Appendix 1

Data sources and methodology

The evidence presented in this book was sourced from various studies and projects conducted previously by CREST as well as analyses done specifically for this text. There are four main studies that were used in this regard.[1] We list each study and the core methodology pursued in the table below, and indicate where the study is cited in the relevant chapters. This is followed by a more extensive discussion of the methodology of the study concerned.

No.	Study	Methodology	Chapters
1	Trends in PhD production and efficiency (2014)	Secondary analysis of HEMIS student data conducted specifically for the book	Results of these analyses are found in Chapters 2–5
2	Study of productive departments in SSH (2013)	Qualitative study (face-to-face group interviews) with 25 selected departments at 13 South African universities	Chapter 6
3	DST retention and progression study (2014)	Secondary analysis of HEMIS student data for pipeline analysis	Chapter 3
		Web-based survey of post-graduate enrolments	Chapter 2
4	Study of doctoral supervision in South Africa (2011)	Web-based survey of SA doctoral supervisors	Chapter 5

Study 1: Trends in PhD production and efficiency

Data tables and analyses

The main sources for the data tables were data extracts prepared by the Department of Higher Education and Training (DHET) from the South African Post-Secondary Education (SAPSE) Information System for the period 1986 to 1999 and the Higher Education Management Information System (HEMIS) for the period 2000 to 2012. In order to establish trends over time, all data from pre-merger institutions were mapped to the post-merger universities.

The only source of data before 1960 that could be obtained was an unpublished doctoral thesis from the University of Pretoria by JG Garbers completed in 1960, titled '*Gradueringstendense in Suid-Afrika 1918–1957*'. Another useful source of data before 1986 was a 1982 publication by the former Pretoria-based *Departement van Nasionale Opvoeding* titled '*Statistiek van Kwalifikasies Toegeken deur die Outonome Universiteite in die Jare 1971–1979*'.

The data elements needed for the national versus international doctoral student comparisons were extracted by the DHET from the national HEMIS database according to the specifications needed for the analyses. International data are only available from the year 2000 in the HEMIS data.

HEMIS replaced the SAPSE Information System in 1999. HEMIS is an electronic database maintained by the DHET. It contains unit record data of qualifications offered by public universities as well as students enrolled by universities and staff employed by universities. Space data as well as financial data are also collected from public universities according to specifications. HEMIS data are audited and are provided to the DHET according to technical specifications.

Efficiency: Cohort analyses

Individual student records for all doctoral students for all the years from 2003 to 2012 were obtained from the audited data sets of HEMIS maintained by the DHET. The data for the analysis were extracted from the student as well as qualification files.

The following variables were used for the analysis of the qualification records: 001 – Qualification code; 003 – Qualification name; 005 – Qualification type; 063 – Institution code; 082 – Qualifier; 089 – Mode of delivery; 588 – Submission; 529 – Collection year

The following variables were used from the individual student records: 001 – Qualification code; 005 – Qualification type; 007 – Student number; RegisterID – Unique random number created for students to replace ID number to mask student identity; 009 – Qualification commencement date; 010 – Entrance category; 011 – Date of birth; 012 – Gender; 013 – Race; 014 – Nationality; 024 – Attendance mode; 025 – Qualification requirements status; 026 – CESM category for first area of specialisation; 027 – CESM category for second area of specialisation; 063 – Institution code; 529 – Collection year; 571 – Age in years; 588 – Submission; 589 – Headcount indicator.

Although an attempt was made to determine trends for full-time and part-time doctoral students, the full- or part-time indicator (Element 079 – Full-time/Part-time student) could not be used for analytical purposes since it was not captured for the majority of PhD student records by the universities.

This was a drawback because the difference in the performance of full-time and part-time students could not be compared. It is recommended that the DHET ensures that universities collect and capture this information accurately in future.

Students from pre-merger institutions were mapped to post-merger institutions. Vista students were mapped to their post-2005 institutions by testing the destination institutions in which they ended up after 2005 by tabulating their post-2005 institutions using their Unique_ID. Patterns in student numbers were used to map VISTA students who completed before 2005 to their 'post-merger' institutions. Doctoral student data for the period 1996 to 2004 were thus reconfigured into the post-merger institutions to enable appropriate trend analyses.

For the purposes of this study, throughput rates/completion rates are defined as the percentage of students who graduate within the period of analysis. Dropout rates are the percentage of students who do not graduate within the period of analysis, including students who discontinue their studies. Since each cohort was analysed until 2012, dropouts were defined as students who had not graduated by 2012. Those who were still registered in 2012, but who did not graduate in 2012, were counted as dropouts, since the 2013 data were not available at the time of the analysis to determine how many of them reregistered in 2013.

To calculate throughput/completion rates one needs to look at data from several years to select a cohort of new entrants of students of a reasonable size and follow their graduation trends over a number of years. In essence, throughput/completion rates give an indication of how successful the universities were in graduating new entrants into the qualification in a particular year (cohort) over the period of analysis. It also provides information on dropouts, who are students who do not reregister after dropping out for the period of analysis and who thus do not obtain the qualification. The data contain a record for each year of registration for each student. A student's enrolment and graduation history thus had to be constructed from several enrolment records for each student. Each student record contains at least one – or, at the most, four – Classification of Education Subject Material (CESM) code/s. HEMIS data element 81 – *Institution Programme Name* – is a valuable data field which, with the CESM categories, assists with the classification of students into a field of study.

The CESM code was used to classify the student enrolment into the major fields of study. The CESM category for first area of specialisation was used to classify the student record into the major field of study. The major field of study was classified for each qualification type in which the student was enrolled separately because the student does not necessarily continue further studies in the same field of study. Due to the fact that students change their fields of study often for the same qualification type, the field

of study of the last enrolment of the student for the particular qualification type was used for the classification of the major field of study. Where the CESM category for first area of specialisation was not completed, the CESM category for the second area of specialisation was used for the classification of the major field of study. The second- and third-order CESMs were trimmed to first-order CESM categories. The CESM categories changed in 2010.

The following recoding was done for the years 2001 to 2009:

CESM	Field of study
1 – agriculture, 6 – computer science and data processing, 10 – home economics, 15 – life sciences and physical sciences, 16 – mathematical sciences	Natural sciences
2 – architecture and environmental design, 8 – engineering and engineering technology	Engineering and technology
9 – healthcare and health sciences	Health sciences
4 – business, commerce and management sciences	Business, economic and management sciences
7 – education	Education
3 – arts, visual and performing, 5 – communication, 12 – languages, linguistics and literature, 13 – law, 14 – libraries and museums, 17 – military sciences, 18 – philosophy, religion and theology, 19 – physical education, health education and leisure, 20 – psychology, 21 – public administration and social services, 22 – social sciences and social studies	Humanities and social sciences

The following recoding was done for the years 2010 to 2013:

CESM	Field of study
1 – agriculture, agricultural operations and related services, 6 – computer and information sciences, 10 – family ecology and consumer sciences, 13 – life sciences, 14 – physical sciences, 15 – mathematics and statistics	Natural sciences
2 – architecture and the built environment, 8 – engineering, 11 – languages, linguistics and literature	Engineering and technology
9 – health professions and related clinical sciences	Health sciences
4 – business, economics and management sciences	Business, economic and management sciences
7 – education	Education
3 – visual and performing arts, 5 – communication, journalism and related studies, 12 – law, 16 – military sciences, 17 – philosophy, religion and theology, 18 – psychology, 19 – public management and services, 20 – social sciences	Humanities and social sciences

The study required that the new entrants in each academic year had to be clearly distinguished from doctoral students who have previously been registered. In the HEMIS data the entrance category variable gives an indication of whether the student was a new entrant or a student who had previously been registered for the qualification. New entrants in doctoral programmes are not always consistently classified by universities. For this reason students who registered in year n who were not registered in year n-1 for doctoral studies were considered new entrants.

Study 2: Study of productive departments in the social sciences and humanities

The second study listed here consisted of case study analyses of highly productive departments in the social sciences and humanities. Thirteen universities were selected to take part in this study: the universities of Cape Town, Johannesburg, KwaZulu-Natal, Pretoria, South Africa, Free State, Western Cape, Witwatersrand, Zululand, Nelson Mandela Metropolitan, Rhodes, North-West and Stellenbosch. Collectively they produced 96% of the higher education system's doctoral graduates in social sciences and humanities over the 10-year period from 2000 to 2009.

The first criterion applied to select individual departments for study was that a university must have produced a total of at least 20 graduates in that field over the 10-year period from 2000 to 2009. This process resulted in a list of 52 departments meeting this criterion. Additional criteria were hence applied to reduce the department selection. This included analysing graduation rates and doctoral enrolments over the 10-year period and checking on the progress of cohorts of new doctoral enrolments over the period 2001 to 2004.

Ten disciplines were subsequently identified for further exploration: Education, Psychology, Public Administration, Political Studies, Economics, Sociology, Religion, Law, English and Social Work. In all, these fields produced 80.9% of the doctorates in social sciences and humanities across South Africa. No field contributed less than 2.5% (Political Studies) to the total, and each field also had an average annual intake of at least 26 (Political Studies) doctoral candidates between 2000 and 2004. Ultimately 25 departments were included for this part of the study.

On-site interviews were subsequently conducted by Gillian Godsell and Johann Louw. All interviews were recorded in audio and transcribed, but interviewers also took notes and these were included in the analysis. Table 1 shows the final list of departments selected for this study.

Study 3: DST retention and progression study

Part 1: Analysis of HEMIS data for pipeline analyses

The quantitative component of this study consisted of a secondary analysis of HEMIS student data. The methodology consisted of measuring the actual numbers of bachelors students who proceed to register for honours studies and so on. However, conversion – interpreted as the change from one degree (lower) to another (higher) – can occur either directly upon completion of the first degree or only some time thereafter.

Table 1.1.1: Departments selected

University	Department/Faculty	Number of participants
North-West University (NWU)	Education	2
North-West University	Social Work	2
Rhodes University (RU)	Education	4
Stellenbosch University (SU)	Public Administration	3
Stellenbosch University	Theology	1
Stellenbosch University	Sociology	6
University of Cape Town (UCT)	Faculty of Law	2
University of Cape Town	Economics	2
University of Cape Town	English	2
University of Cape Town	Political Studies	1
University of Johannesburg (UJ)	Education	2
University of Johannesburg	Psychology	1
University of KwaZulu-Natal (UKZN)	(School of) Accounting, Economics and Finance	1
University of KwaZulu-Natal	English	2
University of KwaZulu-Natal	Religion	4
University of Pretoria (UP)	Law	1
University of Pretoria	Public Management Administration	3
University of Pretoria	Social Work and Criminology	3
University of Pretoria	Theology	3
University of South Africa (UNISA)	Public Administration	2
University of the Free State (UFS)	Theology	3
University of the Western Cape (UWC)	Education	1
Witwatersrand University (Wits)	Psychology	2
Witwatersrand University	Political Studies	1
Witwatersrand University	Sociology	1

The analysis started with a consolidated database of 1 933 681 records that included all students enrolled for any degree at a South African university for the period 2001–2013. The comprehensive database included biographical information that allowed for an in-depth analysis of student retention, progression and completion rates in terms of scientific domain (categorised into six broad domains), gender, race, nationality (categorised into three broad geographical locations) and age group (categorised into three broad groups).

Calculation of retention and completion rates

In calculating retention, we used a definition according to which all students who remained registered (enrolled) for a particular qualification level (e.g. masters) within the system over a given period, irrespective of completion status, could be considered 'retained' for that qualification level. Put differently, retention is the percentage of new entrants in a given base reference year (BRY) who have completed a degree by the end of a reporting year or are still enrolled for the degree in that reporting year.

As a first step, we created 12 data sets for each of three qualification levels (bachelors, honours and doctoral), thereby producing 36 data sets in total. In each data set the BRY corresponds to the particular cohort of new entrants in that year.

Each initial year or BRY focused on new enrolments only, with subsequent years focusing on returning entrants. This ensured that we could uniquely identify and track a selected cohort (group). In calculating the completion rate, the same cohort that was selected for retention was tracked over the same period. However, in calculating the completion rate the focus was on the percentages of the BRY cohort who had completed their programme at the end of each reporting year (up to 2013).

Calculation of progression rates

Progression was defined as the percentage of students who would move from one level of qualification to a higher level of qualification over a given period. In this instance, our focus was solely on the 'traditional' paths a student might make i.e. moving from bachelors to honours, honours to masters, and masters to doctoral. In taking this approach we discarded possible deviations/alternatives (e.g. a masters graduate who might enrol for an honours programme).

In order to calculate progression rates we also created 12 data sets for each of three progression paths (bachelors to honours, honours to masters, and masters to doctoral). Thus, a total of 36 data sets was produced. The reason for creating 12 data sets per progression path was that there are 12 possible BRYs, where a BRY corresponds to the particular cohort of graduates. Similar datasets were also produced for the progression from honours to masters (12 data sets) and masters to doctoral (12 data sets). In each case the BRY focused on all graduates in a lower qualification degree, so that they could be tracked in terms of their enrolment for a next-level qualification in subsequent years.

Part 2: Web-based survey

A web-based (electronic) survey that targeted a large sample of postgraduate students at the main universities in South Africa was conducted as the second phase of the DST retention and progression study. The 10 most productive universities (in terms of masters and doctoral output) produce more than 97% of all postgraduate students in the system. We obtained the co-operation of these universities to conduct a web-based survey that was emailed to all their current (as of 2014) honours, masters and doctoral students.

The methodology is based on the assumption that it is best to conduct the qualitative component by asking students who have already made their

decisions about proceeding with further studies to reflect on the reasons for their decisions. This means, in practice, we asked honours students to reflect on their reasons for deciding to proceed with postgraduate studies while they were still undergraduate students. We asked masters and doctoral students to do the same. In addition we developed a set of questions about future plans and decision-making for all three groupings. We therefore ruled out the need to ask undergraduate students to participate in the survey. This decision was based both on methodological grounds (we did not think we would get reliable responses to any questions about future decisions) as well as logistical grounds. The methodology employed in the web-based survey entailed the following:

- Getting approval from the universities to conduct a web-based survey (the questionnaire was sent to all universities prior to the submission of the survey for ethical approval).
- CREST designed and managed the survey from a central point, but sent a web link to a contact office at each university for distributing to their postgraduate students. The e-mail links to the web-based questionnaire were distributed in batches by each university to their own students, thereby obviating the need for making student e-mail addresses available to CREST (which is one of the concerns in studies of this nature).
- Students then responded (anonymously) to the e-mail request to participate in the survey and the completed questionnaires were captured by CREST.
- All data analyses were done in aggregate form.

Three separate electronic questionnaires were drafted: one each for honours, masters and doctoral students. The questionnaires were almost identical with differences in phrasing here and there. The qualitative survey aimed to ascertain the primary reasons for students' attitudes (if applicable) towards interrupting their studies or considering discontinuing their studies. A few questions were also aimed at identifying which factors affected students' choices towards selecting institutions and programmes.

Study 4: A study of doctoral supervision in South Africa

With the huge growth in doctoral enrolments (a doubling of enrolments between 2000 and 2009), it was inevitable that individual academics (those with doctoral degrees) would increasingly face larger and ultimately unmanageable numbers of students to supervise. But the challenge of supervision is not only a matter of additional volumes of students to supervise. Evidence from various workshops on doctoral supervision clearly shows that supervisors are not only finding the increased numbers

challenging, but – even more importantly – also the reality that a large number of prospective doctoral candidates are woefully underprepared for doctoral studies. Supervisors complain about the fact that many of their doctoral students cannot write scientifically, do not know how to search the literature, lack the required quantitative and qualitative skills to do proper data analysis and so on. In cases where doctoral students are underprepared for the specific demands of doctoral studies, the doctoral supervisor has to devote more time to guiding the student through the doctoral research process. The 'burden of supervision' is therefore both a result of the substantial growth in the numbers of students to be supervised as well as the large proportion of doctoral candidates who are ill-prepared for their doctoral studies.

In order to gather more systematic evidence about these and related issues, CREST designed and administered a web-based survey of doctoral supervisors at South African universities in 2011.

Methodology

A database of PhD supervisors was compiled from information obtained from South African universities during 2010. We identified the most 'research productive' supervisors on the basis of their publication output over the preceding 10 years. This process produced a list of slightly more than 3 000 names of possible respondents. All of these academics were subsequently invited in an e-mail letter to participate in the web survey. The first batch of e-mails was distributed through the online survey system of Stellenbosch University on 31 October 2011. The initial closing date of the survey was 14 November 2011, effectively giving the participants two weeks to complete the questionnaire. Although e-mails were sent to 3 042 supervisors, delivery failed to approximately 924 recipients, indicating that the addresses were no longer in use or invalid or that some mailboxes were full. At the time of the deadline of the survey (23 November 2011) a total of 336 questionnaires had been received of which 5 were incomplete. Out of approximately 2 118 sent invitations (3 042 initial invitations minus 924 failed deliveries), a total of 331 valid responses were received, resulting in an overall response rate of 15%.

Apart from collecting demographic information, included various questions about PhD supervisory approaches and styles as well as monitoring and feedback mechanisms in the supervisor–student relationship.

Sample profile

An analysis of the realised sample shows that 72% of the 324 respondents who specified their gender were male. The mean age of respondents at the time of completing the survey was 55, but it is also interesting to note that

significant numbers of respondents (36% of the sample) were over 60 and even over 65 (16%). In general, it is fair to say that the sample represents a slightly older profile than population characteristics. This is mainly because of the manner in which we had defined our target population, i.e. as the most research-productive academics in the country. The representation of PhD supervisors by scientific field is comparable to the production of doctoral graduates across scientific fields, based on figures for 2010.

In addition, we were specifically interested to establish for how long the respondents had been supervising PhD students. Altogether 69% of respondents said they had been doing so for at least 10 years, of which 29% reported 20 or more years.

This short description of the demographics of our respondents revealed that our typical respondent is a male, in their mid-50s and with significant experience in doctoral supervision.

Notes

1 In Chapter 5 on quality reference is made to two other surveys conducted by CREST. The first, was a survey of postgraduate students at Stellenbosch University in 2003; the second was a tracer study of doctoral graduates in 2010 under commission for ASSAf. Since we have cited only a few results from each study, we decided not to discuss the methodologies of these two studies in detail in this appendix.