Opening the black box of Higher Education with the aid of longitudinal cohort analysis[1]

Pierre de Villiers & Chris van Wyk[2]

Introduction

Not much is known about the performance of students in the South African Higher Education system. Most studies that have been conducted worked with aggregate numbers and not with individual student records. This paper outlines the procedures that must be in place to track the progress of individual students in the Higher Education system by making use of longitudinal datasets. This process is explained with the aid of the student records of the Higher Education Management Information System (HEMIS). In the last part of the paper this procedure is applied to first time first year students of the period 2000-2004 that received NSFAS awards. Their progress will be discussed in detail.

Longitudinal Cohort Analysis

When one has the data of individual students for a couple of years one can track students as a group or cohort over a specified period of time. With longitudinal data cohort analysis you can determine how successful students progressed through the Higher Education System in South Africa and how many of them eventually received a certificate, diploma or degree or even more than one qualification. There is a body of literature where longitudinal cohort analysis is used to understand student performance and to track students through higher education systems (Cunningham, Milam & Statham, 2005; Ewell, Schild & Paulson, 2003; Data Quality Campaign, 2006; L'Orange, 2008; Voorhees & Lee, 2009).

A cohort is a group of students who enter a higher education institution at the same time. In this paper we track those students who entered any university in South Africa for the first time in 2000 over a period of 10 years. Tracking these so called first time first-year students is a useful tool for helping administrators understand student performance and help to identify those students who are qualifying, progressing, exiting and re-entering the system. With cohort analysis the progress of different cohorts can also be compared.

Longitudinal data coverage is a key requirement in order to facilitate the process to track the progress of students through the Higher Education System. *Longitudinal data coverage* is a dataset that includes data gathered on the same student from year to year. With this method one can thus determine exactly how many students of a specific cohort dropped out without any qualification, how many graduated and how many are still in the system that have not

---

[1] Paper read at the biannual conference of the Economic Society of South Africa, Bloemfontein, 25-27 September 2013.
[2] Department of Economics, University of Stellenbosch.

received any qualifications at all. Furthermore, with the availability of unit-level student records key questions can be answered such as: "What is the profile of the students who qualified and with what qualification type? What is the profile of the students who dropped out of the system without any qualifications?"

Steps to identify students of a specific cohort in the longitudinal data system

Figure 1 graphically depicts the steps that must be used to identify the students and track them through the HEMIS and create a master list with records of all these students that flowed through the system from 2000 to 2009.

Figure 1

*The Process to identify the 2000 cohort and track students through the sys*tem
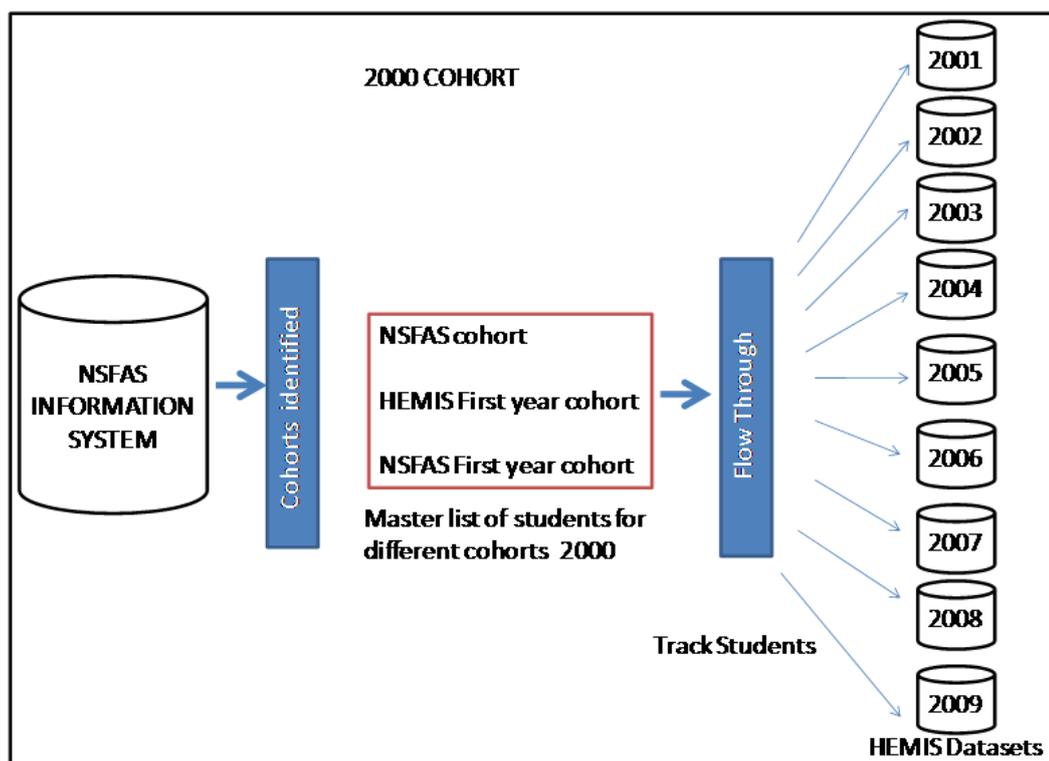
In this study we identified a cohort of students who were first time students in 2000 using the data field in the HEMIS database, called *entcategory* to create a master list. The next step was to determine how successful the 2000 cohort progressed through the higher education system. Therefore you compare the students of the master list with the HEMIS students for each year over time to determine how many students progress from year to year in the following way:

1. Compare the students of the 2000 cohort with the HEMIS students for each year from 2001 to 2009.
2. Identify all the students that flow through the system for each year

3. According to step 2 one can thus determine whether the student is still in the system, received a qualification or dropped out.
4. Store these students that are identified in step 2 for each year as one dataset

Longitudinal Data Coverage in South Africa

The primary focus of this study is the tracking of students through the Higher Education system, using student unit-record datasets captured and maintained by higher education institutions in the Higher Education Management Information System (HEMIS). HEMIS is the key higher education data collection programme designed and implemented to report annually to the Department of Higher Education and Technology (DHET) on the condition of higher education in South Africa. HEMIS collects data on individual students in all higher education institutions in the country. Databases for HEMIS have been in existence long enough to generate long-term enrolment unit-level records. We could obtain the HEMIS datasets from 2000-2009 in order to create a longitudinal dataset of individual students.

The core data elements in the HEMIS enable a more detailed analysis of students over time, such as when a student entered the system, when a student qualified and the type of qualification received together with the different institutions a student attended during the period. Once an individual's student is no more enrolled in HEMIS it means that such a student dropped out of the higher education system without a qualification. Figure 2 shows the number of first time first year students by race for the years 2000-2009. In this paper we only focus on those students who entered university for the first time in 2000 and track them over a period of 10 years.

Figure 2

Cohort of first time first year students in the higher education system of South Africa by race: 2000-2009
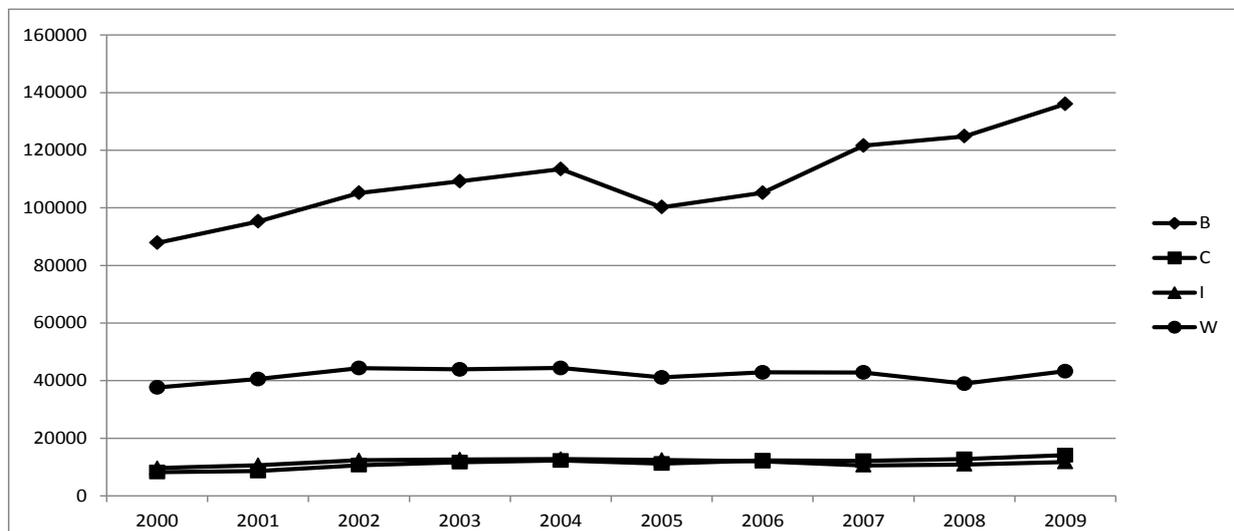
Essential Elements of a Longitudinal Data System
The following key elements are essential in creating a longitudinal data system:

A unique countrywide student identifier
In creating a longitudinal data system it is necessary to link the different datasets that have been collected for individual students for each year through the use of a common field across these datasets. To be able to do this unique identification codes must be assigned to every student. The unique identifier, which is in our case the student's identity number, is used to integrate the datasets of the different years to identify all students in the cohort to develop a single source of standardised individual records.

Furthermore, it is now possible through a student identifier to follow the progress of each student over time, and across institutions or provinces within the country. With unit-level student records it also makes it possible to identify information about a single student across various data sources (e.g. NSFAS and HEMIS) to evaluate the effect of financial support on student performance.

Students progressing through the system
The longitudinal data system facilitates the calculation of progression rate by comparing the master list of the specific cohort with longitudinal student unit-record of HEMIS to determine how many students progressed (remain in the system) from year to year as indicated in Figure 1. The HEMIS datasets are used because it is the official student unit-record system indicating amongst other the student's field of study (CESM), qualification obtained, the entering category, etc.

Students exiting the system without any qualification
Those who exited without any success can be determined by using enrolment from year one minus enrolment (progression) from year two and subtracting all those who graduated at the end of year one. In the discussion of the analysis these students will be referred to as drop outs, meaning that they left the system without any qualifications. However, these students who dropped out after one year can enter the system in the future again even with a different study field or at another institution.

Students receiving a first qualification
The students who obtained a first qualification are calculated by using the requirement qualification field in the HEMIS data set and compare it with the HEMIS longitudinal student unit-records for each student. A student is counted only once irrespective of how many qualifications are obtained over time. This is necessary to ensure that students that obtained

qualifications are not double counted. In the discussion of the analysis these students will be referred to as those that qualified, meaning that they received at least one qualification.

Practical application of tracking students through the HE system

When using longitudinal data (in other words data of the same students for different years) one can track any group of students as they progress through the system. With this presentation we have chosen the students that received NSFAS awards and followed them through the HE system. We decided to use the cohort groups that were first year students for the first time in the period 2000-2004 and received NSFAS awards and tracked them through the HEMIS until 2009.

Success of students that received NSFAS awards

When one looks at the statistics provided by NSFAS it seems as though their students were very successful. There is an incentive built into the scheme for students who are successful in their studies to convert part of the loan into a bursary. Over the period 1996-2011 NSFAS reported that students passed on average 74.4 per cent of the courses for which they entered (NSFAS, 2007; 2010 and NSFAS website available at http://www.nsfas.org.za/profi-statistics.htm). Although courses passed is not the same as obtaining a qualification, this is quite impressive results. For every 25% of courses passed 10% of a student's loan is converted into a bursary – that means up to a maximum of 40% of the loan amount if a student passes all the courses. The above-mentioned pass rate correlates well with the fact that according to NSFAS on average 28.5 per cent of loans were converted into bursaries. However, the Ministerial Committee (2010: 69-70) reported that, of all the students NSFAS funded over the years, 33 per cent are still studying while the other 67 per cent are not at HEIs anymore. Of these students no longer studying, only 28 per cent had graduated, while the remaining 72 per cent had dropped out or did not complete their studies. Thus the NSFAS and Ministerial Committee statistics seem to be contradictory. However, one must bear in mind that courses passed are not the same as obtaining a qualification. This may partly explain these contradictory statistics. With our analysis we try to shed some light on this seemingly contradictory statistics.

Success of NSFAS students using individual data

This section deals with the results of a recent research report (De Villiers, Van Wyk and Van der Berg, 2012) of the progress of first year students for the period 2000-2004 that received NSFAS awards using HEMIS data up to 2009. They investigated how students that received a NSFAS award for the first time in 2000 progressed through the HE system for the period 2000-2009. The same procedure was followed for the cohort groups that received a first

award in 2001 to 2004. To make results comparable they calculated the results for students that were first year students for the first time in those five years.

In 2000 15 345 first-first year students that received NSFAS awards could be linked with the HEMIS database. Table 1 gives an indication how those students progressed through the system. Bear in mind that not all students received NSFAS awards for every year, but the table gives a summary of what happened to the NSFAS students that started in 2000 (irrespective of whether they received an award again or not). In interpreting the tables one must remember that 15 345 students entered the higher education system *at the beginning* of 2000, but 122 obtained some form of qualification *at the end* of 2000. At the end of 2000 2 133 dropped out of the system without any qualifications and 13 090 of the original group continued their studies in *2001*, but did not receive a qualification at the end of 2000. The information in the table thus gives an indication of what the situation was at the end of each year. At the end of 2009 it is not known how many continued their studies in 2010; that is the reason why no values are given for those that continued or dropped out (HEMIS data not available for 2010 and 2011). However, it is known how many obtained a qualification and therefore this could be included in the table. It is important that the numbers given in the 'Qualify' column must be interpreted correctly.

Table 1

Progress of NSFAS 2000 cohort group

Year	Continue	Qualify	Drop Out	Original cohort size
2000	13 090	122	2 133	15 345
2001	11652	224	3 469	15 345
2002	7 864	2 724	4 757	15 345
2003	4 402	5 282	5 661	15 345
2004	2 605	6 735	6 005	15 345
2005	1 672	7 543	6 130	15 345
2006	1 289	8 007	6 049	15 345
2007	1 090	8 288	5 967	15 345
2008	949	8 510	5 886	15 345
2009	n/a	8 678	n/a	15 345

The total qualification of each year in Table 2 was used to calculate the cumulative qualification in the "Qualify" column in Table 1. At the end of 2000 122 students received a qualification and in 2001 another 102 to give a total number of 224. Although some students received a second (or third) qualification, those qualifications were not added to the total, because that would be double counting those students. Table 1 indicates how many *students* obtained at least one qualification and not how many qualifications those students obtained.

From the table one can deduce that 8 678 students of the original 15 345 received at least one qualification in the 10-year period portrayed in the table.

Table2

Qualifications obtained by 2000 NSFAS cohort group

Year	First qualification	Second qualification	All qualifications
2000	122	0	122
2001	102	12	114
2002	2 500	35	2 535
2003	2 558	526	3 048
2004	1 453	491	1 944
2005	808	358	1 166
2006	464	269	733
2007	281	267	548
2008	222	233	455
2009	168	259	427
Total	**8 678**	**2 450**	**11 128**

Figure 3

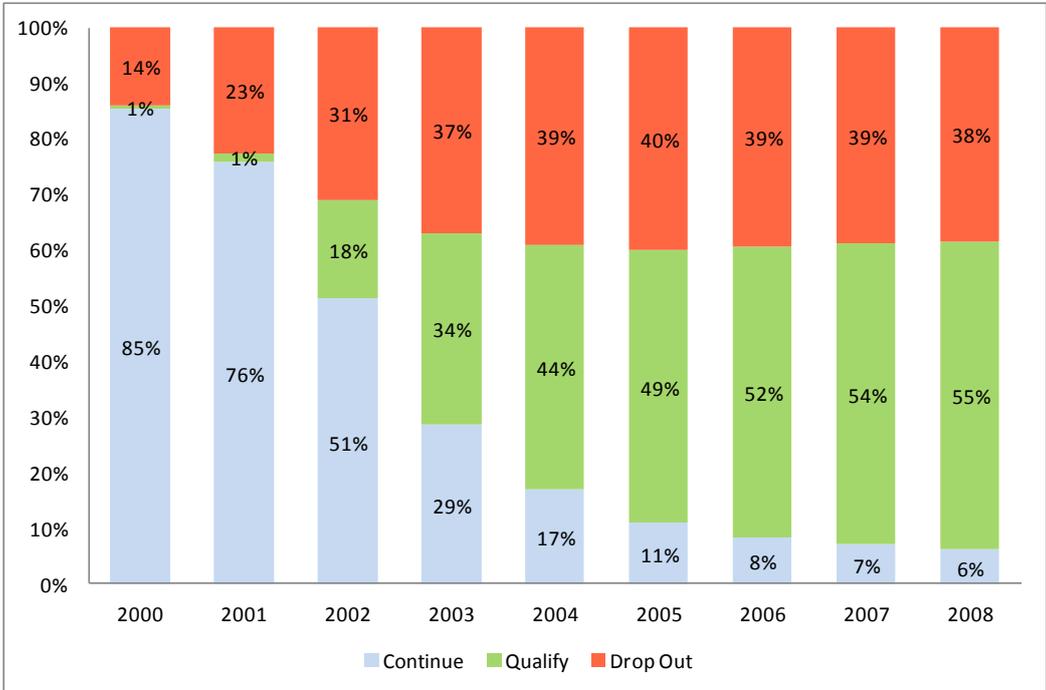*Percentage of 2000 NSFAS cohort group progressing through the system*

Figure 3 is a graphical representation of Table 1, but the numbers are given as percentages. Of the first year students in the 2000 cohort 55 per cent obtained at least one qualification (diploma, certificate or degree) by the end of 2008. With a longitudinal data set it can actually

also be determined how many students received more than one qualification. Of the original group of 15 345 students 2 450 second qualifications were obtained. Of the original cohort group 38 per cent dropped out of the system without any qualification, while 6 per cent were still in the system but had not obtained any qualification.

Interestingly enough this is better than the situation of the non-NSFAS students that started their studies in 2000 (See Figure 4). By 2008 48 per cent of the non-NSFAS students had obtained a qualification, 46 per cent had dropped out without any qualifications and 6 per cent were still in the system without obtaining a qualification. In summary, a greater percentage of NSFAS students obtained a qualification and a smaller percentage of the cohort dropped out of the system without a qualification.

Figure 4

Comparison of the percentage of the 2000 NSFAS and 2000 non-NSFAS cohort groups progressing through the system
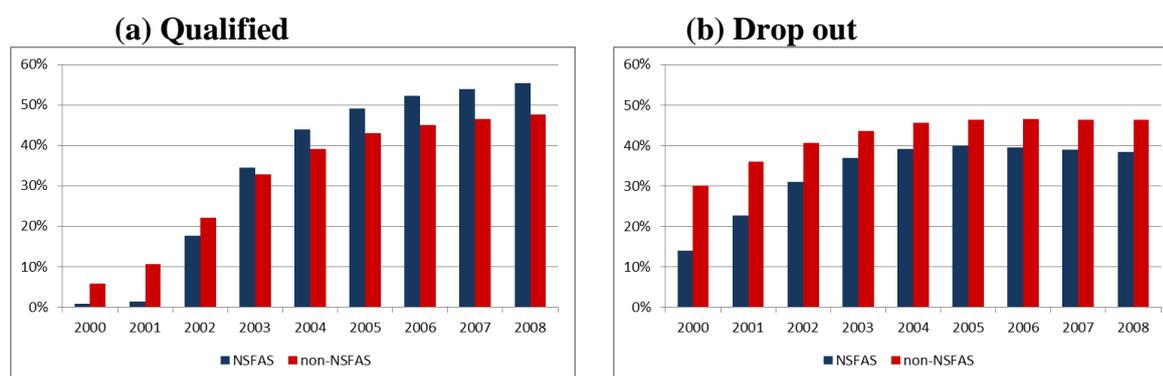
For both groups it is rather surprising that 6% of the original cohort group was still in the system 9 years after they started studying in higher education, but had not obtain any form of qualifications. The other cohort groups for which the analysis was done showed remarkable consistency in success/failure rates, but only the 2000 group will also be discussed in detail.

For example, of the 2004 NSFAS cohort group 38% dropped out without a qualification and 44% obtained at least one qualification. The last percentage is lower than the figure for the 2000 cohort group because fewer years have passed since they started their studies. There was still 18% of the original group in the HE system without obtaining a qualification. As more years pass the percentage with qualifications of this cohort group will definitely increase.

Field of Study of the 2000 NSFAS cohort

The next thing that was investigated is in what fields of study did the NSFAS students received their qualifications. Not surprisingly the majority of qualifications were in the social

sciences – 62.7% while only 37.7% were obtained in natural sciences (See Figure 5). Although the qualifications of each student are available according to its classification of educational subject matter (CESM) category, it was decided to make only the split between natural and social sciences. One of the reasons is that it does not make sense to split the 22 CESM categories according to their funding category, because study field in both natural and social sciences are sometimes grouped in one funding category. To split the qualifications into 22 categories will also make it very difficult to interpret.

Figure 5

Qualifications obtained by 2000 NSFAS cohort group according to field of study
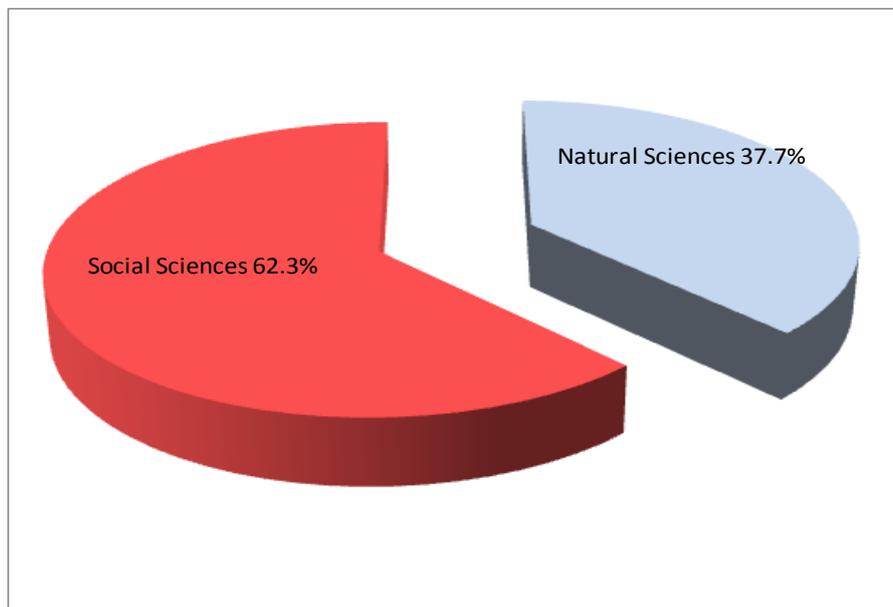

For all the years under discussion the percentage obtaining a degree in the social sciences ranged between 62.3% and 66.7%, while those obtaining a degree in natural sciences ranged between 33.3% and 37.7%. It was thus fairly stable over time.

Type of Qualification of 2000 NSFAS cohort

Another thing that was investigated is what the split was between degrees and diplomas/certificates that NSFAS students obtained. This is summarized in Figure 6 Clearly more diplomas/certificates were obtained than degrees. This is hardly surprising as degrees take at least three years to complete while many diplomas and certificates can be completed within one year. There is a small fraction of the qualifications that could be split between social and natural sciences, but could not be classified as degrees or diplomas/certificates due to data problems. This explains the unknown section in Figure 6.

Figure 6: *Qualifications obtained by 2000 NSFAS cohort group according to type of qualification*
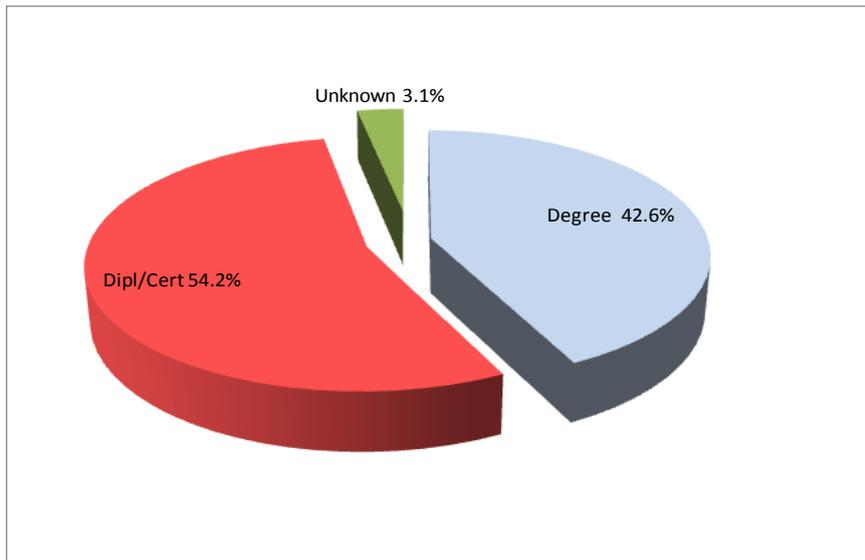

An interesting phenomenon was detected here. With each successive cohort the percentage obtaining a degree increased. It is quite remarkable that of each cohort group that obtained a qualification, the percentage that completed a degree increased from 42.6% of the 2000 cohort group to 54.5% for the 2004 cohort group. This is indeed impressive as a degree is normally on a higher cognitive level than diplomas or certificates.

Type of Qualification per Institution of 2000 NSFAS cohort group

It was also determined how many qualifications were obtained in each institution. From Figure 7 certain clear trends can be seen. Almost two-thirds of all degrees obtained were at traditional universities, while almost no degrees were obtained at universities of technology. Just more than 30% of degrees were obtained at comprehensive universities. The reverse is true for diplomas and certificates. Almost no diplomas or certificates were obtained at traditional universities while almost 70% were obtained at universities of technology. About 26% of diplomas/certificates were obtained at comprehensive universities. Not much changed between these percentages for the other cohort groups.

Figure 7: *Qualifications obtained by 2000 NSFAS cohort group according to type of institution*
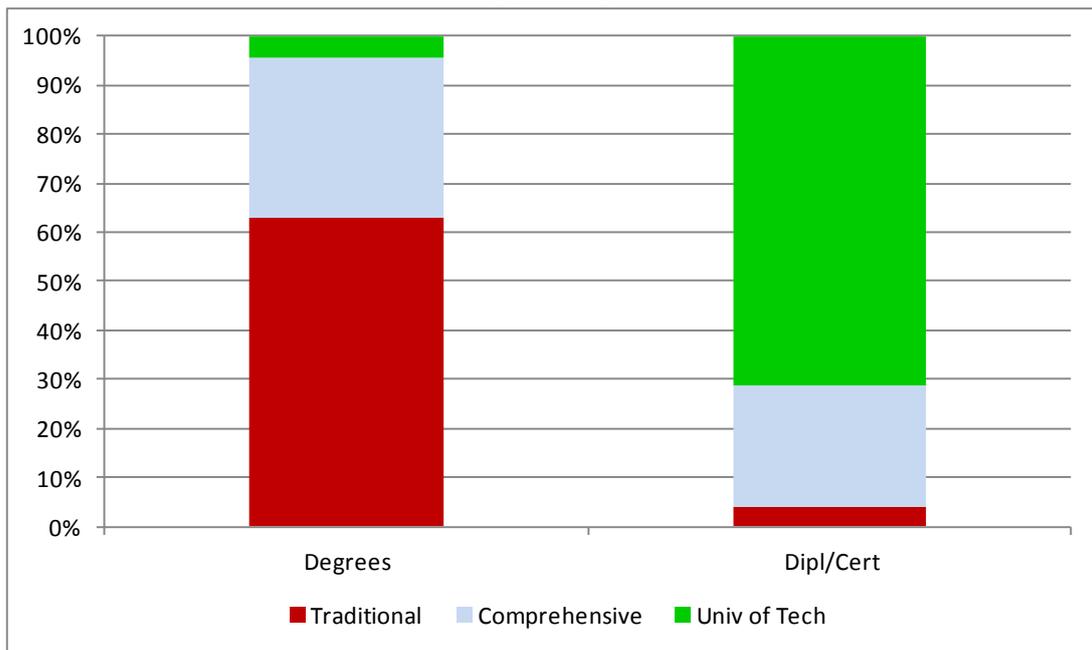
It must be added that with longitudinal datasets one can determine at which individual institution students qualified. One can also calculate how successful students at each institution are and/or can compare the different type of universities on how their students are progressing through the system.

Results of money spent on NSFAS students

From this analysis it seems as though NSFAS students are more successful in the sense that a higher percentage of them obtain qualifications and a smaller percentage drop out of the HE system without qualifications. It does seem as though the financial support allow these students to continue their studies even when not fully successful, whereas non-supported students tend to drop out more easily. 71.2 per cent of the money spent on the 2000 cohort group was spent on successful students (that obtained a qualification) and this dropped to 64.2 per cent of the money spent on the 2004 cohort group that was spent on successful students (mainly due to the shorter period of this cycle 2004-2009). Although it seems that money was spent reasonably efficiently, it did in some instances take too long to identify unsuccessful students that were still receiving an award. Some students received an award for 9 years, without having obtained any qualifications.

Concluding remarks

With the longitudinal datasets it was possible to track 5 cohort groups that received NSFAS awards and started their studies in higher education in the years 2000-2004 and determine

how they progressed through the system by means of the HEMIS datasets of 2000-2009. This analysis required a major research effort due to the large number and size of the datasets. However, it illustrated what is possible with longitudinal datasets.

The results showed that NSFAS has been remarkably successful in terms of graduation rates (those that obtained a degree, diploma or certificate). It is most likely that recipients of NSFAS awards are poorer students and first time generation students. This makes their success in graduating even more remarkable.

From the report it is clear that NSFAS students outperformed the non-NSFAS students in graduating. This is most probably linked to the fact that the financial support lessened the dropout among partly successful students. The trends between the different cohort groups are be very stable over time. For example, the 2004 cohort group has similar graduation rates as the 2000 cohort. It is also encouraging that NSFAS students increasingly obtained degrees rather than diplomas or certificates.

Over the years NSFAS contributed to make higher education more affordable to the poor and also helped HEIs that traditionally serviced poorer communities to balance their books. Without question the scheme contributed positively to make higher education more accessible and affordable to the poor. There can be no question that NSFAS played and is still playing a positive role in making higher education more affordable and thus accessible to the poor.

References

Cunningham AF, Milam J & Statham C, 2005. *Feasibility of a Student Unit Record System Within the Integrated Postsecondary Education Data System*, Research and Development Report. National Center for Education Statistics. March 2005.

Data Quality Campaign, 2006. *Creating a Longitudinal Data System: Using Data to Improve Student Achievement, Available from:* www.DataQualityCampaign.org,

De Villiers P, Van Wyk C, Van der Berg S. 2012. *The first five year project – a cohort study of students awarded NSFAS loans in the first five years 2000-2004*, Research Report for NSFAS, May 2012.

Ewell PT, Schild PR & Paulson K, 2003. *Following the Mobile Student: Can We Develop the Capacity for a Comprehensive Database to Assess Student Progression?* Research Report, Lumina Foundation for Education, *National Center for Higher Education Management Systems.* April 2003

L'Orange H, 2008. *Unit Record Data Systems: A State Perspective, Available from: www.sheeo.org/resources/publications/unit-record-data-systems-state-perspective*

National Student Financial Aid Scheme. *Annual Report. 2007*. Donors. Available at http://www.nsfas.org.za (accessed on 6 October 2009).

National Student Financial Aid Scheme. *Annual Report. 2008*. Available at http://www.nsfas.org.za/resources10/NSFAS_Annual_Report2008.pdf (accessed on 16 February 2011).

National Student Financial Aid Scheme. *Annual Report. 2010*. Available at http://www.nsfas.org.za/docs/annual-reports/2010/AnnualReport2010.pdf (accessed on 16 August 2011).

Voorhees R & Lee J, 2009. *Basics of Longitudinal Cohort Analysis: Principles and Practices of Student Success*. Lumina Foundation for Education