

# Can 'at risk students' be identified based on demographics?

Charles Hacker<sup>a</sup>, Andrew Seagar<sup>b</sup>, and Paul Cassidy<sup>c</sup>  
*Griffith School of Engineering, Griffith University<sup>abc</sup>*  
*a.seagar@griffith.edu.au*

---

## Structured abstract

### BACKGROUND

Student retention and successful completion has become more important to Australian universities, in terms of academic outcomes, quality accountability, and national university rankings. The highest student attrition traditionally occurs within the first year of university study. To improve student first year retention and academic success, early warning and support mechanisms are usually utilised to address student academic performance issues.

The success of students in their first year of university can be quite variable, and it is often difficult to anticipate which students require additional support. These students are classified as 'at risk' students. The 'at risk' students can be identified by monitoring the students' academic performance within early semester assessment items. Students with poor performance in these assessment items can be identified, and faculty and student advisors can better provide these students with support mechanisms in order to minimise the likelihood of student performance issues.

### PURPOSE

The purpose for this study is to determine if other indicators are available to further identify these 'at risk' students. For early identification, the indicators present before students start their university degree will be investigated. The beginning step in this study is to determine if 'at risk' students can be identified based on the student demographic information, that the student provides when enrolling at the university. The chosen demographic subgroups of this study were: age, gender, residency, entry ranking, and chosen engineering discipline.

### DESIGN/METHOD

The investigation determined if student performance issues could be identified from these available demographics. The method adopted was to perform grade statistical analysis on demographic subgroups, from a single population of all first year engineering students. That is, the study consists of tracking the grades within the eight common first year subjects, that all first year engineering students undertake, within all engineering disciplines. Statistical significance tests were carried out for each tested demographic, to determine if poor grades are attributable to certain demographic groups.

### RESULTS

The analysis revealed that neither gender nor residence status of the students were significant in detecting 'at risk' students. The analysis did conclude that both student age and student entry ranking was significant in detecting 'at risk' status. In this case, students who were younger and students who had poorer entry ranking, were more likely to be 'at risk'. The analysis also revealed that students who had not pre-selected an engineering discipline, were more likely to be 'at risk'. Finally, there was no appreciable performance issues between students of different disciplines.

### CONCLUSIONS

It was found that for some of the particular demographics, significant indicators were observed, whereas no significant indicators were observed for others. Based on the study's outcome, it is possible to identify some groups of students, before the start of their studies, who will be most likely be in need of additional support in order to succeed in their first year of university.

### KEYWORDS

Retention rates, student attrition, identifying at risk students, student demographics

---

## Introduction

Student retention (Jamelske, 2009; Scott et al., 2008; Thomas, 2002; Tinto, 2010) and successful completion (Tumen et al., 2008; Rickinson, 1998; McMillan, 2008) has become more important to Australian universities, in terms of academic outcomes, quality accountability, and national university rankings. The highest student attrition traditionally occurs within the first year of university study (Shah et al., 2010, Weldegiorgis, & Awel, 2013; Franssen, & Nijhus, 2011). To improve student first year retention and academic success, early warning and support mechanisms are usually utilised to address student academic performance issues (McKenzie, & Schweitzer, 2001).

The success of students in their first year of university can be quite variable. Often it is difficult to anticipate which students require additional support. Many universities attempt to identify the first year students who are struggling with their studies (classified as 'at risk'). Currently the authors' university implements the 'amber alert' system, that monitors students performance in early semester assessment items. Students with poor performance in these assessment items are identified, so that the faculty and student advisors can better provide these students with support mechanisms, to minimise the likelihood of student performance issues.

Relying solely on detecting 'at risk' students based on student performance in early assessment items, may not be sufficient. It could be the case that students who perform poorly in their initial assessment items, 'give up' and 'drop out' of university, before these students can be targeted to receive the various academic support mechanisms available. A method of detecting these 'at risk' students, before students become discouraged by poor grades, would be beneficial.

The purpose for this study is to determine if other indicators are available to further identify these 'at risk' students. For early identification, the indicators present before students start their university degree will be investigated. The beginning step in this study is to determine if 'at risk' students can be identified based on the student demographic information (Andres, & Carpenter, 1997; Okpala, 2002) that the student provides when enrolling at the university. Here this determination is made *a posteriori* from student grades over the course of the year. Based on the results it will be possible in future to reverse the process and use these indicators in subsequent years in a predictive manner to identify 'at risk' students *a priori*.

## Method

Students starting university provide a variety of information to the university, such as their name, address, date of birth, gender, and entry ranking. Numerous student demographic data can be inferred from this student enrolment information. For this study, the six chosen demographic subgroups chosen were: age, gender, residency (domestic and international student), chosen engineering discipline, and entry ranking (either OP or TER score). Note that a highly performing student is ranked with a low numerical value OP score or a high numerical value TER score, according to whether they finished high school in Queensland or NSW respectively.

The investigation will determine if student performance issues could be identified from these available demographics. The method adopted was to perform statistical analysis on the final grades obtained in 2012 by all the first year engineering students, and to determine if poor grades were more significant within certain demographic groups.

The ability to evaluate all first year engineering students' results is possible at the authors' university, since these first year engineering students undertake a common first year. That is, although the university offers engineering disciplines in Civil Engineering (CV), Mechanical Engineering (ME), Mechatronic Engineering (MT), and Electrical and Electronic Engineering (EE), the first year students undertake common first year subjects. The eight common first year subjects being: Mathematics 1A, Mathematics 1B, Computing and Programming with

MATLAB, Electronics for Engineers, Design and Professional Skills, Fundamentals of Engineering, Engineering Mechanics, and Engineering Materials.

The analysis of the data consisted of three separate tests. The first being a binary *t*-test on the students Gender (male/female) and Residency status (domestic/international). The second being a linear regression on student Entry Ranking and Age, with associated *t*-test on the linearity. Finally, was a multidimensional *Hotelling's Test* on student grades with student chosen engineering discipline (multidimensional due to the eight separate subject grades being a separate dimension).

Each of the binary *t*-tests aims to determine if statistical measures (mean and standard deviation) from two chosen groups of samples taken from the entire cohort show a statistically significant difference under the assumption that the samples in each group are distributed in a normal (Gaussian) fashion. The linear regression *t*-tests aim to determine if a linear relationship, fitted in a least squares sense, has a slope which shows a statistically significant difference from zero (under the assumption that the samples are normally distributed on either sides of the line). The Hotelling's test aims to determine if statistical measures of position and spread in multidimensional data-space (here 8 dimensions) from discipline based groups of samples taken from the entire cohort show a statistically significant difference from one another (again with a assumption of normally distributed samples).

## Binary T-tests

The binary demographic groups of student gender (being male or female) and student residency (being domestic or international) was statistically tested against the average student marks from all subjects. In this test, the entire student sample, from all disciplines, was split into the two specific subgroups. A *Student's t*-test was performed to determine whether the specific demographic groups had a statistically significant difference in their average mark. The test was to determine if the result was statistically significant to either a level of 0.05 or a level of 0.01.

### Gender test

The total student population was  $N = 180$ . This consisted of 172 male and 8 female students. For this test a 0.05 significance occurs if the *t*-value is over 1.973, and a 0.01 significance occurs if the *t*-value is over 2.604. The analysis gave a *t*-value of 0.420. This indicated that the average mark variation between male and female students was not significant at either level.

### Residency test

The total student population was  $N = 180$ . This consisted of 170 domestic students and 10 international students. For this test a 0.05 significance occurs if the *t*-value is over 1.972, and a 0.01 significance occurs if the *t*-value is over 2.602. The analysis gave a *t*-value of 0.787. This indicated that the average mark variation between domestic and international students was not significant.

## Linear regression tests

In the linear regression tests, the demographics were utilised as the independent variable, and the final overall student mark was utilised as the dependant variable. A linear least squares line was fitted to the data, and the analysis was to determine if the slope of the line was significantly different from zero. A *Student's t*-test was performed to test the significance.

The demographics tested with linear regression were the students' entry ranking (OP and TER) with overall student mark, and the student age with overall student mark. It was suspected that the linear regression results could be effected, or skewed, by a large group of low performing students (those getting less than 50%). Thus the *t*-test for linear regression

was performed on the entire student group, as well as the student group gaining an overall mark of 50% or more, and finally on the student group gaining less than 50% overall.

### Student entry ranking (OP and TER) with overall examination performance

The student university entry ranking (OP and TER) score, was expected to be a clear indicator of student performance within their academic first year. The authors' university attracts students from both Queensland and New South Wales. These students obtained OP and TER university entrance rankings (respectively).

To test the level of significance of the entrance ranking, the average overall mark students obtained (for all eight first year subjects), were graphed against the individual student entry ranking. Two graphs result, due to each (OP or TER) ranking.

### Student 'OP' ranking test

The average student mark (from all the eight subjects) with student OP score is graphed in Figure 1. Note that lower OP scores represent higher performing students. Inspection of the figure suggests that (in general) students entering with poorer OP levels, obtain lower overall subject marks, and are more likely to be 'at risk'.

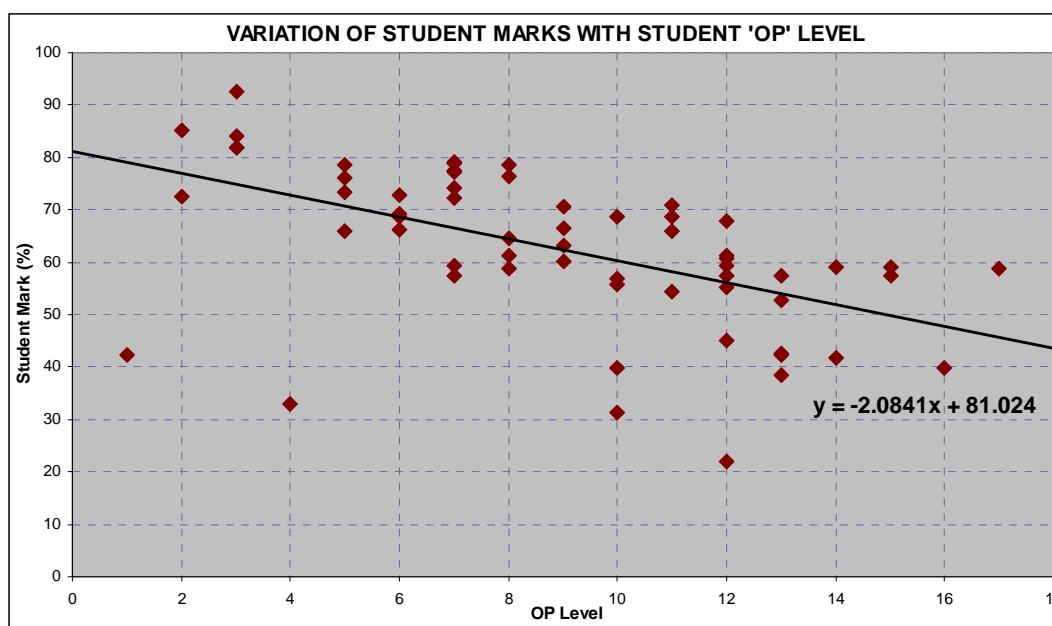


Figure 1: Variation of student marks with 'OP' level

### Entire group regression T-test

The linear regression equation:  $y = -2.084 \times X + 81.02$  was tested for significance. The total population was  $N = 61$ . For this test a 0.05 significance occurs if the t-value is over 2.000, and a 0.01 significance occurs if the t-value is over 2.661. The analysis gave a t-value of 4.950. This indicated the entire group performance was related to their OP score, to a significance level of more than 0.01. The negative slope of the line is consistent with the higher value of OP score being a measure of lower performance.

### Passing student regression T-test

When only the students who gained an overall mark of 50% or more is included in the regression, the linear regression equation is:  $y = -1.9811 \times X + 84.69$ . The total population was  $N = 50$ , and for this test a 0.05 significance occurs if the t-value is over 2.010, and a 0.01 significance occurs if the t-value is over 2.682. The analysis gave a t-value of 7.727. This indicated the performance from this group of students was related to their OP score, to a significance level of more than 0.01.

### Failing student regression T-test

When only the students who gained an overall mark of less than 50% is included in the regression, the linear regression equation is:  $y = 0.139 \times X + 36.53$ . The total population was  $N = 11$ , and for this test a 0.05 significance occurs if the t-value is over 2.262, and a 0.01 significance occurs if the t-value is over 3.249. The analysis gave a t-value of 0.278. This indicated that for this group of students there was no relationship between their performance and their OP score.

### Student 'TER' ranking test

The average student mark (from all eight subjects) with TER score is graphed in Figure 2. Again, the graph of Figure 2 suggests that (in general) students entering with poorer TER levels, obtain lower overall subject marks, and are more likely to be 'at risk'.

### Entire group regression T-test

The linear regression equation:  $y = 0.806 \times X - 4.518$  was tested for significance. The total population was  $N = 92$ . For this test a 0.05 significance occurs if the t-value is over 1.986, and a 0.01 significance occurs if the t-value is over 2.631. The analysis gave a t-value of 4.800. This indicated the entire group performance was related to their TER score, to a significance level of more than 0.01. The positive slope of the line is consistent with the higher value of TER score being a measure of higher performance. Note that although the slope of the line in Figure 2 is different from that in Figure 1, that is because of the different way the OP and TER rankings attribute their numerical scores to student performance. The two set of results and the two figures are in agreement, and confirm that the OP and TER ranking scores are a representative measure and predictor of student performance.

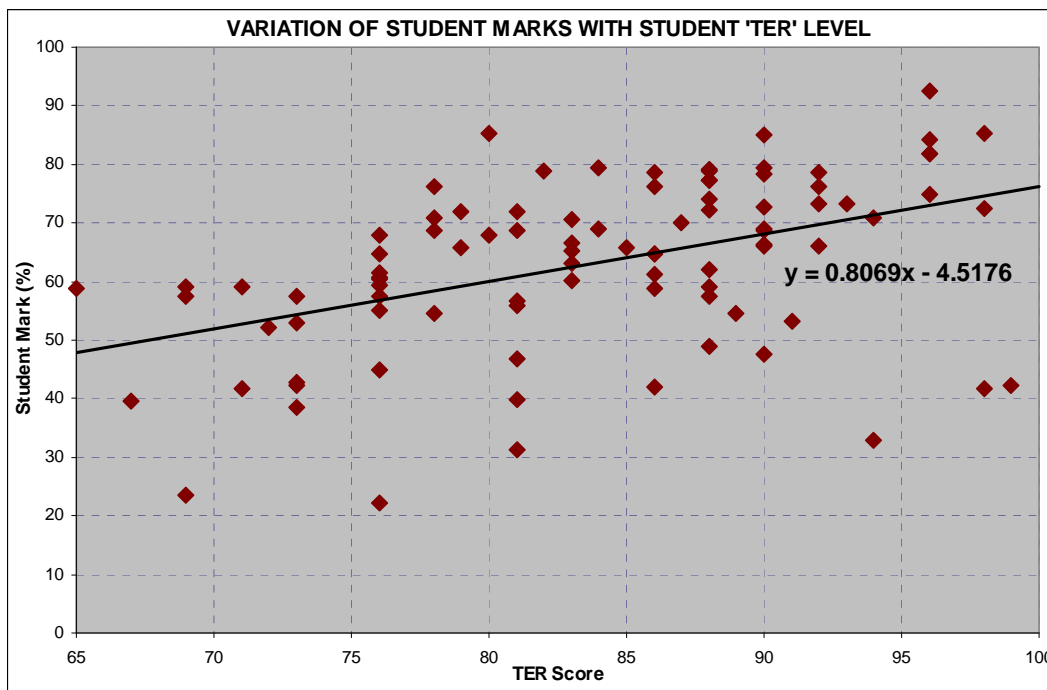


Figure 2: Variation of student marks with 'TER' level

### Passing student regression T-test

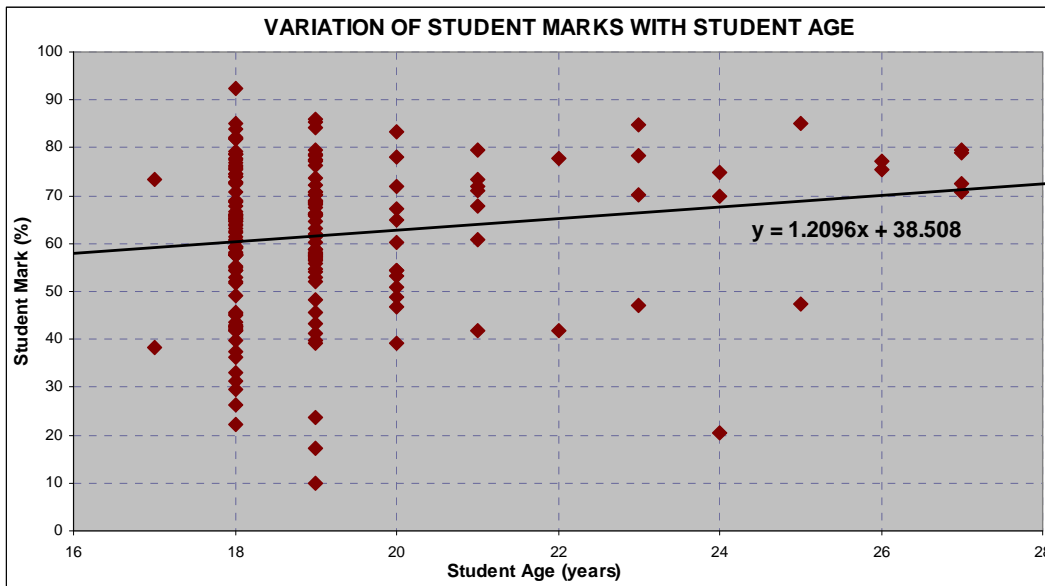
When only the students who gained an overall mark of 50% or more is included in the regression, the linear regression equation is:  $y = 0.737 \times X + 6.165$ . The total population was  $N = 75$ , and for this test a 0.05 significance occurs if the t-value is over 1.992, and a 0.01 significance occurs if the t-value is over 2.644. The analysis gave a t-value of 6.320. This indicated the performance from this group of students was related to their TER score, to a significance level of more than 0.01.

### Failing student regression T-test

When only the students who gained an overall mark of less than 50% is included in the regression, the linear regression equation is:  $y = 0.207 \times X + 22.565$ . The total population was  $N = 17$ , and for this test a 0.05 significance occurs if the t-value is over 2.131, and a 0.01 significance occurs if the t-value is over 2.946. The analysis gave a t-value of 1.095. This indicated that for this group of students there was no relationship between their performance and their TER score.

### Student age with average examination performance

Before the study, it had been long suspected that mature age students were generally higher academically performing students. This anecdotal assumption was tested, by analysing the students overall final first year academic results with student age. The average mark obtained by the students for the eight first year subjects, are graphed against the individual students age in **Figure 3**.



**Figure 3: Variation of Average Student Marks with Student Age**

Although the graph of Figure 3 does give the indication that older (mature age) students are higher performing, the statistical trend line revealed the relationship was of minor importance, due to the line having minimal slope (a 1.2 percentage mark increase for every year of age).

### Entire group age performance T-test

The linear regression equation:  $y = 1.209 \times X + 38.508$  was tested for significance. The total population was  $N = 180$ . For this test a 0.05 significance occurs if the t-value is over 1.973, and a 0.01 significance occurs if the t-value is over 2.603. The analysis gave a t-value of 2.129. This indicated the entire group performance was related to their age, to a significance level of at least 0.05.

### Passing student age performance T-test

When only the students who gained an overall mark of 50% or more is included in the regression, the linear regression equation is:  $y = 1.124 \times X + 46.320$ . The total population was  $N = 142$ , and for this test a 0.05 significance occurs if the t-value is over 1.977, and a 0.01 significance occurs if the t-value is over 2.611. The analysis gave a t-value of 2.986. This indicated the performance from this group of students was related to their age, to a significance level of at least 0.05.

### Failing student age performance T-test

When only the students who gained an overall mark of less than 50% is included in the regression, the linear regression equation is:  $y = 0.203 \times X + 34.558$ . The total population was  $N = 38$ , and for this test a 0.05 significance occurs if the t-value is over 2.028, and a 0.01 significance occurs if the t-value is over 2.719. The analysis gave a t-value of 0.229. This indicated that for this group of students there was no relationship between their performance and their age.

## Multidimensional Hotelling's Test

The *Hotelling's Test* is similar to the *Student's t-test*, but in this case the test allows for each sample to be a multidimensional value. In this study there are eight dimensions, represented by a subject result from each of the eight common first year subjects of the degree program. The test provides a 'T' squared value ( $T^2$ ), that can be utilised as a test of significance. However the method does not utilise the 'T' squared value directly, but instead the value is transformed into a different statistical distribution, called the 'F distribution'. This new distribution provides an 'F' value, that is utilised to test for significance of a hypothesis.

The Hotelling's Test was utilised for testing if there was significant student performance differences between each separate engineering discipline. The engineering disciplines consisted of: Civil Engineering (CV), Mechanical Engineering (ME), Mechatronic Engineering (MT), and Electrical and Electronic Engineering (EE). The common first year also allowed students to enter the engineering degree, without having to yet determined their discipline area (undecided students).

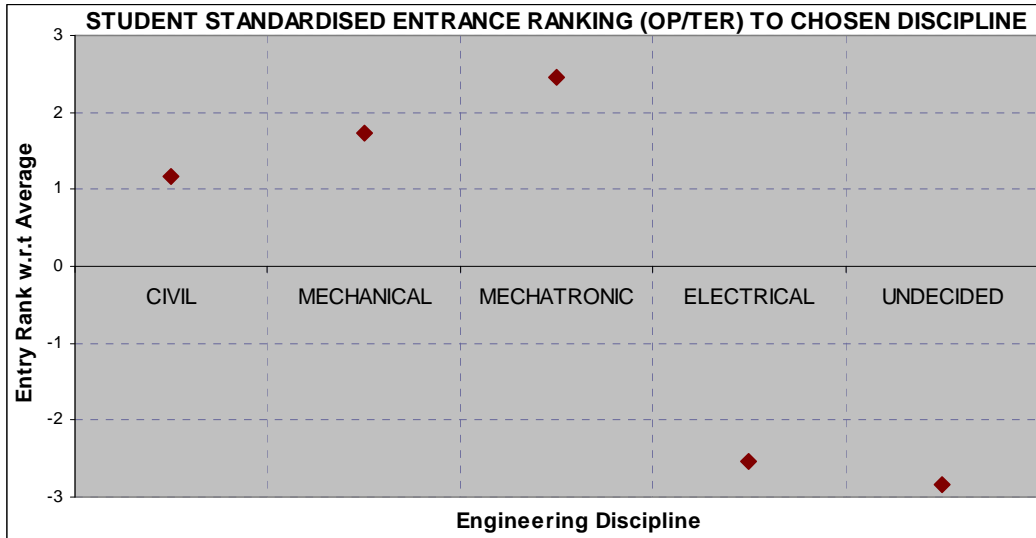
### Entrance ranking and discipline

Since lower entry rankings have been demonstrated (in the above 'Linear Regression Tests') to be a factor for detecting 'at risk' students, it was necessary to determine if lower entry ranking students were more likely to select a specific engineering discipline over another.

The average entry ranking of the students selecting each engineering discipline is given in graph of Figure 4. As indicated from the graph, the entry ranking data revealed that the average entry scores (OP and TER) of the students beginning the first year engineering disciplines, were different between each of the selected disciplines. This suggests a hypothesis that the final average student performance (in the common first year subjects) will be statistically different between students that have selected different engineering disciplines.

Inspection of Figure 4 indicates students pre-selecting the Civil Engineering, Mechanical Engineering, and Mechatronic Engineering disciplines were students with generally higher entrance level scores. It also suggests that students pre-selecting the Electrical and Electronic Engineering, and those students who had not yet chosen their discipline (undecided), could be more likely to have performance issues in their first year studies.

Before proceeding to test that hypothesis it was necessary to first determine if mature age students were more likely to select a particular engineering discipline over another. It has already been determined that, although of minor importance, age affects the students final results and this could bias any results. Inspection of Figure 5 indicates the average age of

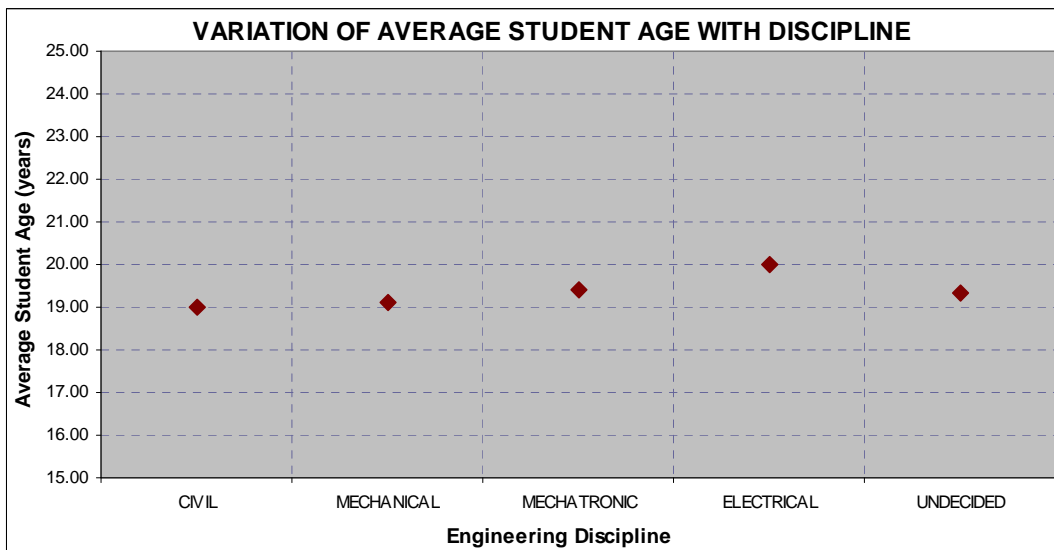


**Figure 4: Student standardised entrance ranking to chosen discipline**

the students within each engineering discipline was between 19 to 20 years of age. It was thus determined that there is not a significant difference in average student age between selected disciplines, and that the results obtained from each discipline would not be biased by the effects of varying student age.

**Comparison of selected disciplines to the undecided students**

To test the hypothesis that the final average grades of the undecided discipline students, will be different to those students who have chosen a discipline, the Hotelling’s Test was performed. In these tests, one demographic group was the students that chose a specific discipline (Civil, Mechanical, Mechatronic, and Electrical), and the other group was the undecided discipline students. The statistical data results are presented in a condensed form within Table 1.



**Figure 5: Variation of Average Student Age with Discipline**

**Table 1: Comparison of Students choosing a Discipline with Undecided Students**

Discipline Comparison		Numbers		Significance level			
		N1 - N2	t2 Value	F Value	0.05	0.01	RESULT
Civil - Undecided	CV - UD	71 - 15	29.11	3.33	2.061	2.751	0.01 Significance



Mechanical - Undecided	ME - UD	58 - 15	42.98	4.48	2.086	2.802	0.01 Significance
Mechatronic - Undecided	MT - UD	18 - 15	31.36	3.03	2.355	3.362	0.05 Significance
Electrical - Undecided	EE - UD	11 - 15	36.16	3.20	2.547	3.790	0.05 Significance

The results show that for all the students that have decided a discipline, the average marks these students obtained is significantly different to the marks obtained by the undecided discipline students.

### Comparison of selected disciplines to the electrical discipline

To test the hypothesis that the final average grades of the Electrical and Electronic (EE) engineering discipline students will be statistically different to the students choosing other disciplines, the Hotelling's Test was again performed. The condensed statistical data results are presented in Table 2.

The results of Table 2 show that there is no significant difference in the average marks obtained by the students choosing the Electrical and Electronic (EE) discipline, with those students choosing other disciplines. Therefore, although the average entry ranking of the students choosing the Electrical and Electronic (EE) discipline is lower than for the other disciplines, the average final marks of these students do not vary significantly from the other disciplines.

**Table 2: Comparison of students choosing the electrical discipline with others**

Discipline Comparison		Numbers			Significance level		
		N1 - N2	t2 Value	F Value	0.05	0.01	RESULT
Civil - Electrical	CV - EE	71 - 11	10.99	1.25	2.067	2.765	Not Significant
Mechanical - Electrical	ME - EE	58 - 11	10.47	1.17	2.096	2.823	Not Significant
Mechatronic - Electrical	MT - EE	18 - 11	9.86	0.91	2.447	3.564	Not Significant

### Finalising comparison between selected disciplines

The *Student Standardised Entrance Ranking* graph (Figure 4) suggests the grade performances between students undertaking the Civil, Mechanical, and Mechatronic disciplines should be similar. To test this hypothesis, the Hotelling's Test is continued on the average student results between each of these remaining disciplines. The results of which are given in Table 3.

As indicated in Table 3, the only disciplines that show a significance (in the eight multidimensional subject marks) between another discipline, is the students undertaking the Civil Engineering and the Mechatronic Engineering discipline. It should be noted that due to the test being performed on the eight multidimensional first year subject marks, this test only

**Table 3: Comparison of the remaining Discipline Choosing Students**

Discipline Comparison		Numbers			Significance level		
		N1 - N2	t2 Value	F Value	0.05	0.01	RESULT
Civil - Mechanical	CV - ME	71 - 58	5.14	0.61	2.016	2.662	Not Significant
Civil - Mechatronic	CV - MT	71 - 18	34.12	3.92	2.056	2.741	0.01 Significance
Mechanical - Mechatronic	ME - MT	58 - 18	17.93	2.03	2.079	2.789	Not Significant

confirms a significant difference occurs between disciplines, and not between subjects. That is, some subjects may have higher marks on average, while some may have lower marks.

## Conclusion

The Binary T-Tests indicated that the average marks of students did not significantly differ between different gender students, and did not significantly differ between the different residency students. Therefore neither gender nor residency was found to be a factor in pre-determining 'at risk' students.

The linear regression tests indicated that both student entry ranking and student age, did significantly effect the final average student performance. Hence both poorer student entry ranking, and students with lower age, are a factor in pre-determining 'at risk' students. However the data also revealed that for the students who obtained an overall mark less than 50%, both the entry ranking and age was not statistically shown to be a factor.

The Hotelling's Test revealed that students who had not pre-selected an engineering discipline (undecided students), obtained significantly different overall results to those students who had chosen a discipline. The conclusion is that undecided discipline students were more likely to be 'at risk'. The Hotelling's Test also indicated that some differences between chosen disciplines could be detected. However there was no obvious pattern in the student mark variation, so this particular result is of little practical value.

Consequently, it was determined that there are some specific student demographics, that can be utilised to pre-identify groups of students who will most likely be in need of additional support, in order to succeed in their first year of university.

## References

- Andres, L., Carpenter, S. (1997). Today's Higher Education Students: Issues of Admission, Retention, Transfer, and Attrition in Relation to Changing Student Demographics. Research Report for British Columbia Council on Admissions and Transfer (BCCAT).
- Franssen, R., Nijhuis, J. (2011). Exploring Student Attrition in Problem-Based Learning: Tutor and Student Perceptions on Student Progress. In: Building Learning Experiences in a Changing World. Ser: Advances in Business Education and Training, 3. Springer, Netherlands. 139-146.
- Jamelske, E. (2009) Measuring the Impact of a University First-Year Experience Program on Student GPA and Retention. *Higher Education*, 57(3). 373-391.
- McKenzie, K., Schweitzer, R. (2001) Who Succeeds at University? Factors Predicting Academic Performance in First Year Australian University Students. *Higher Education Research and Development*, 20(1). 21-33.
- McMillan, J. (2008). University Study in Australia: Persistence Completion and Beyond. LSAY Briefing Reports #18.
- Okpala, C. (2002). Educational Resources, Student Demographics and Achievement Scores. *Journal of Education Finance*, 27(3), 885-907.
- Rickinson, B. (1998) The Relationship Between Undergraduate Student Counselling and Successful Degree Completion. *Studies in Higher Education*, 23(1). 95-102.
- Scott, G., Shah, M., Grebennikov, L., Singh, H. (2008). Improving Student Retention: A University of Western Sydney Case Study. *AAIR Journal*, 14(1). 9-23.
- Shah, M., Nair, C. (2010). Monitoring Student Attrition to Enhance Quality. *Quality Approaches in Higher Education*, 1(2) 24-30.
- Thomas, L. (2002) Student Retention in Higher Education: The Role of Institutional Habitus. *Journal of Educational Policy*, 17(4). 423-442.
- Tinto, V. (2010) From Theory to Action: Exploring the Institutional Conditions for Student Retention. In: *Higher Educational Handbook of Theory and Research*, 25. 51-89.
- Tumen, S., Shulruf, B., Hattie, J. (2008) Student Pathways at the University: Patterns and Predictors of Completion. *Studies in Higher Education*, 33(3). 233-252.
- Weldegiorgis, T., Awel, Y. (2013). Determinants of Student Attrition at College of Business and Economics, Mekelle University: Econometric Investigation. Proc. National Symposium on Establishing, Enhancing and Sustaining Quality Practices in Education, Nekemte Ethiopia, 26-27 April. 110-121.

## **Copyright statement**

Copyright © 2013 Hacker, Seagar and Cassidy: The authors assign to AAEE and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to AAEE to publish this document in full on the World Wide Web (prime sites and mirrors), on Memory Sticks, and in printed form within the AAEE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.