

Approaches to predictive studies: Possibilities and challenges

A. J. Visser

Alternative Admissions Research Project
Centre for Higher Education Development
University of Cape Town
Rondebosch, South Africa
e mail: avisser@ched.uct.ac.za

M. Hanslo

Centre for Higher Education Development
University of Cape Town
Rondebosch, South Africa
e mail: mhanslo@ched.uct.ac.za

Abstract

This article investigates methodological issues associated with predictive studies related to selection and access to Higher Education. These issues are discussed in the context of tests designed and administered by the Alternative Admissions Research Project (AARP) at the University of Cape Town. The aim of the project is to design tests that broaden access to talented students who are not easily identified through the High School Senior Certificate examination system. The tests attempt to provide a mechanism for selection based upon whether the writers have the potential to succeed in the University environment. The article comments on the difficulties and limitations of various approaches to predictive studies. The article argues that a methodology grounded in the survival analysis approach holds particular promise for predictive studies and that it can provide a useful insight into the complex processes of student persistence, retention and attrition rates.

INTRODUCTION

In South Africa the Matriculation Certificate Examination serves as the primary gatekeeper to selective higher education institutions. Access to these higher education institutions is based purely on academic achievement at the school level. This is particularly problematic given applicants' uneven access to resources and vastly different schooling opportunities. This has led to a situation where certain individuals' potential to learn is underestimated and their access to higher education are essentially blocked. As Hill (2002, 2) points out, 'if valid and reliable tests can be developed to assess such potential, the education system will be able to effectively nurture students who are better positioned to benefit from HE'.

In addition, a new system of school assessment in the Further Education and Training Certificate (FETC) is to be phased in from January 2006. In periods of

uncertainty and major change it becomes all the more necessary to develop additional valid instruments to assist in the reliable assessment of applicants who wish to enter selective HE institutions.

One of the core reasons for the Alternative Admissions Research Project's (AARP) existence is grounded in the need for an assessment tool which endeavours to widen access into the University of Cape Town (UCT). It is therefore the primary objective of the AARP to develop a battery of relevant admissions tests that identify educationally talented students with the potential to succeed. In order to realise this central mission it is critical for the AARP to regularly engage in research to assess the impact of its tests. This research most frequently takes the form of predictive studies (Yeld and Haeck 1997; Polakow 1998; Cliff, Yeld and Hanslo 2004). It is important to illustrate that the tests work in selecting students from a diverse group of applicants who effectively engage with university tasks and go on to graduate. This emphasis on graduation becomes all the more central given the South African Department of Education's (DOE) focus on outputs as the benchmarks for continued higher education funding (SAUVCA 2004).

There is an entire body of research focusing on issues of student retention, dropout and throughput from HE (cf. Tinto 1975; Willett and Singer 1991; Pascarella and Terenzini 1991; Murtaugh, Burns and Schuster 1999). Investigations into retention, dropout and throughput are vital to academic planning and central to issues of the implied costs imposed on society, the institution and the individual (DesJardins 2002). Social costs include reduced economic output of non-graduates versus graduates. Predictably the dropout of 'disadvantaged students' may result in further racial and socioeconomic disparity in future generations. The costs to the institutions themselves is that imbalances between intake and graduation comes with major budgetary and financial implications. As mentioned above – the focus by the DOE on outputs puts significant pressure on minimising dropout. For the individual the costs of leaving the institution without graduating imply a loss in potential earnings, less preference in job choice, not to mention the personal and emotional issues related to dropout for academic reasons.

Predictive validity is defined as the extent to which a measure accurately forecasts how a person will think, act, or feel in the future. As applied to the university testing context it refers to the extent to which predictions can be made about the future academic potential of students using scores on a testing instrument tapping particular constructs. Predictive studies in the testing arena most often take the form of correlations and regression analyses. The use of standard statistical tools to measure the predictive validity of a testing instrument in a higher education environment seems relatively straightforward but this is seldom the case.

As reported by Yeld and Visser (2000) there are a number of design and measurement issues which arise in the course of using traditional methods to assess the impact of Academic Development initiatives such as the AARP testing service. Linn (1989) raises the issue of highly selective samples and points out that using

such a sample repeatedly yields a pessimistic view of predictive validity. This is especially true for the AARP where only a self-selected group of applicant writers actually gets admitted and can thus be used for predictive studies. A truncated sample, as described above, will tend to yield a different correlation coefficient in comparison to a sample which was randomly selected.

Studies into the predictive validity of assessment tools often make many hypotheses' which directly influence the outcomes of the study. One example is where regression and correlation studies investigating course level performance are often hampered by excessively small sample sizes as sample sizes diminish in the second and third year respectively. Hall (2003) makes the point that at UCT at present there are 9 786 possible course combinations for which undergraduate and postgraduate students are currently registered across all Faculties. There may be big samples at the first year level but further up the curriculum the samples become extremely truncated. Enlarging samples sizes by combining courses often leads to limitations in the strength of the findings due to the non-fit between course structure and curriculum assessment.

Using only correlation methods to assess the impact of testing instruments in a HE setting also masks the fact that courses change over the years in terms of content, structure, presentation and assessment.

Similarly academic development programmes and course interventions may have impacted positively on the performance of weak incoming students and this serves to further depress the correlation co-efficient when correlating testing instrument scores with course performance scores (Cliff, Hanslo, Herman, Fish, and Visser 2002). Ironically in cases such as this a weak correlation is proof of the success of academic development initiatives.

Defining the criterion of success also has an impact on the findings generated in predictive studies. Many evaluation studies use first year performance as the central criterion but this has been shown to be a poor predictor of success further up the curriculum (Griesel 1999). An alternate route is to simply model graduation or throughput in a simple tabular statistic. A complication that arises is that many students drop out of higher education studies due to non-academic reasons (financial, motivational, health etc.) – these numbers are quite significant and complicate predictive studies which use graduation as the criterion of success as they have not reached the destination event (Yeld and Visser 2000).

Current research by the AARP at UCT explicates the need to look beyond yearly performance and focus on throughput to graduation as a criterion (Polakow 1997, 1998, 1999). With this in mind the AARP has sought to investigate methods which supplement standard statistical methods and are better suited to the HE context.

NEW METHODOLOGIES TO ADDRESS OLD CONCERNS

A technique which overcomes many of the statistical predicaments suggested above is survival analysis. Survival analysis is a useful statistical technique for answering questions that deal with the duration of events. Survival analysis is variously known as time-to-event analysis, event history analysis and reliability analysis and was originally developed by biostatisticians modeling human lifetimes (Cox 1972; Kalbfleisch and Prentice 1980; Miller 1981). It is only fairly recently that this method has gained popularity in other fields such as institutional research where the focus is shifted from mortality to student retention (life cycle) and graduation rates (deaths) (Willett and Singer 1991; Huff and Fang 1999). A powerful advantage of survival analysis is that by constructing hazard models of students' careers, one can investigate not only whether particular groups of interest (e.g. stratified by Race or Gender) drop out, but also when they are most likely to do so (Boonzaaier 2000). The models can be used to study the relative risk of different groups of interest leaving university and they lend themselves to a variety of substantive questions such as: Are students more at risk of leaving during particular stages of their careers? Does the profile of risk differ among groups? To what extent do assessment instruments predict the risk of dropping out?

One of the key features of survival analysis is its usefulness in incorporating censored observations. Censoring occurs when an individual does not experience the event of interest. For the purposes of this article, a student is regarded as censored if they are not excluded due to academic reasons (e.g. a student who transfers to another university or who drops out due to financial difficulties). This type of censoring, where the true exclusion time is unknown, is known as right censoring. Students who were excluded due to graduation, those that left the institution due to financial difficulty or illness, or, those that transferred to other institutions/faculties were classified as censored.

Traditional statistical methods treat censored events with extreme caution. Censored cases are removed from the analysis and the average length of time until exclusion for the remaining cases is examined. This practice creates two problems. It may result in an unacceptably small data set and a biased sample that has a negative effect on the distribution of survival times. This is true in this study where most students never experience exclusion due to academic reasons. Another way to resolve the censoring problem is to impute event times for the censored cases by setting their unknown event times equal to the length of time they've been included in the study. However, imputation underestimates the survival time because the ultimate event times for censored cases are necessarily greater than the imputed value. In contrast to traditional statistical methods, survival analysis incorporates the censored cases and produces an accurate analysis of exclusion times.

EXPLORING SURVIVAL ANALYSIS

A number of assumptions are recognized when performing a survival analysis. The main assumption is that students continuing at university eventually qualify. Thus students who are still in the institution are classified as successful as they have not been excluded on academic grounds. This assumption is worthy of further investigation. Another assumption is the measurement of time. Time to exclusion is a continuous event. Students may be excluded or censored at any time during the year but this information is not readily available. Therefore progress codes, which are produced at the end of each year of study, are used and therefore time to exclusion is viewed as a discrete event.

Survival function

There are two functions that are used to describe survival times – the survival and the hazard function. Let t be the random variable representing time at university until exclusion. The survival function, $S(t)$, is the probability that a student will remain longer than time t at university and for a given population it will simply indicate what percentage of the population is still surviving at a particular time. Graphically, $S(t)$ is represented by a step function.

Consider a sample of n students who are observed until they are censored or excluded. The Kaplan-Meier estimate of survival is

where $S(t)$ — is the probability of surviving during year t , n_t is the number of students at the start of year t and d_t is the number of students who are excluded during year t .

Hazard function

Another measure of survival which allows a researcher to identify particularly precarious times for exclusion is the hazard function, $h(t)$, the probability that a student is excluded at time t , conditional on having remained at the institution up to that time. In the context of our study, it is the percent of students who start a particular year of study and are then excluded in that year. Mathematically,

$$h(t) = \frac{\text{No. excluded}}{\text{No. alive at the start of the year} - \frac{1}{2} \text{No. censored} - \frac{1}{2} \text{No. excluded}}$$

The hazard may increase, decrease, remain constant or show a more complicated process. Because the magnitude of the hazard indicates the risk

associated with each year it is a more convenient function to use compared to the survivor function. Kalbfleisch and Prentice (1980) provides a formal in-depth discussion of the survivor and hazard functions.

USING SURVIVAL ANALYSIS TO COMPUTE THE PREDICTIVE VALIDITY OF AN ADMISSIONS TEST AT UCT

The primary objective of the AARP is to develop a battery of relevant admissions tests that identify educationally talented students with the potential to succeed at UCT. One such test is the PTEEP (Placement Test in English for Educational Purposes). The PTEEP is an English language test and is written voluntarily by students who have applied for admission to any faculty on campus. The test incorporates a combination of multiple choice questions and productive pieces and includes elements of teaching, modeling, and practice. A table is included in the appendix of this article which highlights the skills specifications used in the construction of questions for the PTEEP.

Since the AARP is also tasked with increasing access to UCT the project distinguishes the differences in the educational backgrounds of particular students. The project classifies students with regard to their educational histories into a number of groups – the two most prominent being students who come from an ex-HOA (ex-House of Assembly, or, Model C) educational background and students who come from an ex-DET (ex-Department of Education and Training) educational background. During Apartheid, South African authorities encouraged an educational landscape based on a racially-defined skewing. Ex-DET schools which were attended by black South Africans were under-resourced and undeveloped. Conversely, the mainly white attended ex-HOA schools had access to vast pools of resources.

The project uses the PTEEP to make recommendations to Faculties about which students to select. To investigate the predictive validity of the PTEEP test, it was decided to compare the academic progress of the top performers to that of the poor performers. Students who wrote the PTEEP as applicants were classified as top performers if they obtained a score ranked in the top three deciles (PTEEP Top 30 per cent), and as poor/bottom performers if they obtained a score ranked in the bottom three deciles (PTEEP Bottom 30 per cent).

The top and bottom performers were also compared with a group which did not write the PTEEP but gained access to UCT via the traditional route. These students were admitted solely via meeting the points requirements of a cut-off totaled points score achieved from subjects taken in the Matriculation Certificate Examination or the equivalent thereof. This group will be referred to as the 'Non-PTEEP Group'. This allowed a comparison of the survival times and hazard function of the top PTEEP performers, the bottom PTEEP performers and a large control group of students who met the traditional School Leaving Examination Points score for admissions to study at UCT.

The aim of the study was thus threefold

- (a) A comparison of students from Ex-HOA and Ex-DET backgrounds, and whether they differ with regard to School Leaving Examination Points and PTEEP Scores.
- (b) A comparison of the Top 30 per cent of PTEEP performers with the Bottom 30 per cent of PTEEP performers with regard to their mean survival times and hazard rates over the duration of their time spent at UCT.
- (c) A comparison of the group of students who got access to UCT solely by means of their School Leaving Examination Points results (non-PTEEP group) with the groups mentioned in (b) above, with regard to their mean survival times and hazard rates over the duration of their time spent at UCT.

DESCRIPTION OF DATA SET

The data set consisted of 22 347 undergraduate students from all faculties who attended the University of Cape Town between 1995 to 2002. The 1995 cohort contributed eight years worth of data whereas the 2002 cohort only contributed one year of data. Only first degrees were considered. Students were tracked each year to determine whether they were allowed to progress to their next year of study or whether they were excluded from the University for academic reasons. Progress codes were obtained from a comprehensive student record database. Progress codes were separated into three groups – continued, excluded and withdrew in good standing. A student could withdraw from university having been allowed to progress but without graduating. These students were regarded as having withdrawn in good academic standing, that is, censored because they were not excluded on academic grounds. They were not eliminated from the analysis but included in the analysis for the length of time they remained at university.

Overall the group consisted of 22347 students, 3 698 ex-DET students and 18 649 ex-HOA students. The non-PTEEP group consisted of 18825 students, 2358 ex-DET students and 16467 ex-HOA students. A total of 3522 students wrote the PTEEP: 1340 (36 per cent) of ex-DET students and 2182 (11 per cent) of ex-HOA students wrote the PTEEP either during their application process or as registered students.

DESCRIPTIVE STATISTICS

Before performing the survival analysis, the ex-HOA group and ex-DET group were examined individually. Descriptive statistics for PTEEP performance and School Leaving Examination Points were calculated in order to determine whether the groups could be regarded as similar or not. Table 1 below provides descriptive statistics on PTEEP scores for the groups of interest.

The ex-HOA students outperform the ex-DET students, and for both ex-education groups, those who fall into the Top 30 per cent of their group perform

Table 1: Descriptive statistics on PTEEP score by ex-education department and PTEEP group

Ex-education group	PTEEP group (per cent)	Sample size	Mean PTEEP score	Standard deviation
Ex-DET	Top 30	620	56.80	8.02
Ex-DET	Bottom 30	184	28.20	7.72
Ex-HOA	Top 30	989	75.35	7.14
Ex-HOA	Bottom 30	311	49.37	10.12

Table 2: Descriptive statistics on school leaving examination points by ex-education department and PTEEP group

Ex-education group	PTEEP group (per cent)	Sample size	Mean school leaving examination score	Standard deviation
Ex-DET	Top 30	534	32.69	6.11
Ex-DET	Bottom 30	154	30.05	3.91
Ex-DET	non-PTEEP	2 061	32.28	5.35
Ex-HOA	Top 30	906	37.71	5.83
Ex-HOA	Bottom 30	288	32.67	4.13
Ex-HOA	non-PTEEP	12 789	38.32	5.84

much better than those who are categorised in the Bottom 30 per cent. The statistics clearly show the differential performance in these two groups and justify the need to look at these groups separately.

Table 2 above compares School Leaving Examination scores by ex-education group and PTEEP group. Ex-HOA students generally enter university with higher School Leaving Examination scores compared to ex-DET students. Once again this points to the need to examine survival in these groups separately. The gap in average School Leaving Examination marks between top and bottom PTEEP performers is larger for ex-HOA students compared to ex-DET students. And finally, there is not much difference between School Leaving Examination Points for PTEEP top performers and non-PTEEP students. Because PTEEP top performers and non-PTEEP School Leaving Examination performance is similar, one would expect that survival times would be similar as well. However, the survival analysis shown later clearly illustrate that PTEEP top performers survive longer than those who've entered university on the basis of their School Leaving Examination results only.

SURVIVAL ANALYSIS

The ex-DET set of data is presented in Table 3 below to illustrate in detail the Kaplan-Meier method of calculating survival. It consists of 3698 students who are followed up until exclusion from university due to academic reasons. Survival estimates are calculated at each time of exclusion, that is after each year of study at university.

Table 3: Kaplan-Meier estimates for ex-DET group

Survival time	Number at risk n_t	Number censored	Observed exclusions d_t	Probability of surviving (per cent)	Cumulative probability of surviving (per cent) $S(t)$
0	3698	0	0	100.00	100.00
1	3698	402	414	88.80	88.80
2	2882	306	369	87.20	77.43
3	2207	617	211	90.44	70.03
4	1379	608	119	91.37	63.99
5	652	328	52	92.02	58.88
6	272	177	18	93.38	54.99
7	77	60	1	98.70	54.27
8	16	16	0	100.00	54.27

The survival probability starts at 1 when all subjects are in the study. After the first year survival drops to 88.80 per cent, after the second year it drops to 77.43 per cent, to 70.03 per cent after the third year and it finally drops to 54.27 per cent by the end of the eighth year. The ex-DET Kaplan-Meier probabilities, as well as those for the ex-HOA group, are plotted as a solid step curve in Figure 1.

The survivor functions (and hazard function presented later) are graphically displayed as chains of short line segments since duration time is measured discretely as progress codes are obtained yearly.

It is clear that time to exclusion differs for the two groups. Survival estimates are not that different after the first year of study. However, as time increases, so does the gap between the ex-HOA and ex-DET group. By the end of the eighth year survival has dropped to 54 per cent in the ex-DET group but remains high at 80 per cent in the ex-HOA group. A nonparametric log-rank test revealed that the time to exclusion was significantly different for the two groups ($z=26.26$, $p<0.001$). Therefore survival for the PTEEP Top 30 per cent, PTEEP Bottom 30 per cent and non-PTEEP groups are examined separately for the ex-HOA and ex-DET groups, as they exhibit very different survival patterns at UCT.

The survival functions graphed in Figure 1 above does not allow the identification of particularly risky times for exclusion. Therefore consider the

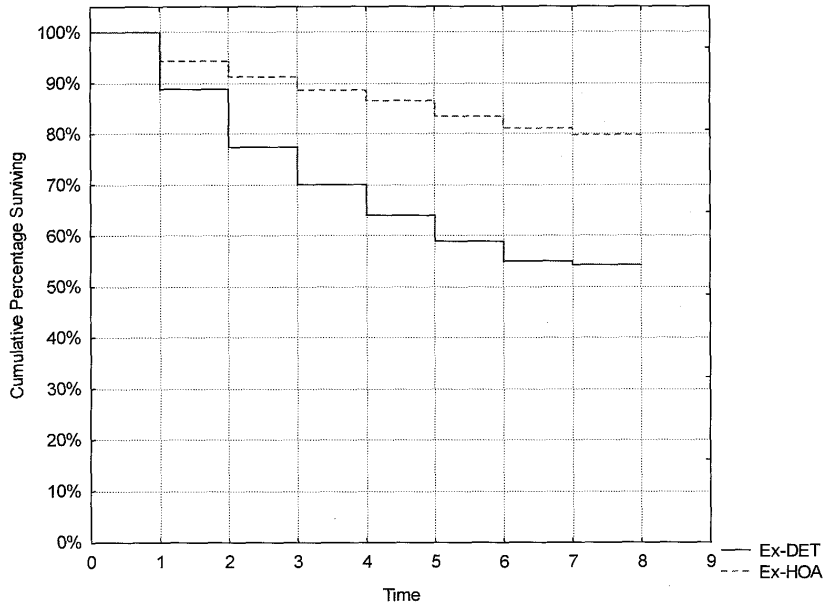


Figure 1: Survival functions by ex-education department

Table 4: Hazard estimates by ex-education department

Year	Ex-HOA	Ex-DET
1	6.45 (n 18649)	12.58 (n 3698)
2	3.68 (n 13720)	14.50 (n 2882)
3	3.93 (n 10536)	11.77 (n 2207)
4	3.55 (n 5638)	11.72 (n 1379)
5	5.36 (n 1510)	11.26 (n 652)
6	4.98 (n 466)	10.32 (n 272)

hazard function which reveals the times at which students are most at risk of being excluded. The empirical hazard function for the two ex-education groups is tabulated in Table 4 and graphed in Figure 2 below.

The magnitude of the hazard represents the risk of exclusion for each of the time periods. In particular for the DET group, the hazard of exclusion is 12.58 per cent in the first year of study indicating that 12.58 per cent of students are likely to be excluded in the first year of study. This rate increases and peaks to 14.50 per cent in the second year indicating that ex-DET students are most at risk of being excluded in their second year. The hazard then decreases until it reaches 2.15 per cent by the end of the seventh year. Because the sample becomes quite small towards the end of the study period these estimates become less reliable. The ex-HOA students are most at risk of being excluded in the first year of study.

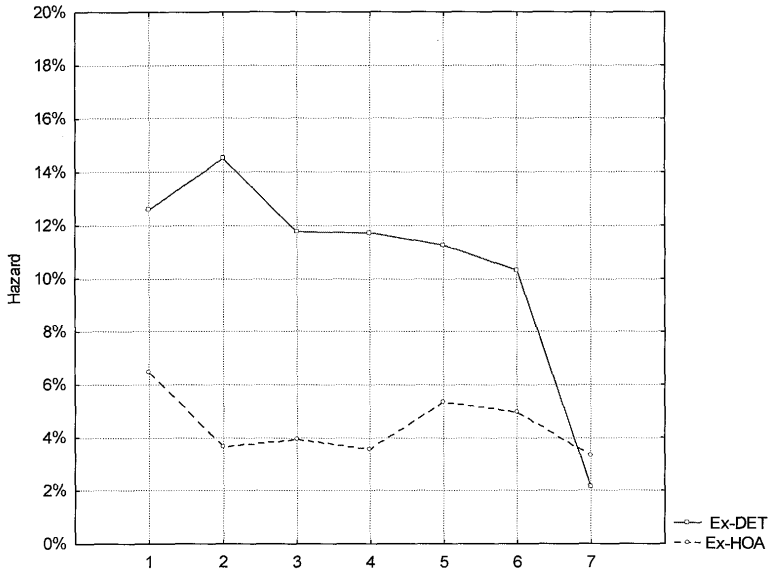


Figure 2: Hazard functions by ex-education department

Comparing the ex-education groups, the hazard of exclusion is consistently higher in the ex-DET group compared to the ex-HOA group. Quite interesting is the sharp decline in the hazard rate for the ex-DET group in the seventh year. It actually falls below the hazard rate for the ex-HOA group. However, the sample size is quite small by this point.

For brevity tables have been omitted and only figures are shown for the sections below. Tables are available from the authors.

Ex-HOA group

The figures which summarise survival for the ex-HOA group is given below. The ex-HOA group only wrote the AARP tests from 1997 onwards and therefore only contributes 6 years worth of information, regarding PTEEP performance, to the study.

Examination of the survival functions for the ex-HOA group reveals that the non-PTEEP group and the PTEEP Top 30 per cent group are quite similar. After the sixth year survival has dropped to the low eighties in the non-PTEEP and PTEEP Top 30 per cent groups respectively. However, the function for the Bottom 30 per cent group indicates that time to exclusion for this group is shorter than that for the non-PTEEP and PTEEP Top 30 per cent group. A log-rank test to compare multiple groups revealed that the groups differ with respect to exclusion time ($\chi^2=43.63$, $p<0.001$).

The hazard functions for the ex-HOA group support the inferences from the survival functions. The hazard rates are very similar for the non-PTEEP group and

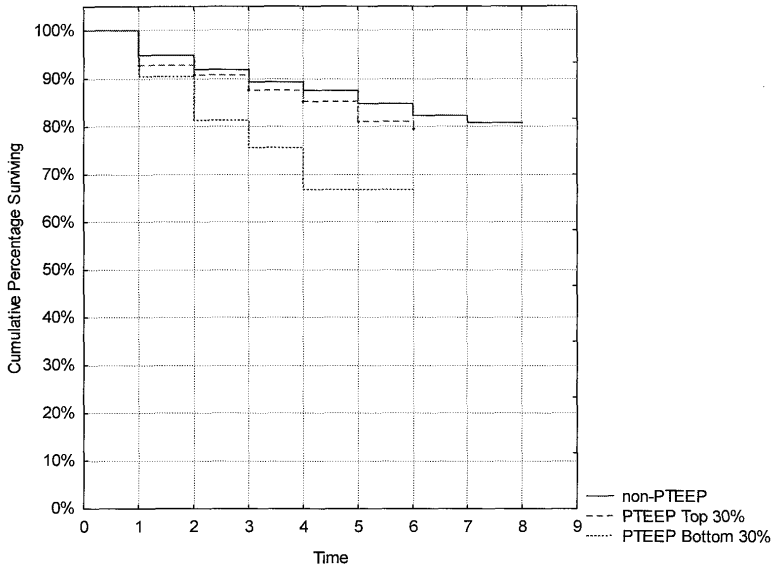


Figure 3: Survival functions of ex-HOA students by PTEEP group

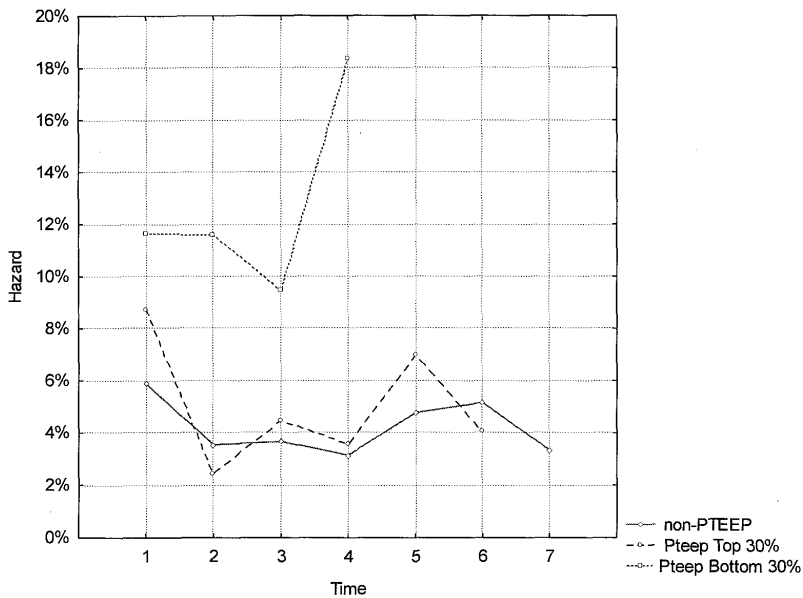


Figure 4: Hazard functions of ex-HOA students by PTEEP group

PTEEP Top performers. However, the probabilities of exclusion is higher for the PTEEP Bottom performers, with the probabilities peaking during the fourth year of study indicating that a PTEEP Bottom performer is most likely to be excluded at this point of their studies. Top performers and non-PTEEP students are most likely to be excluded within the first year of study.

Ex-DET group

Now consider survival for the ex-DET group. The probabilities of survival after each year at university is plotted in Figure 5 below.

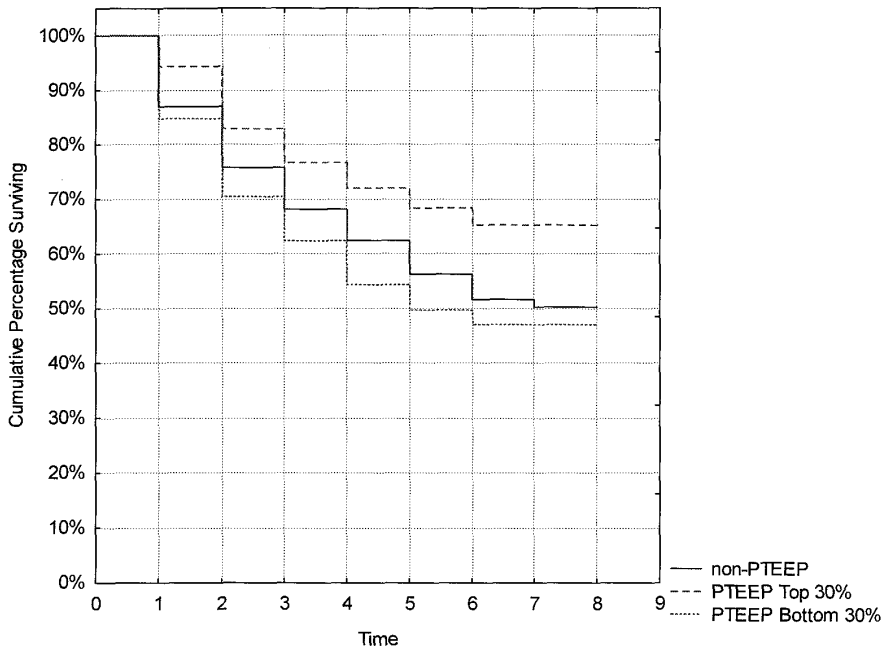


Figure 5: Survival functions of ex-DET students by PTEEP group

The patterns of survival for the PTEEP and non-PTEEP groups differ from the ex-HOA groups'. The differences in time to exclusion of the non-PTEEP, PTEEP Top 30 per cent and Bottom 30 per cent groups are more apparent in the ex-DET analysis compared to the ex-HOA group. The Top 30 per cent group survives longer at university compared to the other two groups of students. In fact, the gaps between probabilities of exclusion appear to widen as time increases. As expected, bottom performers remain a shorter time in the institution before being excluded for academic reasons. The median survival time for this group is 5 years as opposed to 7 years for non-PTEEP students and >8 years for top PTEEP performers. The difference in survival functions of the three groups is statistically significant ($\chi^2=30.84, p<0.001$). This pattern of survival is quite interesting in the light of information contributed by the School Leaving Examination Points system. At the outset of their studies, the top achievers in the PTEEP test and the non-PTEEP students appear to be performing quite similarly when examining School Leaving Examination Points only (see Table 2 above). However, using a survival analysis approach, when one examines survival over time, those identified by the PTEEP as top performers survive longer at university compared to those

who did not write the PTEEP test. Therefore, the university has a better likelihood of identifying a student's risk if the PTEEP is written as opposed to only considering the school leaving points as an access mechanism.

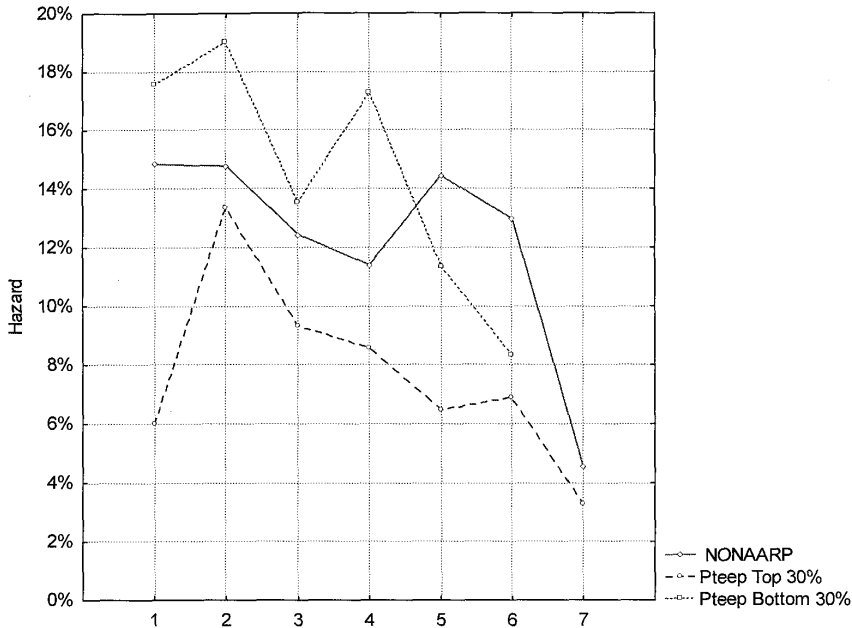


Figure 6: Hazard functions of ex-DET students by PTEEP group

Quite interesting patterns can be seen in the hazard functions for ex-DET students. Generally, the hazard for the top performers of the PTEEP is the lowest over all years. These students are most at risk of leaving during their second year of study but the risk gradually decreases after the second year. PTEEP bottom performers are most likely to be excluded in the second year a significant proportion of students are excluded due to poor academic performance. The non-PTEEP groups' hazard peaks after the first, second and fifth year of study, all indicating increased risk of exclusion.

CONCLUSION

The use of a survival analysis approach to investigating predictive validity overcomes many of the pitfalls associated with traditional approaches such as correlations and regression analyses. Most importantly it allows for the incorporating of censored data which would typically be omitted from a study should standard statistical methods be used.

In this study the survival analysis approach provided useful insight into the attrition patterns of students at the University of Cape Town. It not only answered

questions regarding whether students drop out but also, through the use of the hazard function, it was able to illustrate when the periods of risk were highest. This is particularly useful in the university setting where it is vital to assess when students are most at risk of exclusion from an institution.

By stratifying the population of students into groups of interest, the ex-DET and ex-HOA groups, as well as PTEEP top performers, PTEEP poor performers, and traditional access students, could be compared with regard to survival time at the university.

The initial descriptive statistics and survival analysis by ex-education department suggested looking at the two groups, ex-HOA group and ex-DET group, separately rather than grouping them together. The results of the analysis indicate that even though students might come to university with very similar School Leaving Examination scores, the PTEEP is able to provide useful additional information regarding risk of exclusion especially in the ex-DET group of students, where top PTEEP performers clearly display a lower likelihood of being excluded compared to the bottom PTEEP performers.

The study showed that ex-DET students were most at risk of exclusion during their second year of study, and that ex-HOA student were typically at risk during their first year. However, poor PTEEP performing ex-HOA students differed from top performers and non-PTEEP students in that their hazard of exclusion peaked during the fourth year of study. It may be worthwhile investigating this group to determine possible reasons for this late increase.

The analysis also points to PTEEP performance being as good a predictor of risk of dropout as the traditional Senior Certificate School Leaving Examination Points performance. For the ex-DET group, the PTEEP appears to be an even better predictor when compared to Senior Certificate School Leaving Examination Points.

APPENDIX

The following table shows the specifications on which the PTEEP is based followed by a brief description of each specification area (Cliff, A., N. Yeld and M. Hanslo. 2005.).

Skill assessed	Explanation of skill
Vocabulary	Students' abilities to derive/work out word meanings from their context
Metaphorical Expression	Students' abilities to understand and work with metaphor in language. This includes their capacity to perceive language connotation, word play, ambiguity, idiomatic expressions etc.
Extrapolation, application and inferencing	Students' capacities to draw conclusions and apply insights, either on the basis of what is stated in texts or is implied by these texts.

Skill assessed	Explanation of skill
Understanding the communicative function of sentences	Students' abilities to 'see' how parts of sentences/discourse define other parts; or are examples of ideas; or are supports for arguments; or attempts to persuade
Understanding relations between parts of text	Students' capacities to 'see' the structure and organisation of discourse and argument, by paying attention within and between paragraphs in text to transitions in argument; superordinate and subordinate ideas; introductions and conclusions; logical development
Understanding text genre	Students' abilities to perceive 'audience' in text and purpose in writing, including an ability to understand text register (formality/informality) and tone (e.g. didactic/informative/persuasive)
Separating the essential from the non essential	Students' capacities to 'see' main ideas and supporting detail; statements and examples; facts and opinions; propositions and their arguments; being able to classify, categorise and 'label'
Understanding information presented visually	Students' abilities to understand graphs, tables, diagrams, pictures, maps, flow charts
Understanding basic numerical concepts	Students' abilities to make numerical estimations; comparisons; calculate percentages and fractions; make chronological references and sequence events/processes; do basic computations
Vocabulary	Students' abilities to derive/work out word meanings from their context

REFERENCES

- Boonzaaier, C. 2000. Survival analysis – a useful tool. *Southern African Association for Institutional Research (SAAIR) Newsletter 2 (2)*:1.
- Cliff, A., N. Yeld and M. Hanslo. 2005. Assessing the academic literacy skills of entry level students, using the Placement Test in English for Educational Purposes (PTEEP). Assessment in Education (Under review).
- Cliff, A., M. Hanslo, C. Herman, W. Fish and A. Visser. 2002. Unpublished University of Cape Town Science Faculty Report August 2002.
- Cox, D. R. 1972. Regression models and life tables. *Journal of the Royal Statistical Society 34 (Series B)*: 187–202.
- DesJardins, S. 2002. Studying the timing of student departure from college. Paper presented at the Annual Forum of the Association for Institutional Research (AIR).
- Griesel, H. 1999. Access and the higher education sector. A South African case study on innovative policy and programmes. Draft discussion document for the Eastern and Southern African Regional and National Seminars, 15–19 and 22 June 1999. Prepared for the Department of Education and the Association for the Development of Education in Africa (ADEA).
- Hall, M. 2003. Personal communication. UCT, Centre for Higher Education Development, Board Meeting.

- Hill, C. 2002. Linguistic and cultural diversity: a growing challenge to American higher education. *CSE Technical Report 556*. National Center for Research on Evaluation, Standards and Student Testing (CRESST), UCLA.
- Huff, K. and D. Fang. 1999. When are students most at risk of encountering academic difficulty? A study of the 1992 matriculants to US medical schools. *Academic Medicine* 74 (4): 454-460.
- Kalbfleisch, J. and R. Prentice. 1980. *The statistical analysis of failure time data*. New York: Wiley.
- Linn, R. L. (ed.). 1989. *Educational measurement*. (3rd Edition). New York: ACE and Macmillan.
- Miller R. G. 1981. *Survival analysis*. New York: Wiley.
- Murtaugh, P., L. Burns and J. Schuster. 1999. Predicting the retention of university students. *Research in Higher Education* 40 (3): 355-371.
- Pascarella, E. and P. Terenzini. 1991. *How college affects students: findings and insights from twenty years of research*. San Francisco: Jossey Bass.
- Polakow, D. 1997. *Analysis of the predictive capacity of AARP testing: The 1996 cohort*. Report prepared for the Academic Development Programme November 1997, University of Cape Town.
- . 1998. *Analysis of the predictive capacity of AARP testing: The 1996 cohort second year*. Report prepared for the Academic Development Programme August 1998, University of Cape Town.
- . 1999. *Survival analysis of university tenure, with foci on contrasting probability of exclusion in AARP and non AARP ex DET student groupings*. Report prepared for the Academic Development Programme November 1999, University of Cape Town.
- SAUVCA see South African Vice Chancellors' Universities Association.
- South African Vice Chancellors' Universities Association. 2004. Presentation to the Parliamentary Committee on Education by the South African Vice Chancellors' Association, 1 June 2004. <http://www.pmg.org.za/docs/2004/appendices/040601sauvca.pdf>.
- Tinto, V. 1975. Dropout from higher education: a theoretical synthesis of recent research. *Review of Educational Research* 45:89-125.
- Willett, J. and J. Singer. 1991. From whether to when: New methods for studying student dropout and teacher attrition. *Review of Educational Research* 61:407-450.
- Yeld, N. and W. Haeck. 1997. Educational histories and academic potential: can tests deliver? *Assessment and Evaluation in Higher Education* 22:5-16.
- Yeld, N. and A. Visser. 2000. Impact assessment in academic development. Paper presented at the South African Academic Development Conference, Rhodes University.
- Yeld, N. 2001. *Assessment, equity and language of learning: key issues for higher education selection in South Africa*. Unpublished PhD Thesis, University of Cape Town.