

Southern Africa Labour and Development Research Unit



Assessing the usability of the Western Cape
Graduate Destination Survey for the
analysis of labour market outcomes

by

Nicola Branson and Murray Leibbrandt

About the Author(s)

Nicola Branson is a researcher in the Southern Africa Labour Development Research Unit at the University of Cape Town (UCT). Email: nicola.branson@gmail.com

Murray Leibbrandt is a Professor in the School of Economics at the University of Cape Town and the Director of the Southern Africa Labour and Development Research Unit. He holds the DSD/NRF National Research Chair of Poverty and Inequality Research and is a Principal Investigator on the National Income Dynamics Study.

Acknowledgments

This research was conducted as part of the Labour Market Intelligence Partnership, research consortium headed by the Human Sciences Research Council (South Africa) and funded by the Department of Higher Education and Training (South Africa).

Recommended citation

Branson, N., Leibbrandt, M. (2017). Assessing the usability of the Western Cape Graduate Destination Survey for the analysis of labour market outcomes. A Southern Africa Labour and Development Research Unit Working Paper Number 198. Cape Town: SALDRU, University of Cape Town

ISBN: 978-1-928281-59-7

© Southern Africa Labour and Development Research Unit, UCT, 2017

Working Papers can be downloaded in Adobe Acrobat format from www.saldru.uct.ac.za.
Printed copies of Working Papers are available for R25.00 each plus vat and postage charges.

Orders may be directed to:

The Senior Administrative Officer, SALDRU, University of Cape Town, Private Bag, Rondebosch, 7701,
Tel: (021) 650 5696, Fax: (021) 650 5697, Email: tania.hendricks@uct.ac.za



Assessing the usability of the Western Cape Graduate Destination Survey for the analysis of labour market outcomes

Nicola Branson and Murray Leibbrandt

Saldru Working Paper 198

University of Cape Town

January 2017

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 nonresponse is not random and inferences using data on those who respond will be inaccurate. This study examines response rates in the Western Cape Graduate Destination Study 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, compared rates of continued study to those in the 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 are 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 study. We recommend that focus be taken in 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.

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 comes 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 GDS 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 surveys¹, has resulted in some studies reporting high levels of 'graduate unemployment' when those with diplomas/certificates that 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 and 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 representative of those in non-degree programs and variation across institutions left unexplored.

Another potential contributor to these differences stems from the low response rates that plague GDS. The design of a GDS makes it inherently vulnerable to selection bias. GDS 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 is quickly 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 WCGDS, run by CHEC, attempted to contact all 2010 graduates from the four WC 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 HEMIS database of all 2010 graduates from these institutions to statistically weight the successfully contacted graduates to the actual socio-demographic 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 i.e. 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.

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

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

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, bachelors degree, bachelors degree with diploma, honours degree or higher degree (master or doctorate) (GHS 2014).

The WCGDS was primarily designed to analyze 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 generalized 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.

Expanding the graduate discussions – the type of questions GDS 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 a large return to completing a post schooling qualification. Yet the information contained in household surveys is limited in two respects. First, given the small percentage of graduates in the national population (graduates account for 0.27 percent of the 2011 population (CHEC 2013)), nationally representative household studies do not achieve samples large enough to disaggregate by institution or study program. Second, the information included about graduates is limited and, for cross sectional data, only contains concurrent socioeconomic information. This makes it difficult to disentangle the impact of studying from preexisting 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 to ascertain not only the labour market and further studying trajectories of graduates but also perceived value of their qualification once working, relevance in the workplace and 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 allows investigation into the match between labour force participation and labour shortage areas. Figure 1 shows the employment rates of graduates in the WCGDS by qualification type and area of study. It shows variability in employment rates across study area, particularly among graduates with lower qualification.

The Western Cape Graduate Destination questionnaire is divided into five main sections – time during school, at university, employment experience, further studies and plans for the future – this type of 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 is 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 and 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 UCT and US and CPUT and 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 GDS studies aid us in unpacking many of the puzzles within the higher education sector that are often left untouched due to data limitations. On the other hand, the problem with graduate destination studies is 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.

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 to those who do not respond. In fact, given the large sampling frames generally available to utilize in the design of a GDS – 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.

$$y_i = \mathbf{x}'_i \boldsymbol{\beta}_1 + \varepsilon_i \quad (y_i \text{ observed only if } A_i^* < 0)$$

(1)

$$A_i^* = \mathbf{x}'_i \boldsymbol{\beta}_2 + \mathbf{z}'_i \boldsymbol{\gamma} + v_i \tag{2}$$

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 (1) (\mathbf{x}'_i) in addition to some additional ones (\mathbf{z}'_i). A_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 $\boldsymbol{\beta}_1$ 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, \mathbf{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 realized 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 - i.e. that make ε_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 WCGDS can therefore misguide users to think 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, p 103) points out selection “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 ε_i and v_i 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 (Heckman1979; Maddala, 1986).”

Data

We construct a database of information from the WCGDS with additional institutional and HEMIS records from 2010 and 2012. The WCGDS, run by CHEC, attempted to contact all 2010 graduates from the four WC 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 2014). 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 a response rate of 22.5% was achieved, with rates varying marginally across institutions: CPUT (21.8%), UCT (21.9%), SU (21.6%) and UWC (26.7%). Details of the survey process are available in Du Toit (2014). A few key things 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 on the institutional software. Thus any graduate who had continued their studies would have more up-to-date information.
- Contact details from the National Study Financial Aid Scheme (NSFAS) were only sought for student at CPUT and UWC as these institutions had a large share of students with incomplete contact details.
- A non-random telephone follow up was used to increase the response rates for graduates predominantly from CPUT and UWC. No information is available on who the call center attempted to call, only on the mode of interview (email or phone) 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 DHET and de-identified data 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 to the Census 2011. This provides us with more detailed information for some parts of the analysis.

Results section

In this section we attempt to assess whether the 22% response rate in the WCGDS poses a concern for estimates of employment. First 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 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 respond and those who do not, overall and by institution. It is clear that those with more complete contact details are more likely to respond. For example, 88% of responders have cell phone details compared to 66% of non-responders. Similarly, those who respond are more likely to have an email address. NSFAS information was only used for CPUT and UWC students, we have no information on NSFAS funding for UCT and US student. Again within CPUT and UWC, those with NSFAS cell phone numbers and or emails are more likely to respond. The second section of the table presents information on the type of email address on record for the graduate. First it is worth noting that a much larger share of US and UWC respondents have an institutional email than graduates from UCT and especially CPUT. This is partly a function of the fact that US and UWC graduates can keep their emails active for life while UCT and CPUT graduates (at the time) were only given a limit period where the email remained active. In all cases, except CPUT, those who respond are less likely to have an institutional email than those who do not respond.

Next we look at the characteristics of graduates in 2010. Table 4.1-4.3 presents the average characteristics of responders versus non-responders as per their student records at the time of graduation. Students who responded are more likely to be female, younger and African and less likely to be white and foreign. They are also more likely to have had a bursary, either a NSFAS one or another type. In terms of their matric subject, 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 bachelors 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, CPUT is the least balanced.
- 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 9% points.
- The use of NSFAS information for students at CPUT and UWC is evident in the data, with the share of NSFAS bursary holders 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 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 US responders, no similar difference is observed at CPUT and UWC. Other differences in qualification level across institutions between responders and non-responders indicate that responders are on average more advanced qualifying students at UCT, US and UWC but less highly qualifying students at CPUT.
- GPA differences on the other hand are largest among the 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 program differences are apparent although Science, Engineering and Technology (SET) students appear 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, via telephone or did not respond. Those who responded via the web are much more similar to the non-responders than those who responded via telephone (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 – e.g. female and African – other characteristic differences suggest higher probabilities of employment – higher qualifications and higher GPAs. It is also clear that differences between responders and non-responders differ by institution. Next we look at the probability of not responding in 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 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 US students are about 15 percentage points less likely to respond than CPUT students, and UWC students are 9 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 percent 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 Masters degree have significantly different response rates. Those qualifying with a Masters are 5 percentage points more likely to respond than those with a certificate/diploma qualification. Educational subject matter (CESM) is predictive of response rates in the multivariate framework. All fields are less likely to

² Note that GPA information is only available for a subset of students and this differs by institution. US only has GPA information for bachelors and cert/diplo (excluding PG) and CPUT and UCT for all except doctorates and master's by coursework. UWC has GPA information across the board of qualifications.

respond than SET, with those in the education field 8 percentage points less likely to respond than the SET graduates. Finally GPA is significantly related to response, with respondents with higher GPA 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 US 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 US 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 US. For qualification type, those with certificates or diplomas (including postgraduate diplomas) are less likely to respond at UCT, and those with masters at US 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 CESM. While SET graduates appear to be the most likely to respond, education graduates at US and UWC are 8 percentage points less likely to respond than SET graduates and UCT health sciences graduates are similarly 8 percentage points less likely to respond. Finally, it is worth noting that the relationship between responding and having an institutional email is much stronger for UCT, UWC and CPUT than at Stellenbosch. While UCT, UWC and CPUT graduates with an institutional email are 9.3, 8.5 and 7.6 percentage points less likely to respond than those with other emails, the coefficient at US is only 0.039 (i.e. 3.9 % points). The relationship is however significant at the one percent 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 US, 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 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 asked DHET to link the 2010 graduate database to the HEMIS records for 2012 and they provided us with

information on which students were studying in 2012³. 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 US in the HEMIS 2012 database to the proportion of WCGDS respondents who reported to be studying in 2012. The table shows that, according to the HEMIS database, between 21-23% 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 saying they were studying in 2012 in the questionnaire evidence that the matching worked well. The final column of the table shows the proportion of WCGDS responders that signaled they were studying. These percentages are higher than in the overall graduate sample as represented by the HEMIS 2012 data - 6 % points higher for CPUT and UCT and 11% points higher for US. In addition, among students who were studying, responders are much more likely to be studying at their original institution, while a higher share of non-responders are studying at the University of South Africa. Finally, responders are more likely to be studying towards a Masters 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 masterlist 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 CHEC report notes the high rate of continued studying as a key finding (21% immediately continued to study after graduating in 2010 p36, 31% were studying 1 September 2012 p73) from the 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 p79). 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 2010 graduates to the internal 2012 UCT database and therefore 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 of who was enrolled at UCT in 2012 from the UCT administrative database. Using this information we present a Beckett, Gould, Lillard and Welch (BGLW) 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 differ 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 statistically significantly different relationship 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 the responder group and similarly, the relationship between being a non-NSFAS funded bursary holder and the probability of studying is weaker (0.109-

³ They were unable to match UWC graduates to their database given invalid ID numbers; therefore the analysis proceeds to CPUT, UCT and UWC graduates only.

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 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 study probability differs for those who respond versus do not respond, suggesting that the characteristics of WCGDS responders who are studying in 2012 are 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 studying probability, 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 reflects 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 z_i' that is correlated with response but not correlated with ε_i 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 first round survey completion rates and whether the baseline survey was verified as instruments. In this vein, we argue that it is possible, conditional on many of the characteristics - especially 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 force⁴ 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 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 force⁵. The lambda coefficient is negative and insignificant for UCT, US and UWC and positive and insignificant for CPUT (note there is no relationship between institutional email and response at CPUT). The lack of significance suggests that, conditional on the assumptions of the

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.

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 that UCT, US 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. Collecting 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.

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 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 for 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:
 - Cross checking against external (administrative) data and
 - 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 HEMIS data shows that responders are much more likely to be studying than is evident in the total graduate population.
- This means 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 suggest 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 conducted 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.

Recommendations for a National Graduate Destination Survey

Plans for a National Graduate Destination Survey (NGDS) are currently being discussed. A study of this size bears large costs 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 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 NSFAS information has illustrated the benefit of obtaining contact information from multiple sources. This should be utilised across the board.

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

A recommendation arising from the 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 towards improving the response of the sample chosen rather than attempting to contact all graduates. The Eastern Cape Graduate Destination Study is a good example of this, where they achieved response rates of 47% among Rhodes Students and 37% among students from the University of Fort Hare (Rogan et al. 2015).

d. Linking/triangulating data and findings with other administration data – EMIS, HEMIS, UIF, tax, Census – can be useful for assessing bias in key estimates.

Utilising other existing administrative databases can be useful for three reasons. It can limit the number of questions asked in the survey e.g. 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 characterize the survey process have been found to be important determinants in regressions for the probability of not responding (Falaris and Peters 1998; Hill and Willis 2001). Similarly, for the

Heckman selection model we require an instrument that is a significant determinant of response but 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 to a web based survey could be information on whether the graduate opened the email; to differentiate those who did not receive the survey versus 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 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 (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.
- du Toit, J., Kraak, A., Favish J and 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.
- du Toit, J. (2016) Can we augment web responses with telephonic responses to a graduate destination survey? *Assessment & Evaluation in Higher Education* Vol. 41:4.
- Falaris, E.M., 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, E.M., Peters, H.E. (1998). Survey attrition and schooling choices. *Journal of Human Resources*, 33, 531–554.
- Statistics South Africa (2014). *General Household Survey*. Pretoria: Stats SA.
- Heckman, J. (1979). Sample Selection Bias as a Specification Error, *Econometrica* 47, 153–61.
- Department of Higher Educational and Training (DHET) 2010, 2012. Higher Education Management Information System (HEMIS) 2010, 2012.
- 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, D.H., Willis, R.J. (2001). Reducing panel attrition: a search for effective policy instruments. *Journal of Human Resources* 36, 416–438.
- Maddala, G. S. (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 and Whitfield, K. (2015) Pathways through University and into the Labour Market Report on a graduate tracer study from the Eastern Cape. *LMIP report 18*.
- Van Der Berg, S. and H. Van Broekhuizen (2012). Graduate unemployment in South Africa: A much exaggerated problem, Stellenbosch University Economic Working Papers: 22/12, Department of Economics, University of Stellenbosch, December.
- Centre for Higher Education Consortium (CHEC) (2012) Western Cape Graduate Destination Study data.

Tables and Figures

Table 1: Response rates – overall and by institution

campus	mean	N
CPUT	0.218	7441
UCT	0.219	6165
US	0.216	7380
UWC	0.267	3724
Total	0.225	24710

Notes to Table 1: Table 1 shows the percentage of Western Cape 2010 Graduates who responded to the Western Cape Graduate Study.

Figure 1: Employment by qualification type and subject

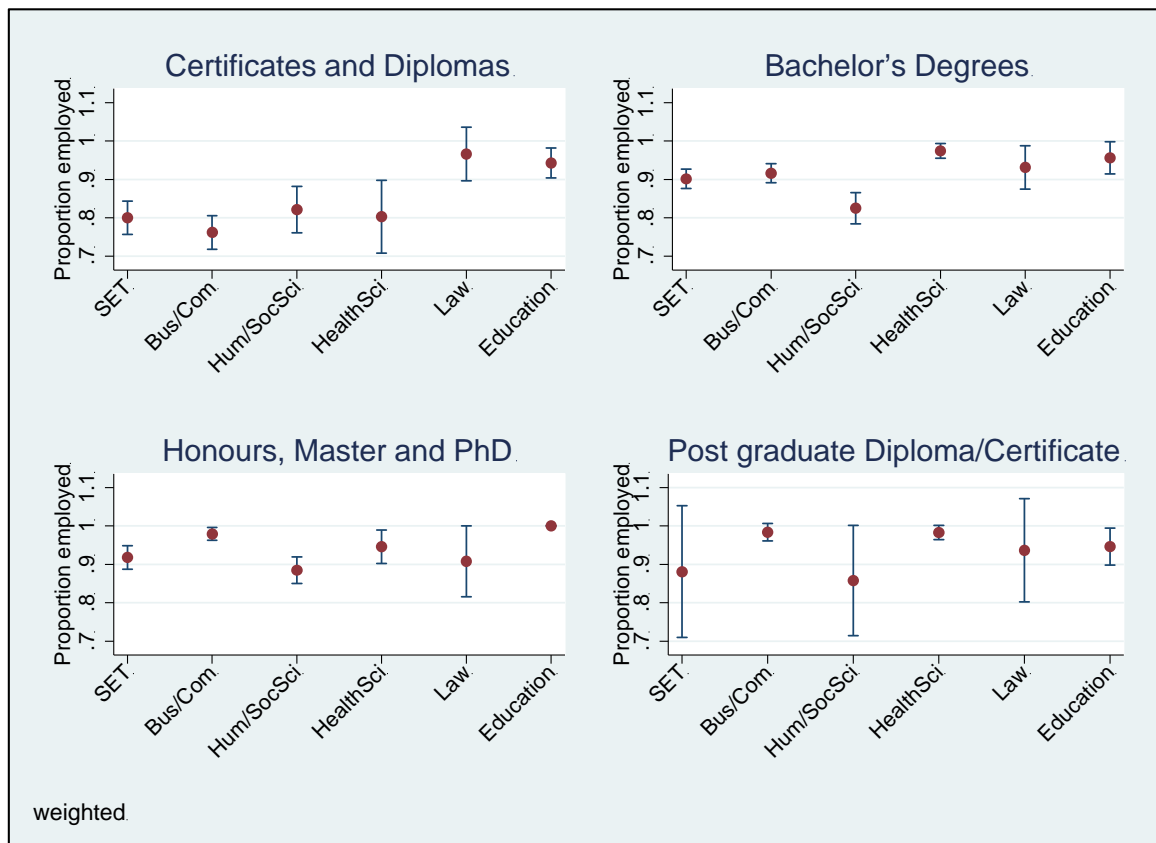


Figure 2: Employment by institution – all qualifications versus Bachelor qualifications only

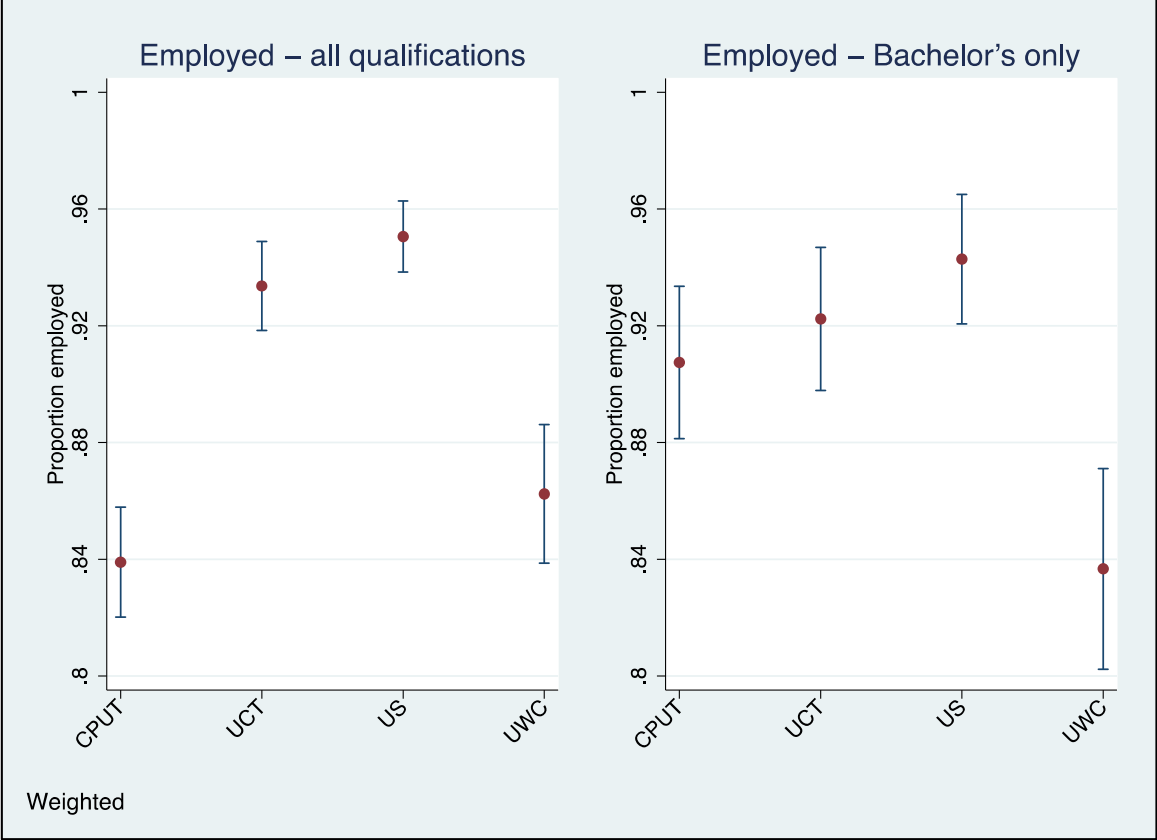


Table 2: Distribution of graduate qualification types by institution

Qualification type	Institution				Total
	CPUT	UCT	US	UWC	
Certificate/diploma	61.36	4.67	1.48	11.12	21.76
Postgraduate certific	0	14.08	19.65	8.59	10.68
Bachelor's	35.79	48.26	44.09	53.14	43.99
Honours	1.42	14.03	17.25	16.86	11.62
Master's	1.28	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

Table 3: Contact details – overall and by institution

	All		CPUT		UCT		US		UWC						
	Non Responder	Responder	Non Responder	Responder	Non Responder	Responder	Non Responder	Responder	Non Responder	Responder					
<i>Contact details available:</i>															
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	0.4	***							0.23	0.29	***
NSFAS email	0.12	0.17	***	0.29	0.41	***							0.24	0.29	***
<i>Email type:</i>															
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	0.3	***	0.02	0.03	*	0.17	0.13	***	0.57	0.47	***	0.79	0.68	***
<i>Sample size</i>	<i>19150</i>	<i>5560</i>		<i>19150</i>	<i>5560</i>		<i>19150</i>	<i>5560</i>		<i>19150</i>	<i>5560</i>		<i>19150</i>	<i>5560</i>	

Notes to Table 3: 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.

Table 4.1: Baseline characteristics by response – demographic and bursary

	All		CPUT only		UCT only		US only		UWC only						
	Non-Resp.	Responder	Non-Resp.	Responder	Non-Resp.	Responder	Non-Resp.	Responder	Non-Resp.	Responder					
<i>Personal characteristics:</i>															
Male	0.57	0.55	**	0.58	0.55	*	0.54	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	***	
<i>Population Group (%)</i>															
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	0.10	0.03	0.03		0.16	0.21	**	
Other bursary	0.23	0.27	***	0.07	0.09	***	0.20	0.24	**	0.37	0.42	***	0.32	0.35	*

Notes to Table 4.1: Table 4.1 presents baseline demographic characteristics and whether the graduate was on a 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			US only			UWC only		
	Non-Resp.	Responder		Non-Resp.	Responder		Non-Resp.	Responder		Non-Resp.	Responder		Non-Resp.	Responder	
Matric results:															
<i>Level of maths taken (%):</i>															
No maths	0.36	0.34	**	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	
<i>Mathematics symbol (%):</i>															
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		0.06	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	
<i>Level of science taken (%):</i>															
No science	0.57	0.53	***	0.62	0.55	***	0.52	0.49	*	0.54	0.49	**	0.63	0.58	**
HG	0.30	0.32	**	0.14	0.14		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	
<i>Science symbol (%):</i>															
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	0.05	
C	0.20	0.20		0.15	0.16		0.22	0.23		0.24	0.24		0.16	0.15	
D	0.21	0.21		0.28	0.24	*	0.13	0.14		0.19	0.19		0.28	0.30	
E	0.19	0.19		0.37	0.37		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 to Table 4.2: 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			US only		UWC only	
	Non-Resp.	Responder	Non-Resp.	Responder		Non-Resp.	Responder		Non-Resp.	Responder	Non-Resp.	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 **
<i>Qualification type</i>												
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 ***	0.09	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 **
Masters	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	0.03		0.02	0.03 *	0.02	0.01 **
<i>Classification of Educational Subject Matter (CESM)</i>												
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 sci	0.18	0.20	0.11	0.13 *		0.24	0.24		0.17	0.19 *	0.22	0.26 **
Health sciences	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	0.16	0.09 *		0.07	0.06		0.09	0.06 *	0.17	0.11 **
Sample size	19150	5560	5816	1625		4817	1348		5788	1592	2729	995

Notes to Table 4.3: Table 4.3 presents the GPA, qualification type and the 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.

Figure 3: GPA by institution and response status – bachelor qualifications only

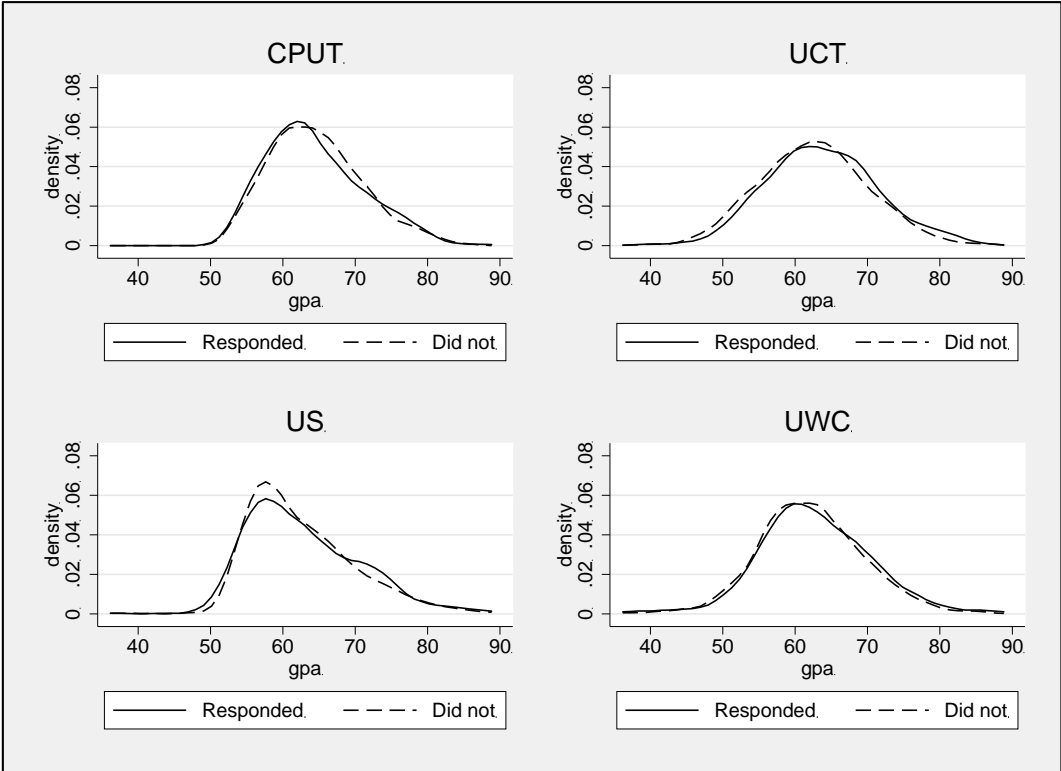


Table 5: Non-response analysis: Estimating the correlates of not responding to the WCGDS

	(1) All	(2) CPUT	(3) UCT	(4) US	(5) UWC		(1) All	(2) CPUT	(3) UCT	(4) US	(5) UWC
Age in 2012	0.009*** (0.003)	0.014*** (0.005)	-0.000 (0.006)	0.007 (0.005)	0.010 (0.007)	CESM (ref: SET)					
Quadratic: Age in 2012	-0.000*** (0.000)	-0.000** (0.000)	-0.000 (0.000)	-0.000* (0.000)	-0.000 (0.000)	Business and commerce	0.034*** (0.007)	0.020* (0.012)	0.037** (0.015)	0.057*** (0.012)	0.041* (0.023)
Age missing	0.142*** (0.028)	0.199*** (0.026)	0.059 (0.102)		0.259*** (0.051)	Human and social sciences	0.019** (0.008)	-0.029* (0.017)	0.041*** (0.014)	0.039*** (0.014)	0.013 (0.021)
Male	0.015*** (0.006)	0.028*** (0.010)	-0.001 (0.011)	0.007 (0.010)	0.036** (0.015)	Health sciences	0.035*** (0.009)	0.027 (0.018)	0.079*** (0.018)	0.011 (0.018)	0.056** (0.022)
Population group (ref: African)						Law	0.047*** (0.013)		0.065*** (0.022)	0.060** (0.024)	0.040 (0.027)
Coloured	0.008 (0.008)	-0.031** (0.013)	-0.001 (0.018)	0.078*** (0.014)	0.039** (0.017)	Education	0.077*** (0.009)	0.054*** (0.017)	-0.003 (0.033)	0.081*** (0.017)	0.082*** (0.025)
Indian/Asian	0.037*** (0.014)	-0.016 (0.043)	0.037* (0.021)	0.043 (0.037)	0.074*** (0.026)	GPA	-0.002*** (0.000)	-0.002*** (0.001)	-0.002*** (0.001)	-0.001 (0.001)	-0.002* (0.001)
White	0.046*** (0.008)	0.052*** (0.015)	0.042*** (0.016)	0.075*** (0.017)	0.055* (0.031)	GPA - missing	-0.163*** (0.038)		-0.198** (0.081)		-0.204 (0.162)
Foreign	0.034** (0.014)	-0.040 (0.084)	-0.045 (0.034)	0.073*** (0.015)	-0.138 (0.164)	Have a NSFAS email contact	-0.053*** (0.011)	-0.061*** (0.014)			-0.016 (0.023)
NSFAS bursary	-0.006 (0.010)	0.003 (0.015)	-0.040* (0.021)	0.002 (0.027)	-0.008 (0.026)	Have an institutional email	0.068*** (0.007)	0.076*** (0.022)	0.093*** (0.012)	0.039*** (0.010)	0.085*** (0.019)
Other bursary	-0.022*** (0.007)	-0.036* (0.020)	-0.015 (0.014)	-0.031*** (0.011)	-0.024 (0.017)	Campus (ref: CPUT)					
Qualification type (Bachelors)						UCT	0.145*** (0.010)				
Certificate/diploma	-0.002 (0.009)	-0.007 (0.010)	0.138*** (0.023)	0.047 (0.035)	0.032 (0.029)	US	0.156*** (0.011)				
Postgraduate certificate/diplom:	0.015 (0.011)		0.035** (0.018)	-0.020 (0.066)	0.018 (0.029)	UWC	0.088*** (0.011)				
Honours degree	-0.007 (0.010)	0.058 (0.045)	0.025 (0.016)	-0.093 (0.072)	-0.014 (0.022)	Observations	24,710	7,430	6,165	7,380	3,714
Master's degree	-0.051*** (0.013)	-0.158 (0.097)	-0.019 (0.020)	-0.140* (0.077)	-0.022 (0.030)	Chi-square test	1724	884.5	281	279.8	359.4
PhD	-0.031 (0.025)		-0.017 (0.047)	-0.114 (0.086)	0.068 (0.062)	Degrees of Freedom	31	22	26	22	26
						P-value	0	0	0	0	0
						Pseudo R-square	0.0739	0.124	0.0492	0.0477	0.115

Notes to Table 5: Table 5 presents regression marginal effects from an attrition probit analyses in addition to a chi-square test of whether the variables in the model are jointly equal to zero, by institution type. Attriters 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 WCGDS rates to the 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	6894	1,450	21%	426	430	27%
UCT	5102	1,143	22%	300	379	28%
US	7380	1,710	23%	492	533	34%
Total	19376	4,303		1218	1342	

Notes to Table 6: 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 21-23% 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 saying 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 that signaled they were studying is far higher among respondents of the WCGDS than in the overall graduate sample, a 6 % points higher for CPUT and UCT and 11% points higher for US. In addition, among students who were studying, responders are much more likely to be studying at their original institution, while non-responders have a higher share studying at the University of South Africa. 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 test for the probability of Studying in 2012 (UCT sub sample only)

	Coefficient	Std Error		Coefficient	Std Error
Age	-0.006***	(0.002)	Attritor	-0.134	(0.086)
Male	0.012	(0.020)			
Population group (ref: White):			Interactions:		
Not specified	0.075*	(0.043)	Att x Age	0.004*	(0.002)
African	-0.023	(0.028)	Att x Male	-0.019	(0.023)
Coloured	-0.006	(0.030)	Att x Race not specified	-0.048	(0.049)
Indian	-0.075*	(0.041)	Att x Coloured	0.010	(0.032)
Foreign	0.025	(0.053)	Att x Indian	-0.010	(0.034)
NSFAS bursary	-0.015	(0.036)	Att x Foreign	0.005	(0.064)
Other bursary	0.109***	(0.025)	Att x NSFAS bursary	-0.022	(0.040)
Qualification type (ref: Bachelor):			Att x Other bursary	-0.074***	(0.029)
Certi/Dipl	0.053	(0.095)	Att x Certi/Dipl	-0.064	(0.102)
Postgrad cert/dipl	-0.081**	(0.037)	Att x Postgrad cert/dipl	0.017	(0.041)
Honours	0.051*	(0.031)	Att x Honours	-0.002	(0.035)
Masters	-0.096***	(0.031)	Att x Masters	0.002	(0.035)
Doctrate	-0.086	(0.063)	Att x Doctrate	0.007	(0.072)
Contact details available:			Att x Business and commerce	0.047	(0.031)
Email	0.013	(0.051)	Att x Human and social sciences	0.046	(0.031)
Cell	0.091**	(0.036)	Att x Health sciences	0.016	(0.048)
Insitutional email	0.142***	(0.030)	Att x Law	0.039	(0.059)
CESM (ref: Sciemce, engineering and technology)			Att x Education	0.047	(0.070)
Business and commerce	-0.162***	(0.027)	Att x Email	0.028	(0.053)
Human and social sciences	-0.113***	(0.027)	Att x Cell	-0.050	(0.038)
Health sciences	-0.174***	(0.043)	Att x Insitutional email	-0.043	(0.033)
Law	-0.045	(0.053)			
Education	-0.173***	(0.063)	Constant	0.338***	(0.079)
Sample Size				6,165	
R-squared				0.098	
Test of the joint significance of the attrition terms:					
F(24, 6117)				2.19	
Prob > F				0.0007	

Notes to table 7: Table 7 presents coefficients and standard errors of the probability of studying at UCT from a linear probability model. The attritor variable takes a value of one for those who do not respond to the WCGDS, 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 criteria

	All	CPUT	UCT	US	UWC
lamda (inverse mills ratio)					
Coefficient	0.023	0.396	-0.108	-0.054	-0.006
Standard Error	(0.031)	(0.650)	(0.140)	(0.065)	(0.224)
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 to Table 8. 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.

southern africa labour and development research unit

The Southern Africa Labour and Development Research Unit (SALDRU) conducts research directed at improving the well-being of South Africa's poor. It was established in 1975. Over the next two decades the unit's research played a central role in documenting the human costs of apartheid. Key projects from this period included the Farm Labour Conference (1976), the Economics of Health Care Conference (1978), and the Second Carnegie Enquiry into Poverty and Development in South Africa (1983-86). At the urging of the African National Congress, from 1992-1994 SALDRU and the World Bank coordinated the Project for Statistics on Living Standards and Development (PSLSD). This project provide baseline data for the implementation of post-apartheid socio-economic policies through South Africa's first non-racial national sample survey.

In the post-apartheid period, SALDRU has continued to gather data and conduct research directed at informing and assessing anti-poverty policy. In line with its historical contribution, SALDRU's researchers continue to conduct research detailing changing patterns of well-being in South Africa and assessing the impact of government policy on the poor. Current research work falls into the following research themes: post-apartheid poverty; employment and migration dynamics; family support structures in an era of rapid social change; public works and public infrastructure programmes, financial strategies of the poor; common property resources and the poor. Key survey projects include the Langeberg Integrated Family Survey (1999), the Khayelitsha/Mitchell's Plain Survey (2000), the ongoing Cape Area Panel Study (2001-) and the Financial Diaries Project.



www.saldru.uct.ac.za

Level 3, School of Economics Building, Middle Campus, University of Cape Town
Private Bag, Rondebosch 7701, Cape Town, South Africa

Tel: +27 (0)21 650 5696

Fax: +27 (0) 21 650 5797

Web: www.saldru.uct.ac.za

