual sociodemographic profile of the 2010 cohort of graduates on the basis on gender, population group, qualification type and institution (HESA 2014). This reweighting does not, however, account for selection bias, that is, for the fact that those who respond and those who do not respond could be systematically different in ways that affect their success in the labour market or other outcomes of interest.

---

1   Surveys typically ask respondents what their highest level of education is and post-schooling categories that could relate to university include certificate or diploma with matric, bachelor's degree, bachelor's degree with diploma, honours degree or higher degree (masters or doctorate) (2014 General Household Survey).

This study analyses the low response rate in the WCGDS and the possible selectivity that ensues. The study further examines non-response in the WCGDS in detail in order to:

- Document some of the procedures used in the survey and their impact on outcomes, and so provide a resource for those using these data and for those planning the National Graduation Destination Survey;
- Describe the characteristics of, and differences in characteristics between, those who responded as opposed to those who did not respond to the survey, and to explore these in a multivariate framework; and
- Propose methods to assess and account for non-response bias using national administrative databases and information on the type of contact details available.

The WCGDS was primarily designed to analyse university experience, further studies, job search, labour market outcomes, and future expectations (Du Toit et al. 2014). Our analysis is primarily focused on employment and further studies. The analysis of selection bias is outcome-specific and therefore cannot be generalised across all outcomes, since selection can be correlated with different (often unobserved) variables that affect the outcome of interest. Thus, similar analyses would be required for each topic of interest, given the substantially low response rate.

2. EXPANDING THE GRADUATE DISCUSSIONS – THE TYPE OF QUESTIONS GRADUATE DESTINATION SURVEYS AIM TO ANSWER

There is much interest in whether those who delay entry into the labour market, and continue to study beyond matric, end up on a better employment trajectory than those who do not. Studies using national household survey data show that a large number return in order to complete a post-schooling qualification. Yet the information contained in household surveys is limited in two respects. Firstly, given the small percentage of graduates in the national population (graduates accounted for 0.27% of the 2011 population – CHEC, 2013), nationally representative household studies do not achieve samples large enough to disaggregate by institution or study programme. Secondly, the information included about graduates is limited and, for cross-sectional data, only contains concurrent socio-economic information. This makes it difficult to disentangle the impact of studying from pre-existing characteristics. Panel studies such as the Cape Area Panel Study and the National Income Dynamics Study go a step further

and provide more detail on the factors associated with who attends and who does not, yet continue to suffer from the same sample-size and hence aggregation issues.

By focusing on the graduate population, graduate destination studies circumvent the sample-size issues. These surveys tend to include questions designed to ascertain not only the labour market and further-study trajectories of graduates, but also the perceived value of their qualification once working, relevance in the workplace, how much the qualification prepared the graduate for work, as well as satisfaction with work obtained. The focus on graduates allows a disaggregation by institution and field of study and also allows investigation into the match between labour force participation and labour-shortage areas. Figure 1 shows the employment rates of graduates in the Western Cape Graduate Destination Survey (WCGDS) by qualification type and area of study.

Figure 1: Employment by qualification type and subject
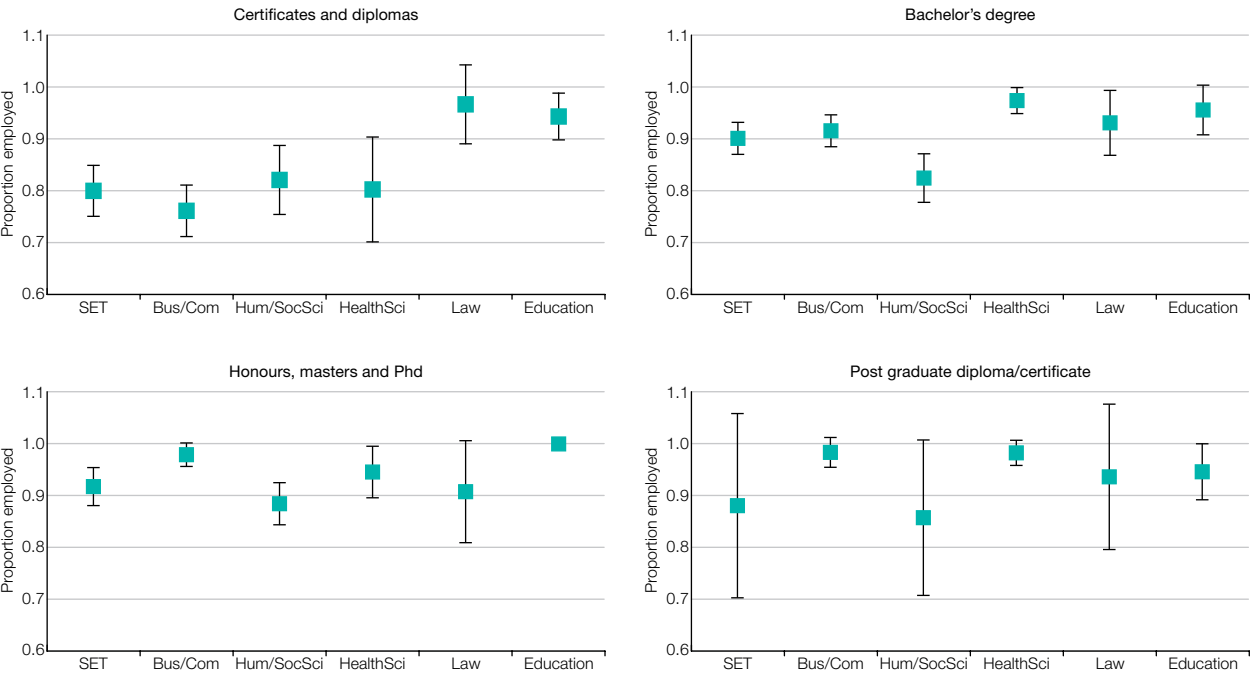
It also shows variability in employment rates across study area, particularly among graduates with lower qualifications.

The WCGDS questionnaire is divided into five main sections – time during school, time at university, employment experience, further studies, and plans for the future – so that the survey could enable analyses to determine the nature, source and success of graduate funding; the impact of career guidance on employment match; different factors improving the odds of employment; and the main job destinations of graduates, be they private or public. It also has the potential to address some of the controversy in the graduate unemployment discussion by tackling head-on the view that graduate unemployment is not a reality but rather stems primarily from data issues. One point of contention lies in the definition of a graduate. Inclusion of diploma and certificate qualifications in addition to degrees decreases the employment and earnings returns substantially (Van den Berg & Van Broekhuizen 2012). Figure 2 presents evidence of this from the WCGDS. The figure shows the employment rate of graduates two years after graduating, by institution. In the left-hand panel, we see employment rates among graduates of all qualification types. Here, there are substantial and statistically significant differences between the rates at the University of Cape Town (UCT) and Stellenbosch University (SU) and

between those at the Cape Peninsula University of Technology (CPUT) and the University of the Western Cape (UWC). Yet, this does not account for the distribution of qualifications by institution. Table 2 shows that the majority of CPUT 2010 graduates qualified with undergraduate certificates and diplomas, while the majority at the other institutions are degrees. The right-hand panel restricts the sample to those with bachelor qualifications only and results in a very different picture. This figure shows that it is not only the difference in return to a college certificate or diploma that is lower, but that there are differences in returns to bachelor degrees across universities.

Thus graduate destination studies aid us in unpacking many of the puzzles within the higher education sector that are often left untouched owing to data limitations. On the other hand, the problem with graduate destination studies is that they are prone to low response rates and a high likelihood that response is linked to different employment trajectories. As a result, it is not clear whether these data can provide unbiased estimates of the labour market trajectories of graduates. This has important implications for the university sector – getting precisely measured information that is wrong can be damaging, especially if part of the reason it is wrong is related to the institution or type of qualification or course.

Figure 2: Employment by institution – all qualifications versus bachelor qualifications only
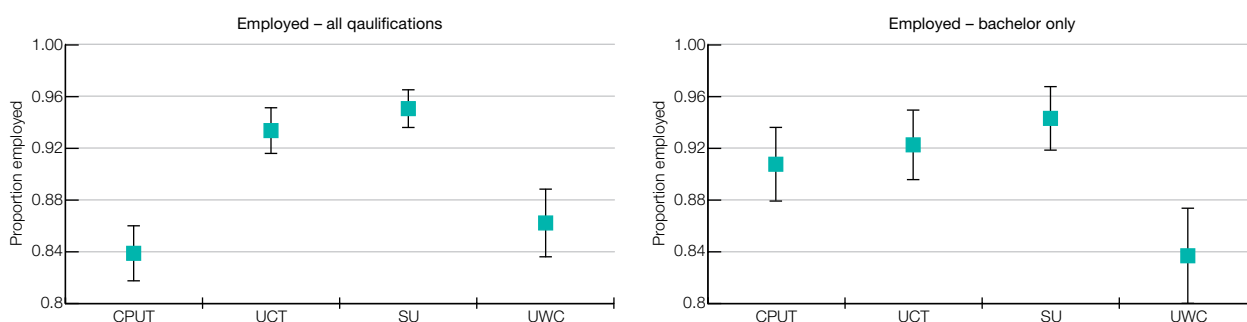
3. SAMPLE SELECTION – MODELLING THE PROBLEM

Sample selection is always a concern in that it reduces sample size and therefore the power of a survey to demonstrate relationships of significant interest. However, the main concern for a survey of this size is that those who respond are different in important ways from those who do not respond. In fact, given the large sampling frames generally available to utilise in the design of a graduate destination survey – all graduates – this concern is doubly problematic, as the realised samples are usually large enough to get precisely measured estimates even if these estimates are in fact wrong. To make this point concretely, we follow Maluccio (2004) in specifying when selection bias becomes a problem and providing a potential approach to correct for this bias.

$$y_i = x_i'\beta_1 + \varepsilon_{(i)} \quad (y_i \text{ observed only if } A_i^* < 0) \quad (1)$$

$$A_i^* = x_i'\beta_2 + z_i'y + v_{(i)} \quad (2)$$

Equation (1) represents the model of interest (in our case, we are interested in the outcome variable, y_i, being employment status). y_i is only observed for those who respond. Equation (2) is the selection equation and depends on the same independent variables as in Equation (1), x_i', in addition to some additional ones (y_i'). is a latent index and, in reality, a student either responds ($A_i = 0$) or does not ($A_i = 1$). If ε_i and v_i are correlated, estimating (1) without accounting for (2) will result in inconsistent β_2 estimates. This is the case where non-response is non-random and our estimates are biased.

From the model, it is clear that an evaluation of non-random responses is model-specific (Maluccio, 2004). When the outcome changes, the explanatory variables, x_i', and the error term, ε_i, change. Therefore ε_i and v_i can be correlated in one model and not in another, resulting in selection bias in one estimation and no selection bias in another.

Statistically weighting the realised sample to reflect the actual socio-demographic profile of the 2010 cohort of graduates on the basis of gender, population group, qualification type, and institution will only reduce the sample bias if the employment outcomes in 2012 of responding graduates defined by these strata (gender, population group, qualification type, and institution) are similar to non-responding graduates of similar baseline characteristics. It does not, however, account for other characteristics that may affect success in the labour market and the odds of responding, that is, that make $\varepsilon_{(i)}$ and $v_{(i)}$ correlated. We will show that contact details, having a bursary, studying status, and nationality are significant determinants of response. These are not taken into consideration with this weighting exercise. To the extent that these and other unobserved characteristics determine employment propensity, the weights provided in the data will not solve the sample-selection problem. Including sample weights in the Western Cape Graduate Destination Survey (WCGDS) can therefore misguide users into thinking that these data are representative and can be used for population estimates.

Another consideration is that observable differences or similarities between responders and non-responders do not necessarily imply that an estimated relationship based on those who respond is biased or not biased. As Maluccio (2004: 103) points out:

> bias could still be a problem even if there were no observable differences between the two groups; it depends on the existence of correlation between the error terms and in equations (1) and (2) shown above. For example, if attrition is selective on observable right-hand-side covariates, and the model is well specified, it may be

possible to condition on those variables allowing consistent estimation of (1) while ignoring (2). This is not an option, however, if there is selection on unobservables. In that case, a possible solution is a standard selection correction methodology (Heckman 1979; Maddala 1986).

4. DATA

We construct a database of information from the Western Cape Graduation Destination Survey (WCGDS), with additional institutional and Higher Education Management Information System (HEMIS) records from 2010 and 2012. The WCGDS, run by the Cape Higher Education Consortium (CHEC), attempted to contact all 2010 graduates from the four Western Cape institutions in 2012, two years after graduating. The sampling master list for the WCGDS contained information from HEMIS 2010 data on all 2010 graduates from these institutions, coupled with institutional contact details collected from institutions during the design of the study (Du Toit 2016). All 2010 graduates were targeted via email, with the sample further increased by contacting an additional subset by phone. The survey relied on voluntary responses, and Table 1 shows that a response rate of 22.5% was achieved, with rates varying marginally across institutions: Cape Peninsula University of Technology (CPUT (21.8%)), University of Cape Town (UCT (21.9%)), University of Stellenbosch (SU (21.6%)) and University of the Western Cape (UWC (26.7%)). Details of the survey process are available in Du Toit (2016). A few key points need to be highlighted for this analysis:

- There were institutional differences in the completeness of contact details;
- In addition, while HEMIS data was from 2010, contact details came from the most up-to-date records kept on the institutional software. Thus,

any graduates who had continued their studies would have more up-to-date information;
- Contact details from the National Student Financial Aid Scheme (NSFAS) were only sought for students at CPUT and UWC, as these institutions had a large share of students with incomplete contact details; and
- A non-random telephonic follow-up was used to increase the response rates for graduates predominantly from CPUT and UWC. No information is available on who the call centre attempted to call, but only on the mode of interview (email or telephonic) for those who did respond. Du Toit (2016) provides an analysis of the effect of the use of telephonic interviews on responses to the employment and job-search questions.

We supplement these data with records on studying status in 2012 from the HEMIS data. Unfortunately, the UWC identification numbers on file were found to be invalid and could not be matched. This match was performed by the Department of Higher Education and Training (DHET) and the de-identified data were returned to us. As a result, we can only calculate the share of graduates studying in 2012 by WCGDS response, institution and field of study. Finally, for the UCT subsample, we have additional institutional information and matched address-code information from Census 2011. This provides us with more detailed information for some parts of the analysis.

5. RESULTS SECTION

In this section, we attempt to assess whether the 22% response rate in the Western Cape Graduation Destination Survey (WCGDS) poses a concern for estimates of employment. Firstly, we compare differences in mean baseline characteristics. Finding some differences, we go on to assess the correlates of response in a multivariate framework and test whether these correlates are jointly significant. Recognising that this only accounts for observable baseline differences, we first use data external to the survey to assess bias in the proportion of students in the sample directly, and, finally, we implement a Heckman selection-type model using institutional email as the exclusion restriction in an attempt to correct the coefficients in our model of employment.

a. Mean differences

Table 3 presents different types of contact details for those who responded and those who did not, both overall and by institution. It is clear that those with more complete contact details were more likely to respond. For example, 88% of responders had cell phone details compared with 66% of non-responders. Similarly, those who responded were more likely to have an email address. National Student Financial Aid Scheme (NSFAS) information was only used for Cape Peninsula University of Technology (CPUT) and University of the Western Cape (UWC) students; we have no information on NSFAS funding for University of Cape Town (UCT) and Stellenbosch University (SU) students. Again, within CPUT and UWC, those with NSFAS cell phone numbers and or email addresses were more likely to respond. The second section of the table presents information on the type of email address on record for the graduate. Firstly, it is worth noting that a much larger share of SU and UWC respondents had an institutional email address than graduates from UCT and especially CPUT. This is partly a

function of the fact that SU and UWC graduates can keep their email addresses active for life, while UCT and CPUT graduates (at the time) were only given a limit period during which the email address remained active. In all cases, except CPUT, those who responded were less likely to have an institutional email address than those who did not respond.

Next, we look at the characteristics of graduates in 2010. Tables 4.1 to 4.3 present the average characteristics of responders versus non-responders as per their student records at the time of graduation. Students who responded were more likely to be female, younger and African and less likely to be white and foreign. They were also more likely to have had a bursary, either an NSFAS one or another type. In terms of their matric subjects, responders were more likely to have taken maths and science, but, of those who did, they were less likely to have achieved an A symbol. Responders were less likely to be qualifying with a postgraduate certificate or diploma or bachelor's degree and more likely to be honours or masters graduates, signalling that they were more advanced students. Responders also had significantly higher grade-point averages on completion of their qualifications. Many of the mean overall differences are significant, but the actual difference in the means is small. The significance is partly a function of the large sample size for those who did not respond.

The overall means hide some institutional differences. We point out some significant differences below:

Individual characteristics:

- UCT characteristics between responders and non-responders are more balanced than in other institutions, with CPUT being the least balanced.

Figure 3: Grade-point average (GPA) by institution and response status – bachelor qualifications only
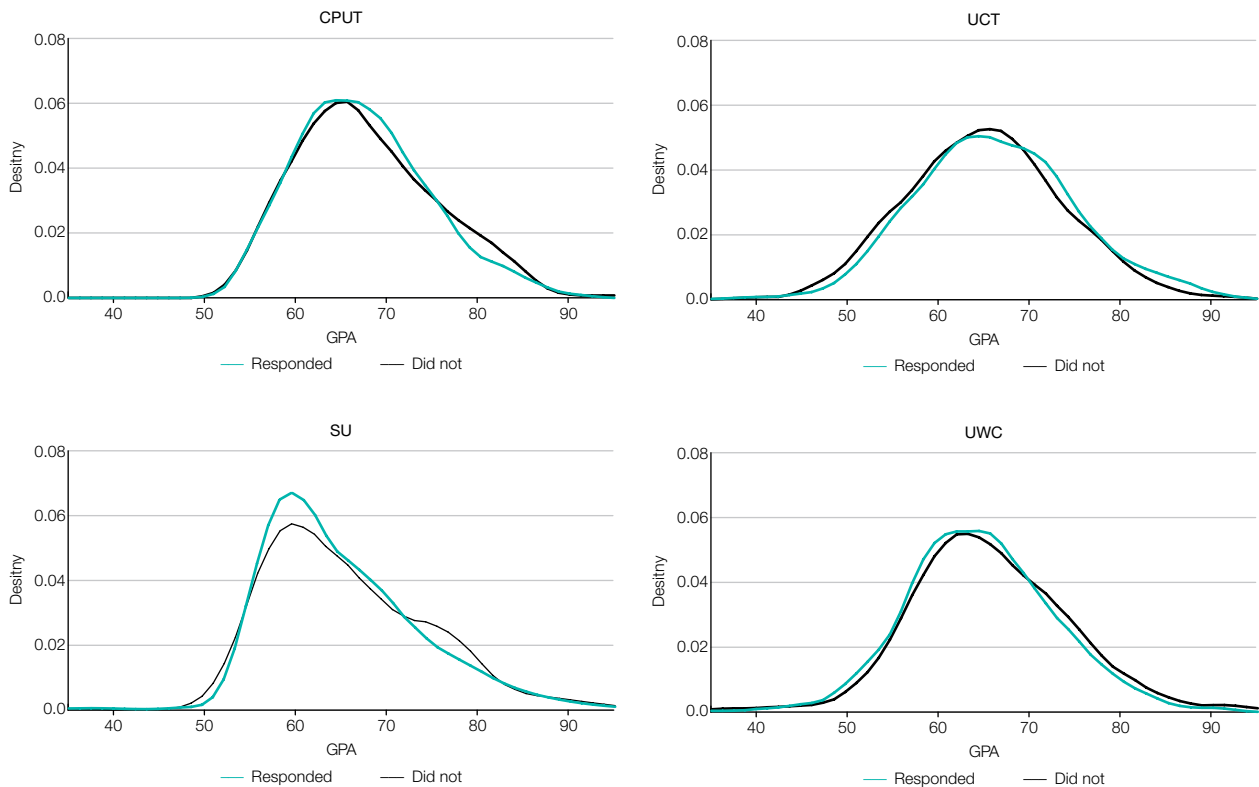
- There is a higher share of females in the responder group at each institution, but especially at CPUT and UWC.
- Responders are significantly more likely to be African (at all except UCT), especially at CPUT where the difference is nine percentage points.
- The use of NSFAS information for students at CPUT and UWC is evident from the data, with the share of NSFAS bursary holders being much higher in the responder group at these institutions.
- There is also a higher share of other bursary holders at all institutions. This could be a result of better contact details or it could signal a stronger connection to the institution.

Exit characteristics:

- These vary across institution by responder status.
- There is a higher share of masters students among UCT and SU responders; no similar difference is observed at CPUT and UWC. Other differences in qualification level across institutions between responders and non-responders indicate that, at UCT, SU and UWC,

responders are, on average, more advanced qualifying students, but are less highly qualifying students at CPUT.
- On the other hand, differences in grade-point average (GPA) are largest among UCT and UWC students. Figure 3 shows that those who responded within UCT have higher GPAs on average – in fact, the whole distribution is shifted to the right. For those at UWC, the average among responders is higher, but the distribution appears more dispersed.[2]
- Few programme differences are apparent, although science, engineering and technology (SET) students appear to overrepresented in the responder group at all institutions.

Another dimension worth noting is the differences in baseline characteristics between those who responded via the Web, those who responded telephonically or those who did not respond. Those who responded via the Web are much more similar

_____

2   Note that GPA information is only available for a subset of students, and this differs by institution. SU only has GPA information for bachelors and certificates/diplomas (excluding postgraduate), and CPUT and UCT for all, except doctorates and masters by coursework. UWC has GPA information across the whole range of qualifications.

to the non-responders than those who responded telephonically (results not shown).

It is not clear from these mean differences between responders and non-responders whether the WCGDS sample would be biased for employment analyses. While some characteristics might suggest lower probabilities of employment among responders – for example, female and African – other characteristic differences suggest higher probabilities of employment – higher qualifications and higher GPAs. It is clear, however, that differences between responders and non-responders vary by institution. Next, we look at the probability of not responding using a multivariate framework.

b. Non-response probit analysis

Table 5 examines baseline characteristics in a multivariate framework. Here, we look at the characteristics associated with not responding to the study – thus the dependent variable is an indicator that the graduate did not respond to the survey and that the coefficients presented are marginal effects from a probit regression. Age, sex and population group are significant predictors of response in this multivariate framework. Whites are 4.6 percentage points less likely to respond than Africans in the pooled model, and foreigners are 3.4 percentage points less likely to respond. Controlling for all characteristics, UCT and SU students are about 15 percentage points less likely to respond than CPUT students, and UWC students are nine percentage points less likely to respond. Availability and type of contact details are strong predictors of response, even after conditioning for multiple characteristics. Those who have cell phone numbers are 20 percentage points more likely to respond. Those who have an institutional email address are 6.8 percentage points less likely to respond, and this coefficient is significant at the one per cent level. We use this variable below as an exclusion restriction in our selection-correction model.

After conditioning on multiple covariates, only those qualifying with a master's degree have

significantly different response rates. Those qualifying with a masters are five percentage points more likely to respond than those with a certificate/diploma qualification. Educational subject matter (classified according to the Classification of Education Subject Matter – CESM) is predictive of response rates in the multivariate framework. All fields are less likely to respond than SET, with those in the education field are eight percentage points less likely to respond than the SET graduates. Finally, GPA is significantly related to response, with respondents with higher GPAs more likely respond.

The chi-square statistic and p-value at the base of the table show that these variables are jointly statistically different from zero at the highest level and therefore suggest that non-response is not random.

Examining the results from similar regressions run for each institution separately, we find some institutional differences. While white graduates and those with less complete contact details from all institutions are less likely to respond, being male is correlated with not responding only at CPUT and UWC, while, at UCT and SU, there is no relationship. The significant positive coefficient on being foreign seen in the overall regressions is driven by a large and significant relationship between being foreign and not responding among SU graduates. No similar relationship is observed for graduates from the other institutions. Having a bursary is positively related with response across all institutions, but only significantly so for UWC and SU. For qualification type, those with certificates or diplomas (including postgraduate diplomas) are less likely to respond at UCT, and those with masters at SU are less likely to respond. There are no additional significant relationships between qualification type and response once other variables are conditioned on.

Together, these factors suggest that those who are more connected to their original institution are more likely to respond. There are some differences across the CESM. While SET graduates appear to be the most likely to respond, education graduates at SU and UWC are eight percentage

points less likely to respond than SET graduates, and UCT health sciences graduates are similarly eight percentage points less likely to respond. Finally, it is worth noting that the relationship between responding and having an institutional email address is much stronger for UCT, UWC and CPUT than at SU. While UCT, UWC and CPUT graduates with an institutional email address are respectively 9.3, 8.5 and 7.6 percentage points less likely to respond than those with other email addresses, the coefficient at SU is only 0.039 (i.e. 3.9 percentage points). The relationship is, however, significant at the one per cent significance level at all institutions.

The chi-square test statistics show that response is non-random for all institutions. One way to adjust for this non-random non-response would be to construct a weight equal to the inverse probability of not responding from the non-response regressions above, and use these in further estimation equations (Falaris, 2003). However, the R-squares in Table 5 are small, especially for UCT and SU, suggesting that, while there is bias as measured on observable baseline characteristics, the impact of a reweighting exercise on these observable characteristics to correct for this bias is going to be small. In addition, this approach would not preclude there being other unobserved or unmeasured characteristics that could bias the results. For example, we do not have baseline information related to labour market prospects for the complete sample. It is very likely that, already on graduating, there are baseline differences in the propensity to be employed. Thus, if non-response was completely explained by these observable characteristics and the model was correctly specified, it would be possible to estimate Model 1, ignoring Model 2. If there are unobservable characteristics, this is no longer possible.

c. Using external data

In this section, we use external data to further assess the extent of non-response bias. We requested the Department of Higher Education and Training (DHET) to link the 2010 graduate database to the Higher Education Management Information

System (HEMIS) records for 2012, and the Department provided us with information on which students were studying in 2012.[3] Given that this is de-identified data, we cannot, however, match this information back into our full database.

Table 6 compares the proportion of 2010 graduates from CPUT, UCT and SU in the HEMIS 2012 database with the proportion of WCGDS respondents who reported that they were studying in 2012. The table shows that, according to the HEMIS database, between 21 and 23% of 2010 graduates were studying in 2012. Column 4 shows the number of WCGDS respondents in the HEMIS database. These numbers are relatively similar to the number of responders who stated in the questionnaire that they were studying in 2012, evidence that the matching worked well. The final column of the table shows the proportion of WCGDS responders who signalled that they were studying. These percentages are higher than in the overall graduate sample as represented by the HEMIS 2012 data – six percentage points higher for CPUT and UCT and 11 percentage points higher for SU. In addition, among students who were studying, responders were much more likely to be studying at their original institution, while a higher share of non-responders were studying at the University of South Africa (Unisa). Finally, responders were more likely to be studying towards a master's degree and less likely to be studying towards a higher certificate or diploma than non-responders. These findings are consistent with the understanding that contact details on the sampling frame master list were constructed from institutional databases at the time of the survey design. Students who continued to study would have more up-to-date contact details and therefore be more likely to respond.

The Cape Higher Education Consortium (CHEC) report notes the high rate of continued study as a key finding (21% immediately continued to study after graduating in 2010 (CHEC 2013: 36), while 31% were studying on 1 September 2012 (CHEC

---

3   It was unable to match UWC graduates to its database owing to invalid identity (ID) numbers; therefore, the analysis proceeds to consider CPUT, UCT and UWC graduates only.

2013: 73)). According to the graduate destination survey (GDS):

> The continuing higher education ratios in the four universities of the Western Cape are high by international comparisons. For example, in Schomburg and Teichler's 2006 graduate destination survey of 12 country cohorts, the continuing higher education of the cohorts investigated varied from 20% in France to 4% in the Czech Republic. As Table 11.9 suggests, the continuing higher education mean of 31% for the four institutions in the Western Cape is excellent by any measure (CHEC 2013: 79).

Table 6 suggests that these numbers are inflated and that the Western Cape continuing higher education ratios are closer to those observed in France.

For UCT, we could match UCT's 2010 graduates to the internal 2012 UCT database, and, therefore, we have all the baseline and WCGDS data in addition to whether students were studying in 2012. The UCT case study illustrates the potential of using external sources to validate information in a graduate destination-type study.

For the UCT subsample, we have information at the individual level (from the UCT administrative database) on who was enrolled at UCT in 2012. Using this information, we present a BGLW (Becketti, Gould, Lillard, Welch) test for the impact of non-response on the probability of studying in 2012. We regress studying status in 2012 on baseline characteristics, an indicator that the student did not respond to the survey, and the interaction of this indicator with baseline characteristics. The logic of the model is to determine whether the relationship between key explanatory variables and the outcome variable differs for responders versus non-responders. Statistically significant interactions between covariates and the non-response indicator indicate that the relationship between the covariate and studying status differs for non-responders. Table 7 shows that, while many of the covariates are significantly related to the probability of studying, only two show

a significantly different relationship statistically between responders and non-responders. These are age and 'other bursary' – the relationship between age and the probability of studying in 2012 is weaker (–0.006 + 0.004) in the non-responder group than in the responder group, and, similarly, the relationship between being a non-NSFAS-funded bursary holder and the probability of studying is weaker (0.109 – 0.074) in the non-responder group. While bursary holders are 10.9 percentage points more likely to be studying among graduates who responded to the WCGDS, bursary holders are only 3.5 percentage points more likely to be studying within the non-responder group. The F-test tests the joint significance of all the attrition terms, and the small p-value indicates that the null hypothesis is rejected – the relationship between the covariates jointly and the probability of studying differs between responders and non-responders.

In this section, we have shown that the composition of the responder sample differs from that of the non-responder sample in terms of the percentage studying and the percentage studying at their previous institution. The UCT-specific analysis shows that, in addition to the compositional difference, the relationship between the determinants of studying and the probability of studying differs for those who responded versus those who did not respond, suggesting that the characteristics of WCGDS responders who were studying in 2012 were not representative of the full studying population. In addition, the covariates included only explain 10% of the probability of studying, suggesting that there are other unobserved characteristics that explain the probability of studying, which, in turn, could also differ by responder status.

What does this mean in terms of bias for employment estimates? Employment rates are calculated as the share of the employed out of the labour force, and the labour force excludes those currently in education. Thus, while a higher share of students studying means a smaller share in the labour force, this does not necessarily mean that those in the labour force are biased in any particular way. On the other hand, if the

differences in characteristics of those studying between responders and non-responders reflect differences in their outside-of-study options – for example, those studying in the one group are studying because they have a lower employment – then the analysis above could suggest a difference in employment propensity. In the next section, we test this using a Heckman selection analysis.

d. Non-response, selection-corrected employment functions

One approach that has been used to attempt to control for unobservable characteristics related to response, is a Heckman selection-correction model (Heckman 1979; Maddala 1986). This approach requires an exclusion restriction, a factor that is correlated with response but not correlated with in Equation (1) – a variable that is not easy to come by. Authors have used quality-of-interview variables as instruments. For example, Maluccio (2004) uses, as instruments, first-round, survey-completion rates and whether the baseline survey was verified. In this vein, we argue that it is possible, conditional on many of the characteristics – especially the CESM – described above, that the institutional email indicator is an appropriate exclusion restriction. Our relationship of interest is employment in 2012. We argue that institutional email will not be related to employment probability when the sample is restricted to the labour force,[4] except possibly through some of the characteristics included in the structural equation (e.g. those with SET qualifications might be less likely to have institutional emails given their interest in technology. However, we control for the CESM in the employment equation).

Table 8 presents the lambda estimates (inverse Mills ratios) from a Heckman selection model of employment using institutional email as the exclusion restriction. The sample is restricted to males in the labour force.[5] The lambda coefficient is negative and insignificant for UCT, SU and UWC, and positive and insignificant for CPUT (note that there is no relationship between institutional email and response at CPUT). The lack of significance suggests that, conditional on the assumptions of the model, selection does not appear to be a significant problem for this model of employment. As a result, there is minimal change in the coefficient estimates (not shown) once 'selection' is accounted for in this way. Ignoring significance, the direction of the lambda coefficients suggests that UCT, SU and UWC graduates who responded to the survey were less likely to be employed, while CPUT graduates were more likely to be employed than those who did not respond to the survey. If the assumptions of this model hold, the results in Table 8, Column 2 present a better estimate of the relationship between employment and graduation characteristics. However, the accuracy of the selection-correction model is strongly dependent on the exclusion criteria. Information about the quality of the interview process at the individual level can provide useful instruments (e.g. Maluccio used whether the baseline survey was verified) and should therefore be collected.

---

4   The analysis in part c shows that institutional email is related to the probability of studying. However, by restricting the sample to those in the labour force we exclude this group.

5   The analysis is restricted to males to avoid the added complication of accounting for female labour market participation decisions.

6. CONCLUSIONS

- Destination studies have a particular type of bias that is inherent to their design and is often overlooked.
- The characteristics of responders and non-responders differ in non-random ways, and this is clear when looking at observable or measureable characteristics.
- The direction of the bias (on employment outcomes) that may result from these observable differences between responders and non-responders is difficult to identify in the Western Cape Graduate Destination Survey (WCGDS) data.
- Weighting on observable predictors of non-response only, accounts for these differences.
- There is a strong likelihood that responders/non-responders also differ in unobservable ways in respect of which it is not possible to adjust with statistical weights.
- Fortunately, there are two useful approaches in the literature to assess representation in terms of unobservables:
    a. Cross-checking against external (administrative) data; and
    b. An approach that allows for some type of control for selection based on unobservable characteristics through the use of a credible instrument.

- Cross-checking the WCGDS against Higher Education Management Information System (HEMIS) data shows that responders are much more likely to be studying than is evident in the total graduate population.
- This means that these data should not be used to estimate levels of continued study or labour market participation.
- Using institutional email as an instrument to correct for selection in an employment equation suggests that selection bias does not appear to be a severe problem for analyses of employment outcomes in the WCGDS data. We therefore have some confidence in the estimates of employment probability from these data.
- However, these findings only apply to employment outcomes (specifically restricted to those in the labour force) and the approach described in this paper would need to be used again for other outcomes (e.g. job satisfaction, job-matching, or studying) of interest and would require appropriate instruments to be found in each case.

7. RECOMMENDATIONS FOR A NATIONAL GRADUATE DESTINATION SURVEY

Plans for a National Graduate Destination Survey (NGDS) are currently being discussed. A study of this size is costly and will be vulnerable to high non-response. One key dimension is the preparation of the initial sampling frame and that good measures be put in place to follow graduates. Given that labour market success is a key question that the NGDS would want to investigate, one of the chief concerns is that those lost to follow-up are different in terms of their labour market outcomes. This study has assessed the extent of this concern in the Western Cape Graduate Destination Survey (WCGDS), and, in doing so, highlights potential solutions for a more successful study.

a. Response rates are strongly related to initial contact information

Response rates are related to the completeness of contact information. Quality of contact details varied significantly by institution. It would therefore be useful to standardise information across institutions and collect multiple contacts for students. The use of National Student Financial Aid Scheme (NSFAS) information has illustrated the benefit of obtaining contact information from multiple sources. This should be utilised across the board.

b. Use a better-prepared and consistent sampling frame with more comprehensive baseline information

Planning ahead and investing in the information collected at baseline will result in more useful information that can be used to assess selection bias. Including a short baseline survey that collects information on post-studying plans, possibly as part of the graduation process, and compiling master lists from all available institutional databases at the time of graduate exit, would be highly beneficial.

Notifying students, while they are studying, of the intended survey and its importance could also improve response. Collecting information on home postal code or other longer-term socio-economic markers (e.g. parental education) would also aid analysis.

A recommendation, arising from the present analysis, for future studies of this nature is to collect as much information as possible on the survey process itself so that an appropriate range of variables can be considered as controls for selection bias.

c. A well-designed sample can focus resources and improve response rates

Attempting to contact all graduates in a national project is an immense task. The use of a well-designed sample would be as successful in collecting the information required and would allow the focus of resources to be directed towards improving the response of the sample chosen rather than attempting to contact all graduates. The Eastern Cape graduate tracer study is a good example of this, where response rates of 47% were achieved among Rhodes University students and 37% among students from the University of Fort Hare (Rogan et al. 2015).

d. Linking/triangulating data and findings with other administrative data can be useful for assessing bias in key estimates

Utilising other existing administrative databases (Education Management Information Systems [EMIS], Higher Education Management Information System [HEMIS], Unemployment Insurance Fund [UIF], tax and Census) can be useful for three

reasons: it can limit the number of questions asked in the survey, for example about school, school neighbourhood, etc.; it can be used to increase the information available at baseline; and it can be used to cross-validate information collected in the survey. It would be worth including a request for more generalised data linkage from students while studying so as to avoid ethical concerns around linking different data sources.

e. Record information about the survey process at an individual level

Many of the techniques available to assess non-response bias rely on information about the survey process. In a selection-on-observables approach, we assume that all influential predictors of non-response have been included in the non-response probit. If there is enough information at the individual level, more of the non-response probability variation will be explained by the covariates, and the data can therefore be reweighted to better reflect the population. Measures that characterise the survey process have been found to be important determinants in regressions for the probability of not responding (Falaris & Peters 1998; Hill &

Willis 2001). Similarly, for the Heckman selection model, we require an instrument that is a significant determinant of response but which is uncorrelated with the outcome(s) of interest. Information about the quality of the baseline information or interview process (at the individual level) could be used for this. One simple inclusion in a Web-based survey could be information on whether the graduate opened the email, so as to differentiate those who did not receive the survey from those who decided not to complete it. Therefore, we advise that as much unit-level information on the interview process as possible be kept.

f. Appropriateness of questionnaire design

While the present analysis did not focus on the contents of the WCGDS questionnaire, it should be noted that it is not an appropriate instrument for labour market analyses. A more comprehensive labour market module, including income, would be a useful tool for the current policy discussions. In addition, in using an unfolding module design, care should be taken that all responders answer the same core questions.

REFERENCES

Bhorat, H (2004) Labour market challenges in the post-apartheid South Africa. *South African Journal of Economics* 72(5): 940–977

CHEC (Cape Higher Education Consortium) (2012) Western Cape Graduate Destination Survey data

CHEC (Cape Higher Education Consortium) (2013) Pathways from University to Work: A Graduate Destination Survey of the 2010 Cohort of Graduates from the Western Cape. A Cape Higher Education Consortium (CHEC) Study

DHET (Department of Higher Education and Training) (2010, 2012) Higher Education Management Information System (HEMIS) 2010, 2012

Du Toit, J (2016) Can we augment Web responses with telephonic responses to a graduate destination survey? *Assessment & Evaluation in Higher Education*, 41(4)

Du Toit, J, Kraak, A, Favish, J & Fletcher, L (2014) From study to work: Methodological challenges of a graduate destination survey in the Western Cape, South Africa. *Assessment & Evaluation in Higher Education* 39(7): 853–864

Falaris, EM (2003) The effect of survey attrition in longitudinal surveys: Evidence from Peru, Côte d'Ivoire and Vietnam. *Journal of Development Economics* 70(1): 133–157

Falaris, EM & Peters, HE (1998) Survey attrition and schooling choices. *Journal of Human Resources* 33:531–554

Heckman, J (1979). Sample selection bias as a specification error. *Econometrica* 47:153–161

Higher Education South Africa (2014) Proposal for a National Graduate Destination Survey (NGDS). Annexure 1 from consultative meeting for HESA's National Graduate Destination Survey (NGDS). Johannesburg, October

Hill, DH & Willis, RJ (2001) Reducing panel attrition: A search for effective policy instruments. *Journal of Human Resources* 36: 416–438

Maddala, GS (1986) *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press

Maluccio, J (2004) Using quality of interview information to assess nonrandom attrition bias in developing-country panel data. *Review of Development Economics* 8(1): 91–109

Rogan, M, Reynolds, J, Du Plessis, U, Bally, R & Whitfield, K (2015) Pathways through University and into the Labour Market. Report on a Graduate Tracer Study from the Eastern Cape. *LMIP Report 18*

Statistics South Africa. (2014) *General Household Survey*. Pretoria: Stats SA

Van Der Berg, S & Van Broekhuizen, H (2012) Graduate unemployment in South Africa: A much exaggerated problem. Stellenbosch University Economic Working Papers: 22/12, Department of Economics, University of Stellenbosch, December

TABLES

Table 1: Response rates – overall and by institution

Campus	Mean	N
CPUT	0.218	7 441
UCT	0.219	6 165
SU	0.216	7 380
UWC	0.269	3 724
Total	0.225	24 710

Notes: Table 1 shows the percentage of Western Cape 2010 graduates who responded to the Western Cape Graduate Destination Survey (WCGDS).

Table 2: Distribution of graduate qualification types by institution

	Institution				
Qualification type	CPUT	UCT	SU	UWC	Total
Certificate/diploma	61.36	4.67	1.48	11.12	21.76
Postgraduate certificate	0	14.08	19.65	8.59	10.68
Bachelors	35.79	48.26	44.09	53.14	43.99
Honours	1.42	14.03	17.25	16.86	11.62
Masters	1.29	16.37	15.18	8.73	10.32
Doctorate	0.15	2.6	2.36	1.56	1.63
Total	100	100	100	100	100

	All			CPUT			UCT			SU			UWC		
	Non-responder	Responder		Non-responder	Responder		Non-responder	Responder		Non-responder	Responder		Non-responder	Responder	
Contact details available:															
Cell	0.66	0.88	***	0.30	0.68	***	0.81	0.92	***	0.90	0.99	***	0.67	0.96	***
Landline	0.57	0.55		0.58	0.47	***	0.74	0.76		0.33	0.39	***	0.74	0.68	***
Email	0.71	0.77	***	0.14	0.23	***	0.87	0.96	***	1.00	1.00		1.00	1.00	
NSFAS cell	0.13	0.17	***	0.31	0.40	***							0.23	0.29	***
NSFAS email	0.12	0.17	***	0.29	0.41	***							0.24	0.29	***
Email type:0.01															
CPUT	0.01	0.01		0.02	0.03	*	0.00	0.00		0.00	0.00		0.00	0.00	
UCT	0.04	0.03	***	0.00	0.00		0.17	0.12	***	0.00	0.00		0.00	0.00	
UWC	0.11	0.12	**	0.00	0.00		0.00	0.00		0.00	0.00		0.78	0.68	***
SUN	0.17	0.14	***	0.00	0.00		0.00	0.01		0.56	0.47	***	0.00	0.00	
Institutional	0.34	0.30	***	0.02	0.00	*	0.13	0.13	***	0.57	0.47	***	0.79	0.68	***
Sample size	19 150	5 560		19 150	5 560		19 150	5 560		19 150	5 560		19 150	5 560	

Notes: Table 3 presents the proportion of graduates with different types of contact details by response status and institution in 2010. Responders are those 2010 graduates who responded to the WCGDS; non-responders are those who did not. Statistical differences between these means are indicated with * p<0.05; ** p<0.01; *** p<0.001. Institutional email type indicates that the graduate had an institutional email. Means presented are not weighted.

	All			CPUT only			UCT only			SU only			UWC only		
	Non-responder	Responder		Non-responder	Responder		Non-responder	Responder		Non-responder	Responder		Non-responder	Responder	
Personal characteristics:															
Male	0.57	0.55	**	0.58	0.55	*	0.56	0.53		0.57	0.55		0.61	0.57	**
Age in 2012	29.40	29.13	**	28.93	27.94	***	28.23	28.10		29.86	30.71	***	31.28	29.66	***
Population group (%):															
African	0.30	0.36	***	0.44	0.53	***	0.25	0.25		0.16	0.21	***	0.38	0.44	***
Coloured	0.27	0.26		0.35	0.34		0.17	0.16		0.17	0.15	**	0.48	0.45	
Indian	0.04	0.03		0.01	0.01		0.08	0.07		0.01	0.01		0.08	0.05	**
White	0.39	0.34	***	0.19	0.12	***	0.50	0.52		0.65	0.62	**	0.05	0.05	
Foreign-born	0.12	0.10	***	0.08	0.06	**	0.19	0.20		0.09	0.07	**	0.13	0.09	**
NSFAS bursary	0.11	0.14	***	0.18	0.25	***	0.09	0.09		0.03	0.03		0.16	0.11	**
Other bursaries	0.23	0.07	***	0.07	0.09	***	0.20	0.20	**	0.42	0.37	***	0.32	0.35	*

*Notes: Table 4.1 presents baseline demographic characteristics and whether the graduate was on a National Student Financial Aid Scheme (NSFAS) or other type of bursary, by response status and institution in 2010. Responders are those 2010 graduates who responded to the WCGDS; non-responders are those who did not. Statistical differences between these means are indicated with * p<0.05; ** p<0.01; *** p<0.001. Means presented are not weighted.*

Table 4.2: Baseline characteristics by response – matric results

	All			CPUT only			UCT only			SU only			UWC only		
	Non-responder	Responder		Non-responder	Responder		Non-responder	Responder		Non-responder	Responder		Non-responder	Responder	
Level of months taken (%):															
No maths	0.36	0.36	**	0.36	0.31	***	0.36	0.36		0.32	0.33		0.43	0.36	***
HG	0.33	0.34	*	0.13	0.14		0.51	0.53		0.46	0.50	**	0.16	0.16	
SG	0.31	0.31		0.49	0.53	**	0.13	0.10	*	0.22	0.17	***	0.40	0.46	**
LG	0.01	0.01	*	0.01	0.02		0.00	0.00		0.00	0.00		0.01	0.02	*
Other	0.00	0.00		0.00	0.00		0.00	0.00		0.00	0.00		0.00	0.00	
Mathematics symbol (%):															
A	0.26	0.24	**	0.08	0.07		0.43	0.39	*	0.35	0.36		0.09	0.09	
B	0.15	0.15		0.09	0.08		0.19	0.19		0.20	0.19		0.10	0.11	
C	0.18	0.18		0.17	0.16		0.18	0.20		0.19	0.20		0.16	0.15	
D	0.17	0.18		0.22	0.22		0.12	0.13		0.16	0.15		0.21	0.23	
E	0.15	0.16	**	0.25	0.25		0.16	0.07		0.08	0.08		0.24	0.25	
F+	0.09	0.10		0.19	0.21		0.01	0.01		0.02	0.01		0.20	0.18	
Level of science taken (%):															
No science	0.57	0.53	***	0.62	0.55	***	0.52	0.49	*	0.54	0.049	**	0.63	0.58	**
HG	0.30	0.32	**	0.14	.014	**	0.45	0.47		0.40	0.45	***	0.16	0.19	**
SG	0.13	0.15	***	0.24	0.30	***	0.03	0.03		0.06	0.06		0.21	0.23	
LG	0.00	0.00		0.00	0.01		0.00	0.00		0.00	0.00		0.00	0.00	
Science symbol (%):															
A	0.20	0.18	**	0.02	0.02		0.36	0.32	*	0.26	0.27		0.03	0.03	
B	0.16	0.15		0.06	0.04		0.23	0.24		0.21	0.22		0.06	0.06	
C	0.20	0.20		0.15	0.16		0.22	0.23		0.24	0.24		0.16	0.16	
D	0.21	0.19		0.28	0.24	*	0.13	0.14		0.19	0.19		0.28	0.28	
E	0.19	0.19		0.37	0.39		0.06	0.07		0.09	0.08		0.34	0.34	
F+	0.05	0.07	**	0.12	0.16	**	0.01	0.01		0.01	0.01		0.13	0.13	

Notes: Table 4.2 presents the matriculation mathematics and science marks of 2010 graduates by response status and institution in 2010. Responders are those 2010 graduates who responded to the WCGDS; non-responders are those who did not. Statistical differences between these means are indicated with * p<0.05; ** p<0.01; *** p<0.001. Means presented are not weighted.

Table 4.3: Baseline characteristics – qualification level and type

	All			CPUT only			UCT only			SU only			UWC only		
	Non-responder	Responder		Non-responder	Responder		Non-responder	Responder		Non-responder	Responder		Non-responder	Responder	
University graduation results:															
Grade-point average	63.86	64.32	***	64.16	63.97		64.52	65.72	***	62.60	63.04		63.37	63.98	***
Qualification type:															
Certificate/diploma	0.22	0.22		0.60	0.67	***	0.05	0.02	***	0.01	0.02		0.12	0.09	**
PG certificate/diploma	0.11	0.09	***	0.00	0.00		0.15	0.12	**	0.21	0.16	***	0.19	0.07	
Bachelor	0.44	0.42	**	0.37	0.32	***	0.48	0.48		0.45	0.41	**	0.52	0.55	
Honours	0.11	0.12	**	0.02	0.00	***	0.14	0.15		0.17	0.18		0.16	0.19	**
Master	0.10	0.12	***	0.01	0.01	*	0.15	0.20	***	0.14	0.20	***	0.09	0.09	
Doctorate	0.02	0.02		0.00	0.00	*	0.02	0.02	*	0.02	0.03	*	0.02	0.01	***
Classification of Education Subject Matter (CESM):															
Science, eng and tech	0.25	0.30		0.29	0.33	*	0.25	0.32		0.24	0.29	*	0.18	0.24	**
Com and bus	0.29	0.26		0.36	0.36	*	0.29	0.28		0.28	0.24	*	0.14	0.14	**
Human and soc science	0.18	0.20		0.11	0.11	*	0.24	0.24		0.17	0.19	*	0.22	0.26	**
Health science	0.13	0.13		0.09	0.09	*	0.09	0.06		0.18	0.20	*	0.19	0.17	**
Law	0.04	0.03		0.00	0.00	*	0.05	0.04		0.04	0.03	*	0.09	0.09	**
Education	0.12	0.08		0.09	0.16	*	0.07	0.06		0.09	0.06	*	0.17	0.11	*
Sample size	19 150	5 560		5 816	1 625		4 817	1 348		5 788	1 592		2 729	995	

*Notes: Table 4.3 presents the grade-point average (GPA), qualification type and the Classification of Educational Subject Matter (CESM) of 2010 graduates, by response status and institution in 2010. Responders are those 2010 graduates who responded to the WCGDS; non-responders are those who did not. Statistical differences between these means are indicated with * p<0.05; ** p<0.01; *** p<0.001. Means presented are not weighted.*

Table 5: Non-response analysis: Estimating the correlates of not responding to the WCGDS

	(1) All	(2) CPUT	(3) UCT	(4) SU	(5) UWC
	Non-responder	Non-responder	Non-responder	Non-responder	Non-responder
Age in 2012	0.0009*** (0.003)	0.014*** (0.005)	–0.000 (0.006)	0.007 (0.005)	0.010 (0.007)
Quadratic: Age in 2012	–0.000*** (0.000)	–0.000** (0.000)	–0.000 (0.000)	–0.000* (0.000)	–0.000 (0.000)
Age missing	0.142*** (0.003)	0.199*** (0.026)	0.059 (0.102		0.259*** (0.051)
Male	0.015*** (0.006)	0.028*** (0.010)	–0.001 (0.011)	0.007 (0.010)	0.036** (0.015)
Populations group (ref: African)					
Coloured	0.008 (0.008)	–0.031** (0.013)	–0.001 (0.018)	0.078*** (0.014)	0.039*** (0.017)
Indian/Asian	0.037*** (0.014)	–0.031** (0.013)	–0.001 (0.018)	0.078*** (0.014)	0.039** (0.017)
White	0.046*** (0.008)	0.052*** (0.015)	0.042*** (*0.016)	0.075*** (0.017)	0.055* (0.031)
Foreign	0.034** (0.008)	–0.040 (0.084)	–0.045 (0.034)	0.073*** (0.015)	–0.138 (0.164)
NSFAS bursary	–0.006 (0.010)	0.003 (0.015)	–0.040* (0.021)	0.002 (0.027)	–0.08 (0.026)
Other bursary	–0.022*** (0.007)	–0.036* (0.020	–0.015 (0.014)	–0.031*** (0.011)	–0.024 (0.017)
Qualifications type (bachelors)					
Certificate/diploma	–0.002 (0.009)	–0.007 (0.010)	0.138*** (0.023)	0.047 (0.035)	0.032 (0.029)
Postgraduate certificate/diploma	0.015 (0.011)	0.035** (0.018)	–0.020 (0.066)	0.020 (0.066)	0.018 (0.029)
Honours degree	–0.007 (0.010)	0.058 (0.045)	0.025 (0.016)	–0.093 (0.072)	–0.014 (0.022)
Master degree	–0.051*** (0.013)	–0.158 (0.097)	–0.019 (0.020)	–0.140* (0.077)	–0.022 (0.030)
Phd	–0.31 (0.025)		–0.017 (0.047)	–0.114 (0.086)	0.068 (0.062)
CESM (ref: SET)					
Business and commerce	0.034*** (0.007)	0.020* (0.012)	0.037** (0.015)	0.057*** (0.012)	0.041* (0.023)
Human and social sciences	0.019** (0.008	–0.029* (0.017)	0.041*** (0.014)	0.039*** (0.014)	0.013 (0.021)
Health sciences	0.035*** (0.009)	0.027 (0.018)	0.079*** (0.018)	0.011 (0.018)	0.056** (0.022)
Law	0.047*** (0.013)		0.065*** (0.022)	0.060** (0.024)	0.040 (0.027)
Education	0.077*** (0.009)	0.054*** (0.017)	–0.003 (0.033)	0.081*** (0.017)	0.082*** (0.025)
GPA	–0.002*** (0.000)	–0.002*** (0.001)	–0.002*** (0.001)	–0.001 (0.001)	–0.002* (0.001)
GPA – missing	–0.163*** (0.007)		–0.198** (0.081)		–0.204 (0.162)
Have an NSFAS email contact	–0.053*** (0.007)	–0.061*** (0.014)			–0.016 (0.023)

	(1) All	(2) CPUT	(3) UCT	(4) SU	(5) UWC
	Non-responder	Non-responder	Non-responder	Non-responder	Non-responder
Have an institutional email	0.068*** (0.007)	0.076*** (0.022)	0.093*** (0.012)	0.039*** (0.010)	0.085*** (0.019)
Campus (ref: CPUT):					
UCT	0.145*** 0(0.010)				
SU	0.156*** 0(0.011)				
UWC	0.088***				
Observations	(0.011)				
Chi-square test	24 710	7 430	6 165	7 380	3 715
Degrees of Freedom	1 724	884.5	281	279.8	359.4
P-value	31	22	26	22	26
Pseudo R-square	0 0.0739	0 0.124	0 0.0492	0 0.0492	0 0.115

*Notes: Table 5 presents regression marginal effects from an attrition probit analysis in addition to a chi-square test of whether the variables in the model are jointly equal to zero, by institution type. Attritors are those 2010 graduates who did not respond to the WCGDS. Statistical differences between these means are indicated with * p<0.05; ** p<0.01; *** p<0.001. Regression results are not weighted.*

Table 6: Studying status in 2012 – comparing the Western Cape Graduate Destination Survey (WCGDS) rates with the Higher Education Management Information System (HEMIS) 2012 database

	All 2010 graduates			WCGDS responders		
	Number with ID numbers	In 2012 HEMIS database		In HEMIS 2012 database	Studying at university according to Q4_1	
	#	#	%	#	#	%
CPUT	6 894	1 450	21%	426	430	27%
UCT	5 102	1 143	22%	300	379	28%
SU	7 380	1 710	23%	492	533	34%
Total	19 376	4 303		1 218	1 342	

Notes: Graduates with valid ID numbers were linked to the HEMIS 2012 database. UWC ID records were invalid and therefore UWC is not included in this assessment. The table shows that between 21 to 23% of 2010 graduates were studying in 2012. Column 4 shows the number of WCGDS respondents in the HEMIS database. These numbers are relatively similar to the number of responders stating that they were studying in 2012 in the questionnaire suggesting that the matching worked. The final column of the table shows that the proportion of responders who signalled that they were studying is far higher among respondents in the WCGDS than in the overall graduate sample – at six percentage points higher for CPUT and UCT and 11 percentage points higher for SU. In addition, among students who were studying, responders were much more likely to be studying at their original institution, while non-responders had a higher share studying at the University of South Africa (Unisa). They are also more likely to be studying towards a master's degree and less likely to be studying towards a higher certificate or diploma.

Table 7: BGLW (Becketti, Gould, Lillard, Welch) test for the probability of studying in 2012 (UCT subsample only)

	Coefficient	Std error
Age	−0.006***	(0.002)
Male	0.012	(0.020)
Population group (ref: White):		
Not specified	0.075*	(0.043)
African	−0.023	(0.028)
Coloured	−0.006	(0.030)
Indian	−0.075*	(0.041)
Foreign	0.025	(0.053)
NSFAS bursary	−0.015	(0.036)
Other bursary	0.109***	(0.025)
Qualification type (ref: Bachelor):		
Certificate/diploma	0.053	(0.095)
Postgraduate certificate/diploma	−0.081**	(0.037)
Honours	0.051*	(0.031)
Masters	−0.096***	(0.031)
Doctorate	−0.086	(0.063)
Contact details available:		
Email	0.013	(0.051)
Cell	0.091**	(0.036)
Institutional email	0.142***	(0.030)
CESM (ref: Science, engineering and technology):		
Business and commerce	−0.162***	(0.027)
Human and social sciences	−0.113***	(0.027)
Health sciences	−0.174***	(0.043)
Law	−0.045	(0.053)
Education	−0.173***	(0.063)
Attrition	−0.134	(0.086)
Interactions"		
Att x Age	0.004*	(0.002)
Att x Male	−0.019	(0.023)
Att x Race not specified	−0.048	(0.049)
Att x Coloured	0.010	(0.032)
Att x Indian	−0.010	(0.034)
Att x Foreign	−0.005	(0.064)
Att x NSFAS bursary	−0.022	(0.040)
Att x Other bursary	−0.074***	(0.029)
Att x Certificate/diploma	−0.064	(0.102)
Att x Postgraduate certificate/diploma	0.017	(0.041)
Att x Honours	−0.002	(0.035)
Att x Masters	0.002	(0.035)
Att x Doctorate	0.007	(0.072)

	Coefficient	Std error
Att x Business and commerce	0.047	(0.031)
Att x Human and social sciences	0.046	(0.031)
Att x Health sciences	0.016	(0.048)
Att x Law	0.039	(0.059)
Att x Education	0.047	(0.070)
Att x Email	0.028	(0.053)
Att x Cell	–0.050	(0.038)
Att x Institutional email	–0.043	(0.033)
Constant	0.338***	(0.079)
Sample size	**6 165**	
R-squared	**0.098**	
Test of the joint significance of the attrition terms: **F(24.6 117)** **Prob > F**	**2.19** **0.0007**	

*Notes: Table 7 presents coefficients and standard errors for the probability of studying at UCT from a linear probability model. The attritor variable takes a value of one for those who did not respond to the WCGDS, and zero otherwise. The sample is restricted to 2010 UCT graduates. * $p<0.05$; ** $p<0.01$; *** $p<0.001$*

Table 8: Lambda and institutional email coefficients from a Heckman selection model of employment using institutional email as the exclusion criterion

	All	CPUT	UCT	SU	UWC
Lambda (inverse mills ratio)					
Coefficient	0.023	0.396	−0.108	−0.054	−0.006
Standard error	(0.031)	(0.650)	(0.140)	(0.065)	(0.024)
Institutional email					
Coefficient	−0.365***	−0.173***	−0.321***	−0.385***	−0.271***
Standard error	(0.039)	(0.165)	(0.087)	(0.056)	(0.089)

Notes: The employment regression was restricted to males in the labour force. Covariates included were age, race, NSFAS and other bursary, contact information available, qualification type, subject material and GPA.

Assessing the Usability of the Western Cape Graduate Destination Survey for the Analysis of Labour Market Outcomes

This report examines the response rate of one graduate destination study. The Western Cape Graduate Destination Survey in 2012 showed that only 22% of all 2010 university graduates from the four Western Cape universities were successfully interviewed. Graduate destination studies can provide information about how graduates transition into work. The low rate of response to these graduate destination studies raises the concern that the non-response is not random and that inferences using data on those who respond will be inaccurate. It was found that those who successfully responded to the survey were more likely to be studying in 2012. While those that responded have some systematically different baseline information that signals that response is not random, this has a limited impact for an equation of employment. Focus should be directed at preparing and standardising the sampling frame, keeping detailed records of the survey process, and linking graduate destination study data with administrative resources to assess bias and supplement the survey information obtained.

About the LMIP

The Labour Market Intelligence Partnership (LMIP) is a collaboration between the Department of Higher Education and Training, and a Human Sciences Research Council-led national research consortium. It aims to provide research to support the development of a credible institutional mechanism for skills plannning in South Africa. For further information and resources on skills planning and the South African post-school sector and labour market, visit http://www.lmip.org.za.