

Profiling students using an institutional information portal: a descriptive study of the Bachelor of Arts degree students, University of South Africa

Omwoyo Bosire Onyancha¹

University of South Africa, Department of Information Science
onyanob@unisa.ac.za

Received 15 August 2010

Accepted 19 November 2010

Using data mining techniques, this study examines the Bachelor of Arts (General) degree's data available in the University of South Africa's institutional information and analysis portal (IP) maintained by the Department of Information and Strategic Analysis (DISA). The purpose of this was to draw a demographic profile of the students and demonstrate the potential use of an IP in monitoring and evaluating the performance of individual qualifications as far as registrations, cancellations and graduation rates are concerned. Data were analysed in order to determine the students' age, gender, occupational, home language and geographic distributions and the relationships between the "incoming", "re-entering", "degree completed" and "graduation" headcounts. It was observed, among other findings, that the BA(G) degree attracts students with diverse characteristics; there is a general continued decline in the number of students registering as well as completing the qualification; the number of students cancelling registrations in BA(G) has continued to grow since 2005; and that there is a significant positive correlation between (a) the "incoming" and "graduation" headcounts; (b) "incoming" and "degree completed" headcounts; (c) "degree completed" and "graduation" headcounts; and (d) "graduation" and "total registration" headcounts. Other findings as well as conclusions and recommendations are offered.

Keywords: Data mining; profiling students; institutional information portals

I Introduction

It is widely acknowledged that knowing one's clients plays a huge role in the development of effective customer care programmes which lead to customer satisfaction and retention, thus enabling one to have a competitive advantage over one's competitors. Schenck and Chiaravalle (2007: 62) argue that "by knowing your customer geographics, you know *where* to reach your customers; by knowing their demographics, you know *who* to reach; and based on their behaviour patterns, you know *how* to reach them". For a sound understanding of one's customers, Hosey (2008) suggests that one needs to ask oneself questions such as the following: Who exactly are one's customers? What are their behaviours? What issues or challenges do they face? What is the best medium to reach them? And, what are their profiles? Typically, in the process of preparing the statistical profiles of the customers, companies or organisations collect such data as age and sex, place of residence, and occupation and income. Schenck and Chiaravalle (2007) postulate that there are three categories of facts and figures that one should take into consideration when developing a customer profile. They include geographics, demographics and behavioural patterns. However the type of data would depend on the type of organization and purpose for which the data are required. For instance, in their book on consumer behaviour in tourism, Swarbrooke and Horner (1999:170) observe that tourism organizations require a wide variety of quantitative and qualitative data on tourist behaviour, including the following: statistical profiles of tourists; statistical records of tourist behaviour; how tourists make purchasing decisions; who makes the purchasing decision; when the purchasing decision is made; consumer perceptions; tourist satisfaction; and cultural and national differences in tourist behaviour. In the case of community or public libraries, the user profile would include such information as the number of people living in a community; sex; age distribution of each sex; marital status; different cultures and languages; occupation and income; unemployment rate; and educational and literacy levels (Evans 2000).

Contestably, at higher institutions of learning, the student is portrayed as the customer (Brannigan *et al* and Clare as cited in Vidanapathirana, Morais & Dorabawila 2001:32; van Lill 2005). These authors argue that university students are customers because of their capacity to influence the programme mix by way of choosing relevant courses and rejecting others, which fail to make acceptable service quality standards, particularly where the students are called upon to pay for higher education. While supporting the need to study student profiles, Vidanapathirana, Morais and Dorabawila (2001) opine that the profiles of adult students impact on their expectations and perceptions pertaining to the quality of services offered by the faculty. In their article entitled "*Delivery and program administration of the Bachelor of Arts degree in Social*

1. Omwoyo Bosire Onyancha (PhD) is a Senior Lecturer in the Department of Information Science, University of South Africa, Pretoria.

Sciences of the Open University of Sri Lanka – Assessment of service quality”, the authors argue that distance learners who are, to a large extent, mature adults have unique demands different from students learning in conventional classrooms. This point is emphasized by Sorensen, Baylen and Sweeney (1999) who outline the following questions that reflect the uniqueness of distance learners’ demands:

- How will the access created by distance education affect student profiles?
- How will we provide student services (financial aid, counselling, etc.) to distant students?
- How will we assist students who are deficient in the technology skills necessary to be successful?
- How will we meet the needs of students who want education and services offered in their homes or at their worksites?
- In looking at library resources, should the university invest more in the acquisition of electronic resources and less in the acquisition of traditional paper resources?
- What about registration and student advising in a distance environment?
- How can we assure equity in access to library resources for distant students?
- What infrastructure will be needed to provide technical support to students?

Obviously, none of these questions can be properly addressed without a clear understanding of the demographic, geographic and behavioural characteristics of the students. In other words, a comprehensible and critical knowledge of the students’ statistical data such as the total number of registered students in a qualification as well as their age, gender, race, location, home language and nationality will greatly impact not only the quality of services offered to them but also the medium through which the services are rendered, among other issues related to teaching and learning. The emergence of the information and communication technologies (ICTS) has made the aforementioned statistical data readily available through such tools as information portals.

2 About institutional information portals

The term “portal” means different things to different people as witnessed in a great number of definitions offered by various writers (Dias in Van Brakel, 2003). Opinion is therefore divided on what constitutes a portal. Expounding on the difference between a “static Web site” and a portal, Van Brakel (2003:593) asserts that “simply affixing the word ‘My’ to a system and adding a personal logon feature, definitely does not metamorphose a static Web site into a portal”. A portal differs from a static Web site in the sense that a portal “enables users to customize, personalize and tailor resources and services according to their unique needs and preferences” (Katz & Goldstein in Van Brakel 2003:594). These features are characteristic of what has become to be commonly termed “information portals (IPs)”, sometimes referred to as enterprise information portals (EIPs). Terming them as types of the new information systems, Chou, Hsu, Yeh and Ho (2005:527) note that IPs have “emerged as one of the most effective ways of distributing, sharing business information, reusing knowledge, helping employees spend more time innovating and less time locating information”. Vaca-Vink (2005: definition section) defines an IP as a “concept for a Web site that serves as a single gateway to a company’s information and knowledge base for employees and possibly for customers, business partners, and the general public as well”. In order to clarify this definition, Vaca-Vink (2005) uses one of the models of an IP and observes that an IP should possess the following elements/characteristics:

- **Access/search:** Access/search allows a user to obtain all the information needed (but no more) in the desired context. For example, a loan officer does not need marketing information to approve a loan. An EIP makes sure the loan officer obtains only the information needed by him or her to perform his or her functions.
- **Categorization:** An IP categorises all information so that it is delivered to the user in the context needed (an apt example would be the subject structure on Yahoo).
- **Collaboration:** An IP allows individuals to collaborate regardless of geographical location.
- **Personalization:** The information provided to individuals using an IP is personalized to that person’s role, preferences and habits.
- **Expertise and profiling:** Expertise and profiling are essential for the collaboration element of an IP. Individuals in an enterprise are profiled according to their experience and competencies. If an individual needs to collaborate with others, he or she can choose those who are qualified for the project.
- **Application integration:** Software applications are integrated to allow individuals to deliver, access and share information regardless of the type of applications used.
- **Security:** This provides information to users on the basis of security clearance. The user logs on and is given access only to information that the user is authorized to access.

It is worth noting that the above outline of what constitutes the elements of an IP, however, does not take into account the different types of information portals. Broadly speaking, there are two types of IPs, namely public and corporate portals, which Van Brakel (2003) refers to respectively as horizontal and vertical portals. Whereas horizontal portals are open systems which normally do not require authentication for the general public whom they serve, the vertical portals “are limited by specific authentication requirements, such as those reserved for employees of a company or institute”

(Van Brakel 2003:597). It is in this latter category of portals that institutional portals or campus portals (e.g. university and governmental portals) are classified. According to Dias in Van Brakel (2003:597), vertical portals can be further categorized into the following:

- a. Portals with the emphasis on decision support (assisting employees to gain access to corporate information in order to make consistent business decisions): information or content portals, business portals, and decision processing portals.
- b. Portals with the emphasis on collaborative processing (using groupware and office automation systems, and dealing with information from the traditional supply chain, as well as information produced by groups or individuals out of this chain): collaborative portals, and expertise portals.
- c. Decision support and collaborative processing portals (a combination): knowledge management portals, and enterprise information portals.

The huge amount of literature on IPs indicates that they are increasingly becoming important tools for strategic management of enterprises as well as the management of information generated within (e.g. staff profiles) and outside (e.g. customer and company profiles) a given institution. Since their inception in the late 1990s (Rusch 2002; Chou & Chou 2002), IPs have become not only tools to support strategic management of enterprises but also as part and parcel of the management functions of various enterprises which aspire to have a competitive advantage over their competitors. Institutions of higher education (HEIs) have not been left behind. Data mined from these portals assist institutions to make informed decisions about various aspects of their operations which affect, among other things, the products, customers, company image, personnel skills and services. The HEIs, and more particularly universities, are using or developing IPs to assist them in managing education-, training- and/or research-related programmes. Zhu, Wang and Ju (2004) provide examples of several universities that were using or developing IPs at the time, most of them based in the West. In Africa, university portals are a recent phenomenon, whereby intranets are largely precursors.

3 About Unisa's institutional information and analysis portal

The University of South Africa (Unisa), an open and distance learning (ODL) institution, through its Department of Information and Strategic Analysis (DISA) launched its institutional information and analysis portal in August 2007. The portal was implemented in September the same year. Using the Higher Education Data Analyzer (HEDA), the portal provides, among other information, the following:

- Institutional Overview 2004-7: Aggregated graphic dashboards and reports on enrolments, exam results, etc.
- Institutional Information, 2004-7: Current and past student registrations, assignments, examinations, graduations, success and graduation rates, and staff information
- HEMIS Statutory Information, 2000 - 2006: comprising official, audited student and staff information
- Narrative reports, institutional research, strategic analyses and presentations

Specifically, the institutional information component of the portal provides a total of 58 fields representing different variables as shown in Figure 1. The portal allows a combination of two or more fields. For instance, in order for one to extract data on the number of female students with disabilities in the Bachelor of Arts (General) qualification, one would select three fields, namely Qualification, Disability and Gender.

Underscoring the importance of the portal, Unisa's DISA (2007), in its e-notice to staff and faculty of Unisa, notes the following:

“The previous lack of availability of reliable information was a tremendous obstacle to effective management. The new *Institutional Information and Analysis Portal* provides an immediate solution.”

4 Purpose of the study

Broadly, this paper reports the findings of an exploratory study of the BA(G) qualification at Unisa. The paper draws the demographic profile of the students using the institutional information portal (IP) to demonstrate the potential use of an IP in monitoring and evaluating the performance of individual qualifications as far as registrations, cancellations and throughput rates are concerned. Specifically, the study focused on the following objectives:

- a. To examine the students' demographic and geographic diversity
- b. To determine the rate and type of cancellations of modules in the qualification
- c. To assess the trends of student registration and graduation
- d. To compare BA(G) registrations and those of the most registered qualifications in the College of Human Sciences

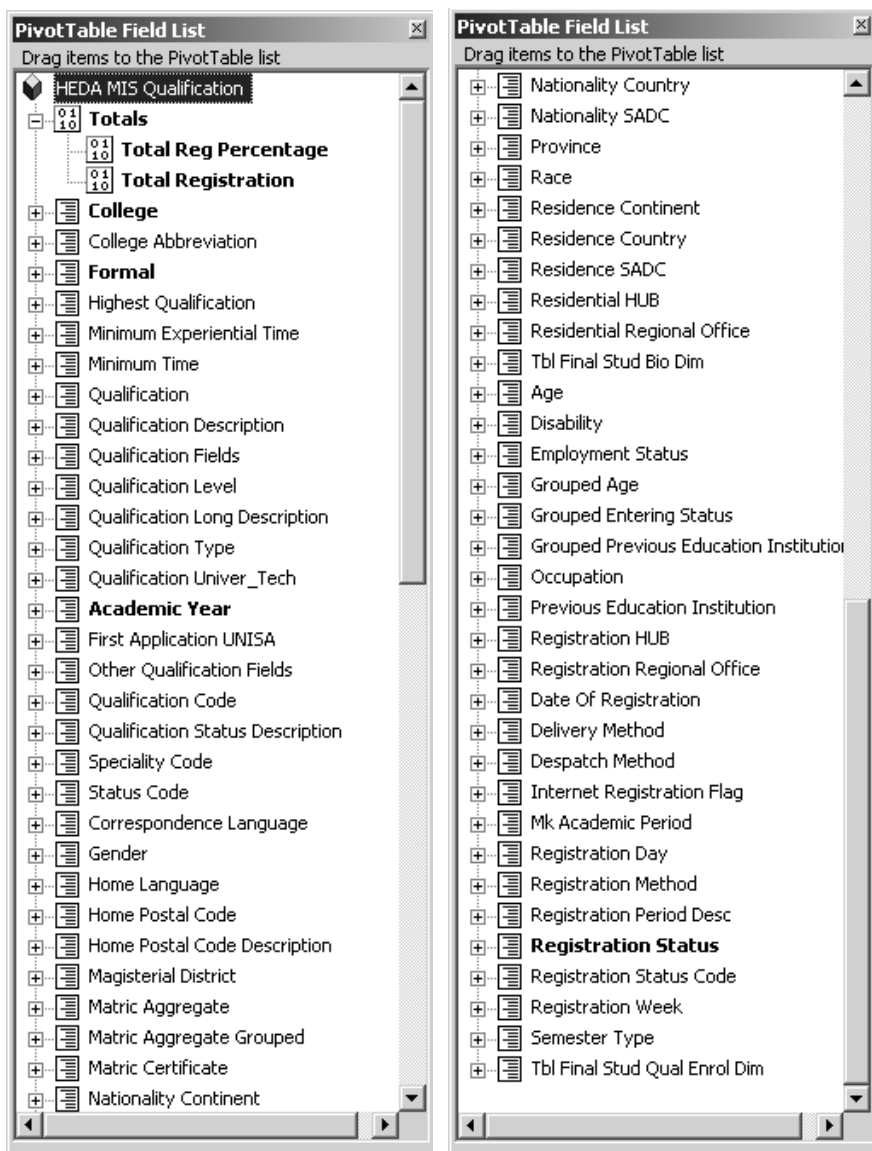


Figure 1 Unisa'sIP's Pivot Table Field List for the provisional qualification registrations

5 Methods and materials

Data mining techniques or approaches were adopted to conduct this study. Lee, Changchien, Wang and Shen (2006:147) argue that there is tremendous amount of data resident in large numerous databases, and there are thus huge amounts of knowledge hidden in the database. The authors further observe that "it is an important course to obtain the valuable information in the database" and in that regard they see the main purpose of data mining as "to discover knowledge from a huge data set". The basis for and importance of data mining are also underscored by Jianhua and Deyi (1998:348) who observe the following:

With massive amounts of data stored in databases, mining information and knowledge in databases has become an important issue in recent research ... Several emerging applications in information providing services, such as data warehousing and on-line services over the Internet, also call for various data mining and knowledge discovery techniques to understand user behavior better, to improve the service provided, and to increase the business opportunities.

There are two purposes of knowledge discovery, namely verification and discovery (Jianhua & Deyi 1998:349; Anane 2001). Whereas verification is meant to verify the user's hypothesis, in discovery, the system finds new patterns. The discovery goal can further be subdivided into prediction (in which case the system finds patterns for the purpose of predicting the future behaviour of some entities) and description (whereby the system finds patterns for the purpose of presenting them to a user in a human-understandable form) (Jianhua & Deyi 1998:349). According to Anane (2001:302), the "output of the data mining process can take several forms [which] include facts, classification, trends, association, or relation between two or more variables". Some of the methods of data mining that have been proposed by various writers include database search, generalization, knowledge representation and mathematical or statistical modelling.

This study adopted a descriptive approach of data mining to explain and discuss patterns showing up in student registrations, cancellations and throughout in the BA(G) qualification as well as to develop a student profile based on various variables. The use of information portals to assess phenomena has been employed by Boshoff (2005) to find out 'The representation of women academics in higher education in South Africa'. Boshoff (2005) collected data from the South African Post-Secondary Education (SAPSE) system of the National Department of Education to determine, among others, the permanent instructional/research staff by sex and by: sector, race group, age group, rank and qualification.

This study followed the four steps of the data mining process, as proposed by Simoudis in Anane (2001).

Step 1: Data selection: The source of data was Unisa's institutional information and analysis portal. The study extracted relevant data from the provisional and HEMIS (Higher Education Management Information System) reports. The focus of the study was the total number of registration headcounts in various qualifications in the CHS, student headcount by age, gender, race, nationality, citizen-resident status, and home language. These variables were selected because they have a huge impact on teaching and learning, particularly in an ODL environment. Data extraction was limited to the academic years between 2004 and 2010 for the provisional reports and 2005 and 2008 for the HEMIS reports. Only formal registration and cancellation headcounts are included in the analysis at this stage, omitting the temporary registration headcounts of the provisional reports. It was deemed necessary to work with the provisional headcounts because they include students who are not active but registered at Unisa, as opposed to the HEMIS headcounts which consist of only the active students. Active students refer to those students who have submitted at least the first assignment. HEMIS reports do not include the students who register at Unisa but fail to submit any assignment.

Step 2: Data transformation: This step was intended to derive new attributes from the extracted data. Firstly, upon the capture of data in the Microsoft Excel worksheets, total headcounts were obtained and the percentage contribution of the variables in question was computed. Because it would have been erroneous to measure the performance of each variable on the basis of the total headcounts since the yearly registrations included "re-entering" students, we opted to report the performance of each variable using the percentage contributions of each variable. Take the following example:

Table I Data transformation???

Continent	2004	2005	2006	2007	2008	2009	2010	TOTAL
Africa	10 225	9 228	7 933	7 177	6 707	6 298	5 190	52 758
Asia	55	58	27	21	18	15	12	206
Europe	183	165	138	121	109	80	73	869
Oceania	6	6	4	5	5	5	5	36
The Americas	31	21	21	17	19	18	15	142
Total	10 500	9 479	8 126	7 341	6 858	6 416	5 295	54 015

Instead of describing the contributions of each region in terms of the total number in the last column (for that would have been erroneous), we converted the figures for each region in column 9 into percentages of 54 015 and used the generated percentages to compare the dispersion of students in the five regions (see Fig 1).

Further analysis of the data involved correlational tests between the graduation and registration headcounts. A Pearson product moment correlation coefficient (r) was computed using the Pearson Correla-

tion function (shown below) in order to determine the relationship between two given variables as explained in section 6.5.

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

Four sets of data were entered into the Microsoft Excel worksheet and the aforementioned function applied to the sets to generate the Pearson correlation coefficient. Basically, the Pearson function returns the Pearson product moment correlation coefficient (r), a dimensionless index that ranges from -1.0 to 1.0, inclusive, and reflects the extent of a linear relationship between any given two data sets, whereby a negative coefficient implies a negative relationship, while a positive coefficient implies a positive correlation. The closer the coefficient is to 1, the stronger the relationship will be between any two data sets.

Step 3: Data mining: The data were then critically analyzed in order to discover the existence of any patterns or new knowledge about the performance of the BA(G) qualification.

Step 4: Results interpretation: This step involved the interpretation of the results and conclusions drawn from the data in order to propose further areas of study as well as offer suggestions about the areas of improvement as far as the IP is concerned.

This study was also informed by a review of the literature as published on the internet sources and in books, journals and other periodicals.

6 Results and discussion

The results and discussion of the findings are provided in accordance with the above-mentioned specific objectives (see section 3). The section is divided into the following subsections:

- Trend of registration in BA(G) vis-à-vis other qualification in the CHS
- Internationality of the BA(G) qualification
- BA(G) qualification's student profile
- Students' entering status
- Completion and graduation status
- Course cancellations in BA(G)

6.1 Trend of registration in BA(G) vis-à-vis other qualifications in the CHS

The BA(G) qualification attracts the largest number of students in the College of Human Sciences (CHS) at Unisa. Out of over 350 qualifications offered in the CHS, the BA(G) degree comprises approximately 3.28% of the total number of registered students in the entire University and about 12.5% of the College's total registered students. Table 2 compares the trend of registration and number of students registered in the top ten qualifications in the CHS between 2004 and 2009. The table shows that, out of the qualifications that recorded the highest number of registered students, the BA(G) degree witnessed the highest registration in 2004 (i.e. 10 500). That figure declined by 10.8% from 10 500 in 2004 to 9 479 in 2005. Apart from the year 2006, which recorded a decreased percentage of 16.7% from the previous year, each subsequent year recorded a decrease in the number of registered students by between 6.7 and 10.7%. Again, it was noted that among the ten top qualifications, the Bachelor of Arts qualifications were dominant. These included the BA (Health Science and Social Services), BA (Communication Science) and BA (Human and Social Studies).

Table 2 Trend of student registration for the BA(G) qualification

Qualification	2004	2005	2006	2007	2008	2009
BA (GENERAL)	10 500	9 479	8 126	7 341	6 858	6 416
BA (HEALTH SC & SOC.SERVICES)	3 587	4 142	4 628	5 157	5 689	5 798
HIGH DIP ADULT BASIC ED & TR	4 266	3 559	3 266	3 312	3 759	4 927
BACHELOR OF SOCIAL WORK	1 256	1 830	2 592	3 689	5 298	6 535
BA (COMMUNICATION SCIENCE)	2 549	2 805	3 110	3 429	3 723	3 889
BED (ECD: FOUNDATION PHASE)	1 564	1 909	2 250	2 738	3 386	4 384
BA (HUMAN AND SOCIAL STUDIES)	1 343	1 703	2 229	2 831	3 351	3 311
BED (INT & SENIOR PHASE)	927	1 176	1 448	2 040	3 090	4 976
ACE (LIFE ORIENTATION)	516	709	1 089	1 647	3 707	4 430
POSTGRAD CERTIFICATE IN EDU (SPFET)	1 463	1 480	1 694	1 856	2 295	2 591

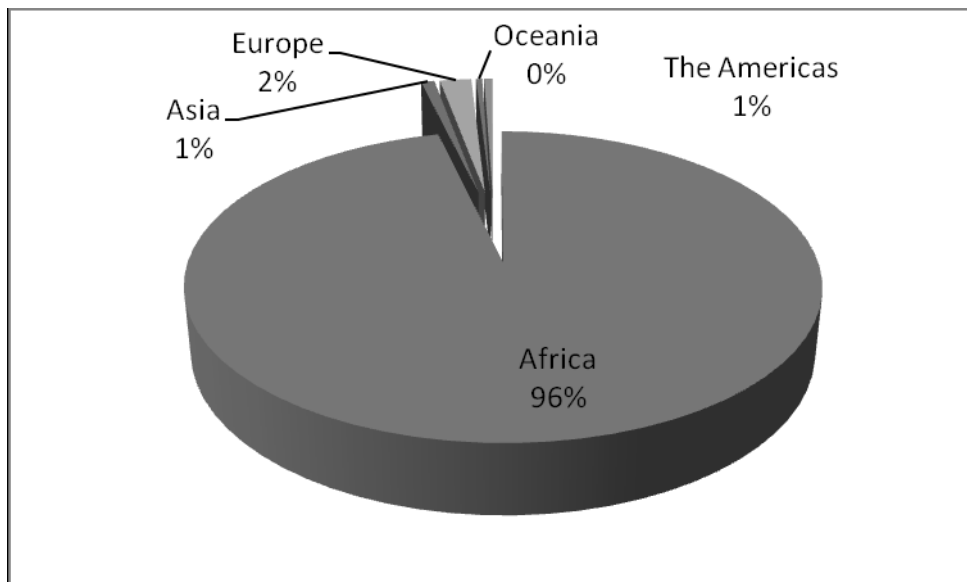


Figure 2 Student distribution according to continent of residence

6.2 Internationality of the BA(G) qualification

An analysis of the student's nationality status revealed that most students are African (98%) and European (2%). The BA(G) degree attracts students from five continental regions, that is, Africa, Asia, Europe, Oceania and the Americas. We also sought to determine where the students are located as a student may be a national of country A but reside in country B. In that respect, it was noted that most students reside in Africa (96%) followed by Europe (2%), Asia (1%) and the Americas (1%) as shown in Figure 1. Although a Pearson correlation test returned a correlation coefficient (r) of 0.999991, implying no significant difference between the continent of nationality and continent of residence, there were nevertheless some students who have nationality in one continent but reside in another. For instance, whereas the students with African nationality comprised 98%, the percentage of students who actually reside in Africa is 96%. This implies that some African students reside in foreign countries. It is surprising that students in developed countries such as those based in Europe would take BA(G) studies at Unisa which is based in a developing country. Is the qualification superior to those offered by universities in foreign countries? Is it unique? Or, do students simply wish to identify with the home institutions? An exposition of the factors influencing such preferences on the part of students, we believe, would assist the responsible persons to market the qualification more widely.

6.3 BA(G) qualification's student profile

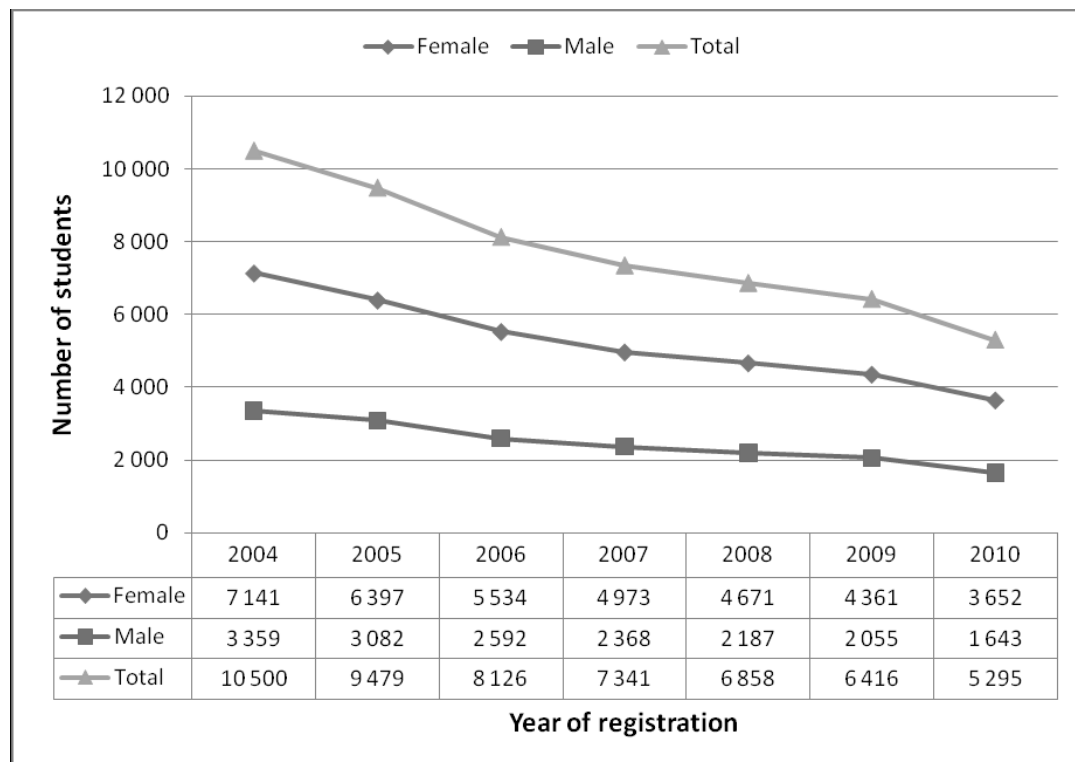
This section draws up a profile of BA(G) qualification students in terms of geographical distribution (country of residence), gender, occupation, age and home language, because these attributes have a massive impact on the way in which the students can be reached and tutored, especially in an ODL environment (see Table 3).

6.3.1 Geographical location (country of residence)

The geographical distribution of students reveals that the majority of students reside in South Africa, which accounted for 88.38% of the student population. In the second position was Mauritius, which contributed a mere 4.06%, followed by Zimbabwe (1.27%), the United Kingdom [UK] (0.98%), Namibia (0.62%), Kenya (0.52%), Botswana (0.38%), Swaziland (0.31%), the USA (0.31%) and Australia (0.27%), to name but a few. It was also noted that the BA students are scattered throughout the world and more specifically in 89 countries. Twenty-seven (27) of these are African countries. There was an entry designated as "Other African" which had two students. This category would at most constitute two countries which could bring the total number of African countries to 29 out of the possible 53 independent African countries. It therefore follows that BA students are spread across 54.7% of the African countries, which in turn means that the qualification's uptake or penetration in the African region seems to be average. Factors leading to this minimal performance require investigation. The factors may include language because only countries from English-speaking Africa featured in the top-ranking countries.

Table 3 Students' countries of residence

No.	Country	%	No.	Country	%
1	South Africa	88.38	26	Malawi	0.06
2	Mauritius	4.06	27	Turkey	0.05
3	Zimbabwe	1.27	28	Greece	0.05
4	United Kingdom	0.98	29	Uganda	0.05
5	Namibia	0.62	30	China	0.05
6	Kenya	0.52	31	Ethiopia	0.04
7	Botswana	0.38	32	South Korea	0.04
8	Swaziland	0.31	33	Belgium	0.04
9	USA	0.31	34	Austria	0.04
10	Australia	0.27	35	Italy	0.03
11	Zambia	0.22	36	Spain	0.03
12	Israel	0.17	37	Norway	0.03
13	Germany	0.16	38	Qatar	0.03
14	England (UK)	0.15	39	Scotland (UK)	0.03
15	Mozambique	0.15	40	Iran	0.03
16	Canada	0.15	41	Netherlands	0.03
17	United Arab Emirates	0.14	42	Japan	0.02
18	New Zealand	0.12	43	Oman	0.02
19	France	0.09	44	India	0.02
20	Taiwan	0.09	45	Portugal	0.02
21	Ireland	0.08	46	Tanzania	0.02
22	Hong Kong	0.08	47	Argentina	0.02
23	Switzerland	0.07	48	Bahrain	0.02
24	Saudi Arabia	0.06	49	Chile	0.02
25	Angola	0.06	50	Nigeria	0.02

**Figure 3** Distribution of students according to gender

6.3.2 Gender

Figure 3 shows the distribution of students by gender. It is apparent from the illustration that the BA students are predominantly female. The females accounted for a higher percentage contribution than their male counterparts throughout the years under investigation, that is, 2004 to 2010. In fact, the number of female students was more than double the number of males. For instance, out of the total 10 500 students registered in 2004, 68.01% were females, while males accounted for 31.99%. The percentage distribution of females and males, in that order, between 2005 and 2010 was as follows: 2005 (67.49%, 32.51%); 2006 (68.10%, 32.90%); 2007 (67.74%, 32.26%); 2008 (68.11%, 32.89%); 2009 (67.97%, 32.03); and 2010 (68.97%, 31.03%). Another pattern worth mentioning is the rate of decline in the number of registered students in both categories (i.e. females and males) in line with the total registrations in each year. The number of female students declined from 7 141 in 2004 to 3652 in 2010, a percentage decrease of 95.54% while the number of male students decreased by 104.44% from 3 359 students in 2004 to 1 643 in 2010. The total number of registered students decreased from 10 500 in 2004 to 5 295 in 2010, a percentage decrease of 98.30%. This trend is worrisome and therefore requires investigation to ascertain the factors leading to the huge decline in the number of students registering for the BA(G) qualification at Unisa.

6.3.3 Occupation

A total of 90 different entries were made under the field tag "occupation" of the BA(G) qualification. Table 4 provides the top 30 most frequent entries which reveal that the BA(G) degree attracts a variety of persons working in different sectors as well as various ranks. The most represented occupations include secondary education teaching which accounted for 26.54%, followed by full-time student at Unisa (19.36%), unemployed (8.41%), clerical or related worker (6.30%) and full-time student (elsewhere) (4.57%). Included in the list of occupations were rank titles such as managers/administrators. It is not surprising to find natural scientists listed among the professionals studying BA(G) modules, for example, nurses, dieticians, etc.; computer specialist, medical doctor/dentist, etc; forester/conservationist; and engineer. Ordinarily, one would expect this group of students to take natural science-based subjects or courses. However, among the subjects offered under the BA qualification are biology, chemistry, computer concepts, and information systems. Although students may not major in any of these subjects, they enrol to study the subjects in order to upgrade their knowledge or their workplaces demand the kind of knowledge related to the subject content of the modules in question or to satisfy the qualification requirements.

Table 4 Student's occupation

RANK	OCCUPATION	%	RANK	OCCUPATION	%
1	Secondary education teacher	26.54	16	Permanent force member	0.44
2	Full-time student at Unisa	19.36	17	Religious worker	0.41
3	Unemployed	8.41	18	Technician (other)	0.39
4	Clerical or related worker	6.30	19	Prison service worker	0.36
5	Full-time student (elsewhere)	4.57	20	Protection service (other)	0.34
6	Manager/administrator	2.63	21	Lecturer/professor	0.32
7	Housewife	1.71	22	Technical worker (other)	0.30
8	Sales worker	1.34	23	Medical doctor/dentist, etc.	0.27
9	Nurses, dieticians etc.	1.17	24	Librarian/archivist	0.26
10	Writer/artist/musician	0.78	25	Retired	0.25
11	Computer specialist	0.77	26	Research worker	0.24
12	Police officer/detective	0.67	26	Forester/conservationist	0.24
13	Service worker (other)	0.65	26	Engineer	0.24
14	Personnel officer	0.52	27	Social/recreation worker	0.22
15	Accountant/auditor	0.45	28	Jurist	0.18

An analysis of the employment status of students revealed that a large number of students (i.e. 47.52%) are employed, while full-time students comprised 23.93% and the percentage of the unemployed students was 10.37%. The rest of the students (18.18%) did not provide information on their status.

5.3.4 Age

A student's age is a vital factor to be considered in an ODL institution when implementing such policies as those that touch on the medium of communication with students, curriculum development, assessment, methods of tuition and the application of information and communication technologies in education. For instance, a study conducted by Ross-Owen (2003) concluded that a "significant positive relationship exists between age and GPA [grade point average]". There are also concerns about the low uptake of information and communication technologies by elderly students, a situation that may hinder effective tuition in an ODL institution such as Unisa. Duvall and Schwartz (200) consider age as one of the potential barriers to academic achievement, particularly in technology-assisted learning. In this regard, "age" was examined in order to profile the age difference between the BA degree students. It was noted that the majority of students are between 20 and 29 years old, followed closely by those who are between 30 and 39 years old. The 40 to 49 age group comprised 23%. The youngest student was 16, and the oldest 88. This wide range of the students' ages is indeed challenging because their retention rate may vary accordingly. It was not clear from the data collected whether or not this pattern has impacted in any way on student performance. It would be worth conducting a study to assess the extent to which age affects student performance.

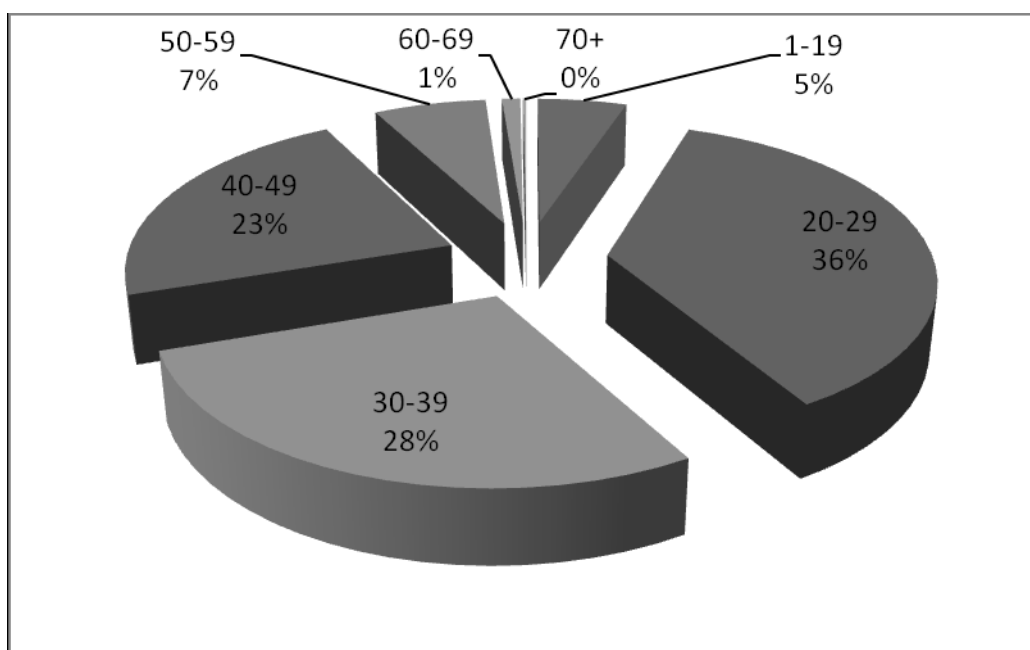


Figure 4 Distribution of students according to age

6.3.5 Home language

Table 5 shows the number of students' home languages. It should be mentioned at the outset that this information, like most of the personal information included in the institutional portal, was supplied by students. It is not surprising therefore to have the English language topping the list of home languages throughout the entire period under investigation, namely 2004 to 2010. It is probable that students did not select the correct option considering that majority of them are African nationals. There are very few people with an African origin who would have the English language as their home or first language. Essentially, home language is critical in student performance because it is often not the language of instruction. Unisa's policy recognizes two languages as the official languages of instruction, namely English and Afrikaans. Overall, the home languages that dominated the top positions are English, Northern Sotho, Zulu, Afrikaans, Tswana, Xhosa and Southern Sotho, to name a few. It should be noted that besides English and French, all the top-ranking languages are South African. Home languages belonging to other African countries were categorized under "other African languages". One other entry that emerged in the list of home languages was Afrikaans/English, which implied a representation of multilingual students. The provision of this option by Unisa could be attributed to the recognized languages of instruction at the University. Outside Africa, the home languages include French, Portuguese, German, Spanish, Greek, Italian and Hebrew.

Table 5 Distribution of students according to home language

Language	2004	2005	2006	2007	2008	2009	2010
English	3 540	3 425	3 086	2 899	2 843	2 791	2 398
Northern Sotho	1 557	1 297	1 134	966	805	659	493
Zulu	1 118	1 011	863	760	731	685	531
Afrikaans	927	832	679	612	550	602	513
Tswana	750	653	498	406	353	282	226
Xhosa	671	544	436	378	314	281	225
Tsonga	372	299	261	240	239	210	170
Southern Sotho	338	296	226	177	170	125	96
French	132	154	151	181	188	207	159
Venda	211	187	174	139	136	136	113
Swati	164	131	110	111	91	81	64
Ndebele	142	123	98	97	93	75	63
Other African languages	148	130	102	78	77	64	51
Afr./Eng.	98	88	82	77	79	71	60
Other foreign languages	114	104	77	64	61	61	47
Shona	80	89	56	64	48	27	33
Portuguese	68	49	41	40	34	24	21
German	22	26	25	21	21	18	19
Spanish	14	14	12	8	4	4	4
Greek	13	10	5	9	7	3	1
Italian	8	5	5	5	3	3	4
Hebrew	7	6	2	3	4	1	1
Ndonga	6	5	2	1	4	1	1
Unknown		1	1	5	3	5	2
TOTAL	10 500	9 479	8 126	7 341	6 858	6 416	5 295

6.4 Students' entering status

The field tag "entering status" refers to status of students at the time of registration in terms of whether or not they are new to the qualification or continuing students. Incoming students in this study consist of the following:

- Those registering for the first time for a BA(G) qualification. This group is designated as "New Registrations" in Table 6.
- Those transferring from other qualifications within Unisa to BA(G), referred to as "Others" in Table 5.
- Those transferring from other institutions and registering at Unisa for a BA(G) degree, designated as "Transfers (Other Institutions)"

The "Re-Entering Unisa" category refers to students who are reregistering for the same BA(G) qualification at Unisa.

Table 6 demonstrates that, as in the case with the general trend of registration between 2004 and 2010, whereby the number of students has continued to decrease, the number of new students registering in the BA(G) degree is declining. Nevertheless, there were occasional increases. For instance, the number of new registrations increased from 918 in 2005 to 951 in 2007, while the number of students transferring from other institutions increased from 82 in 2009 to 117 in 2010. The total number of "new registrations" between 2004 and 2010 was 6 617, while the "other" and "transfers (other institutions)" categories recorded a total of 171 and 1 303 students, respectively. The most worrisome trend, however, involves the decline in the number of students reregistering in the qualification. Apparently, the number of students dropping out has continued to grow. There are several possible reasons for this kind of pattern. As we will see

shortly, this trend may be attributed to the huge number of students cancelling modules in the qualification for financial, administrative and voluntary reasons (see Fig 6). It could also be caused by students opting to register in different qualifications within and outside Unisa. It would be prudent to investigate the factors that are contributing to the decrease in the number of students reregistering for the BA(G) degree.

Table 6 Students' entering status

Title needed	2004	2005	2006	2007	2008	2009	2010
New registrations	1 138	1 241	918	951	978	855	536
Other	49	37	28	15	17	1	24
Re-entering (Unisa)	8 993	7 902	6 971	6 222	5 740	5 478	4 618
Transfers (other institutions)	320	299	209	153	123	82	117
Total	10 500	9 479	8 126	7 341	6 858	6 416	5 295

6.5 Completion and graduation status

The field tag "qualification status description" provides, *inter alia*, the information about the number of students who have provisionally completed the degree; those who have not completed it; final-year students; and preliminary completions. For comparison purposes, we have limited our analysis of the data to the years 2004 to 2008 because the HEMIS reports provided graduation headcount statistics for only the five years. Figure 5 paints a "sorry" picture of the trend of "in-coming", "degree completed" and "graduation" headcounts where

- "In-coming" is a combination of "new registrations", "other" and "transfers (other institutions)" shown in Table 6;
- "Degree completed" refers to the number of students who have completed their studies in BA(G) and have graduated or await graduation.
- "Graduation headcount" refers to the DoE's official and audited figures of the number of students who have actually graduated from Unisa with a BA(G) degree.

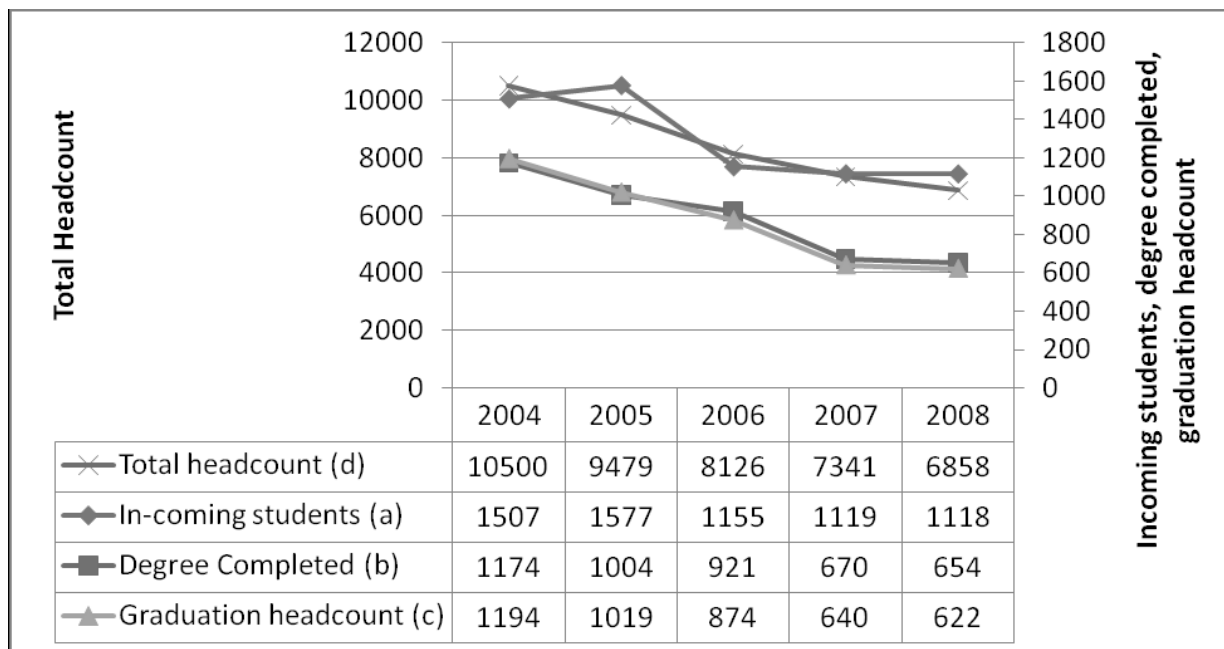


Figure 5 Relationship between headcounts of incoming, completing and graduating students

It was noted that the DoE's audited graduation figures are higher than those labelled as "degree completed" mainly because some students who were designated as "final-year students" and "preliminary completion" in the portal are not included in the "degree completed" category of students. It is possible that by the time the graduations take place, more students would have completed and been confirmed as graduands, thereby raising the number of graduation headcounts.

Generally, the comparison of the four headcounts (i.e. incoming [a], degree completed [b], graduation [c], and total headcounts [d]) shows a similar trend, whereby each has been on the decline since 2004. In fact, the rate of decline for the three components (i.e. a, b, and c) is proportionate to the trend witnessed in the total headcounts per year. In addition to the trend shown in Figure 5, a Pearson correlation test was performed to assess the relationship between the headcounts. The Pearson correlation coefficient (r) that was generated for each pair of data sets (i.e. a and b ; a and c ; b and d ; and c and d) is shown in Table 7.

Table 7 Pearson correlation coefficient (r) for each pair of data sets representing headcounts

Pearson product moment correlation coefficient (r)				
	a	b	c	d
a	-	0.832	0.871	-
b	0.832	-	-	0.973
c	0.871	-	-	0.987
d	-	0.987	0.987	-

Table 7 indicates the existence of strong relationships between the variables. There were significant positive relationships between the following:

- "Incoming" and "degree completed" ($r=0.832$). This implies that the number of incoming students affects the number of the students completing the degree.
- "Incoming" and "graduation headcount" ($r=0.871$). The number of incoming students impacts positively on the graduation (throughput) rate.
- "Degree completed" and "total headcount" ($r=0.987$). There was an almost perfect relationship between these two, a situation that implies that in each academic year, the number of graduating students is a function of the total number of students registered for a BA(G).
- "Degree completed" and "graduation headcount" ($r=0.987$). Ordinarily this relationship should produce a Pearson correlation coefficient value of 1, implying a perfect relationship because the number of students completing a degree each year should be similar to the HEMIS headcount. However, for the reasons outlined above, Unisa's provisional headcount is lower than that of the DoE – hence the existence of an "imperfect" relationship between the two.

6.5 Course cancellations in BA(G)

Data on the course cancellations were available for the years 2005 to 2008, as shown in Figure 6. Three types of cancellations were provided, namely academic, financial and voluntary cancellation. Administrative cancellation refers to situations in which a student is mistakenly registered but later on cancelled and reregistered, say, in the right module. Financial cancellation is caused by the student's failure to clear study fees and the University therefore cancels his or her registration in a module, while the student voluntarily cancels his or her registration for various reasons. Throughout the entire period of investigation, the financial cancellation (FC) was dominant, with 1 703 students cancelling their courses in BA(G) in 2005, 1 503 in 2006, 2 142 in 2007 and 1 962 in 2008. Voluntary cancellations (VC), which equally recorded high values, were as follows: 2005 (974), 2006 (1 275), 2007 (1 282), and 2008 (1 099), while the pattern witnessed in administrative cancellations (AC) was as follows: 2005 (381), 2006 (457), 2007 (292), and 2008 (329). The total number of cancellations in each year was 3 058 in 2005, 3 235 in 2006, 3 716 in 2007 and 3 390 in 2008. Except for one year (i.e. 2008), in which there was a decline in the number of students cancelling modules, there has been a continued increase since 2005.

Of concern among the three types of cancellations are FC and VC, because AC is usually solved when a student is reregistered correctly. In other words, students in the AC category will continue their studies even within the same year that their registration was cancelled. If we therefore exclude these students from the rest of the students whose registrations are cancelled per annum, the BA(G)'s loss of students due to financial and voluntary withdrawal is as follows: 2005 (2 677), 2006 (2 778), 2007 (3 424) and 2008 (3 061).

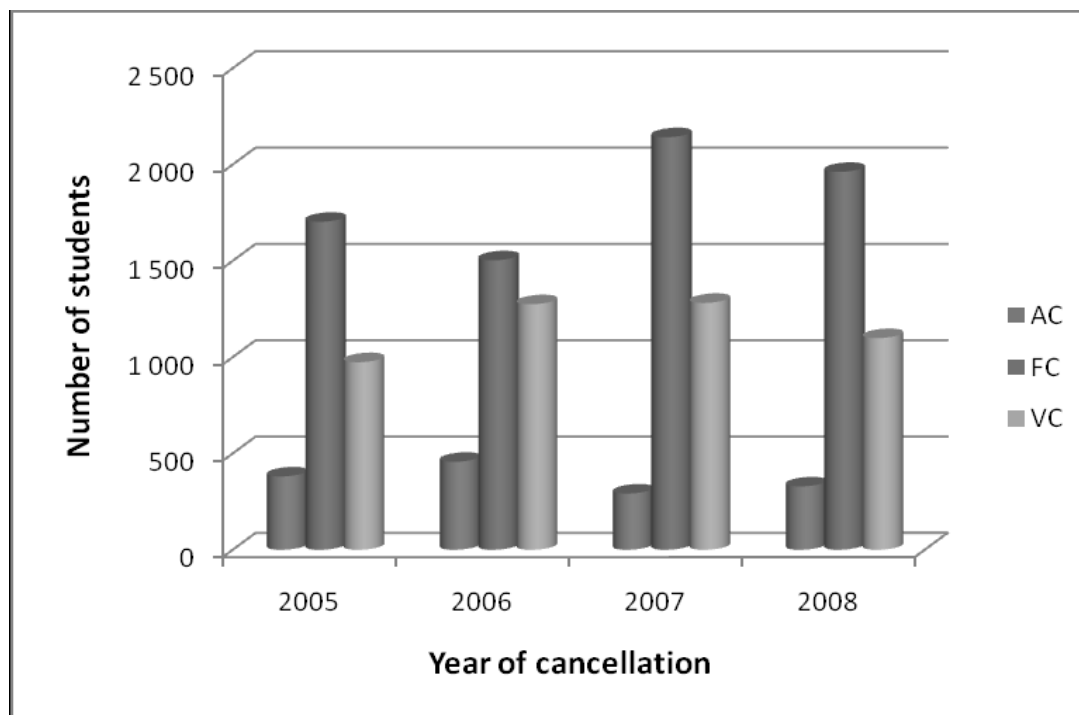


Figure 6 Cancellation status and trend

Key: AC – administrative cancellation
 FC – financial cancellation
 VC – voluntary cancellation

7 Conclusions and recommendations

The findings of this study have revealed that the Bachelor of Arts (General) qualification or degree caters for a variety of students, the majority of whom reside in South Africa. There is diversity in language, age and occupations, as well as the geographical regions in which students reside. These factors, we believe, impact greatly on teaching and learning, particularly in an ODL institution such as Unisa. Because the extent to which these factors impact on student performance was not clear, we can only advise that lecturers take these factors into consideration when discharging their duties. In regard to the degree's penetration in Africa, it was noted that since 2004, students from only 54.7% of the 53 independent African countries have registered in the BA(G) programme. Some countries recorded numbers as low as one student. There is a need for aggressive marketing of Unisa programmes if the institution is to achieve its vision of being an African university for the service of humanity. Although the graduation headcount of students correlates with the total number of students registered in the qualification, there should be concerns about the small percentage of students actually graduating each year. Whereas the percentage of students completing the degree ranged between 9.13% and 11.18%, the DoE's official figures of the BA(G) graduates ranged between 8.72% and 11.37% of the total number of students registered for a BA(G) between 2005 and 2008. The performance of the qualification with regard to retaining students is also worrisome because many students do not come back for the same qualification. The number of students reregistering in the qualification has continued to decrease, while the number of those deregistering for various reasons (e.g. financial and voluntary cancellations) has continued to grow.

With regard to the importance of information portals in institutions of higher learning, we observe that IPs can only be useful in knowledge management as well as strategic management of enterprises and their offerings if the information they provide is filtered and interpreted to support the decision making and policy formulation processes in the affected organization or institution. The provision of information in its raw form (e.g. statistical data) in IPs is only the first step taken in the processes of effective management of academic programmes in HEIs. Informed decision-making processes can only occur if data mining techniques are combined with analytical skills to discern emerging patterns in the available data/information. Qualitative assessment and interpretation of the extracted data are essential when explaining the emerging patterns. As Cloete and Snyman (2004) argue, IPs (or "enterprise portals", as they term them) can only be

valuable technologies, if several components, namely knowledge-sharing culture, knowledge-sharing policies, organizational processes, performance measurement and business strategies are favourable, implemented or permissible in an organization. We recommend linkage of Unisa's information portal to external portals (including open access portals) in South Africa (e.g. the DoE's statistical reports) for knowledge-sharing purposes.

Acknowledgements

This study was commissioned by Dr. Britta Zawada, the Director of the School of Education, Languages and Communications. The assistance rendered to the author by the Department of Information and Strategic Analysis is also appreciated.

References

- Anane, R. 2001. Data mining and serial documents. *Computers and the Humanities*, 35: 299-314.
- Boshoff, N. 2005. The representation of women academics in higher education in South Africa: progress in the pipeline? *South African Journal of Higher Education*, 19(2): 359-377.
- Chou, DC. & Chou, AY. 2002. Healthcare information portal: a Web technology for the healthcare community. *Technology in Society*, 24:317-330.
- Chou, T., Hsu, L., Yeh, Y & Ho, C. 2005. Towards a framework of the performance evaluation of SMEs' industry portals. *Industrial Management and Data Systems*, 105(4):527-544.
- Cloete, L & Snyman, RMM. 2004. Are enterprise portals – knowledge management? *South African Journal of Libraries and Information Science*, 70(1):47-57.
- Duval, CK & Schwartz, RG. 2000. Distance education: relationship between academic performance and technology-adept adult students. *Education and Information Technologies*, 5(3):177-187.
- Hosey, KY. (2008). Top ten marketing mistakes to avoid. [Online]. <http://www.378consulting.com/>. (Accessed 7 March 2010).
- Jianhua, F & Deyi, L. 1998. An overview of data mining and knowledge discovery. *Journal of Computer Science and Technology*, 13(4): 348-368.
- Lee, C-F, Changchien, S.W., Wang, W-T & Shen, J-J. 2006. A data mining approach to database compression. *Information Systems Frontiers*, 8: 147-161.
- Van Lill, D. 2005. Grooming great graduates. *South African Journal of Higher Education*, 19(5): 969-989.
- Ross-Owen, T. 2003. Retention implications of a relationship between age and GPA. [Online]. <http://www.encyclopedia.com/doc/1G1-103563740.html>. (Accessed 6 April 2010).
- Schenck, BF & Chiaravalle, B. 2007. *Branding for dummies*. Indianapolis, Indiana: Wiley Publishing.
- Sorensen, C., Baylen, D & Sweeney, J. 1999. Distance education and emerging technologies: implications for the future of postsecondary education. [Online]. <http://www.cedu.niu.edu/~sorensen/presentations/aeect99.htm>. (Accessed 22 March 2010).
- Swarbrooke, J & Horner, S. 1999. *Consumer behaviour in tourism*. Oxford: Butterworth-Heinemann.
- University of South Africa. Department of Information Strategic Analysis. 2007. Launch of Unisa's Institutional Information and Analysis Portal. Internal memo sent to staff and faculty, 20 August 2007.
- Vaca-Vink, G. 2005. Enterprise information portal. [Online]. http://searchsap.techtarget.com/sDefinition/0,,sid21_gci533557,00.html. (Accessed 22 March 2010).
- Van Brakel, PA. 2003. Information portals: a strategy for importing external content. *The Electronic Library*, 21(6):591-600.
- Vidanapathirana, U., Morais, N & Dorabawila, S. 2001. Delivery and program administration of the Bachelor of Arts degree in Social Sciences of the Open University of Sri Lanka – Assessment of service quality. *Open University of Sri Lanka Journal*, 3:30-56.
- Zhu, F, Wang, A & Ju, Y. 2004. A framework to develop a university information portal. In: Proceedings of 2004 International Conference on Information Acquisition. [Online]. <http://ieeexplore.ieee.org/iel5/9454/30011/01373422.pdf>. (Accessed 22 March 2010).