PREDICTING STUDENT SUCCESS

IN OPEN AND DISTANCE LEARNING

2006

Ormond Simpson

The Open University, UK

Abstract

This paper reviews some of the ways in which student success can be predicted in both conventional and distance education. Predicting such success is particularly important for varying new students where the pre-course start information available is sometimes slight and withdrawal often occurs very early in a course. It suggests that in such cases statistical methods involving logistic regression analysis are the most useful rather than questionnaires or tutors' opinions. Identifying students with low probability of success allows support to be targeted on them. However there are ethical dilemmas to do so

motivation such as 'Self-Efficacy theory', 'Achievement Goal theory' 'Interest Development 3 model' and so on.

Such methods are clearly important for the insights they offer into students' intellectu fectu
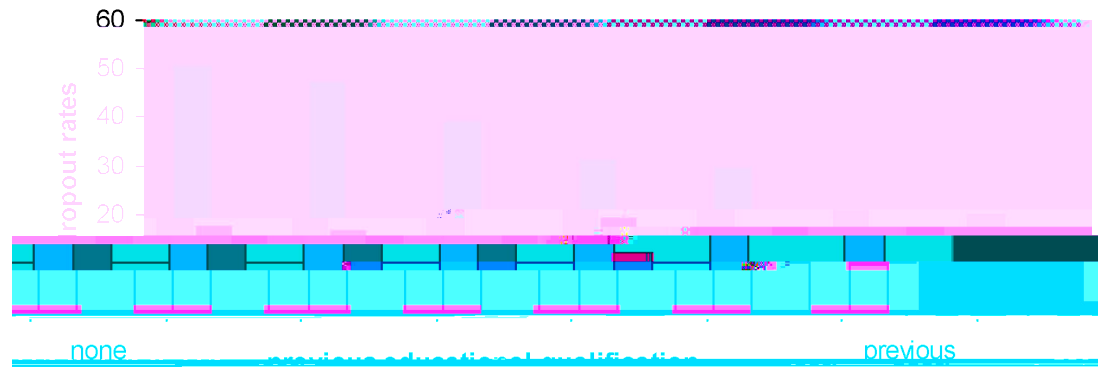
Figure 1 Dropout rates versus previous educational qualifications for first year UKOU students

The results showed a clear increase in student success amongst students targeted for extra support and the project was continued for a number of years. Nevertheless institutional inertia and lack of attention

	None	1-10 hrs	11-15 hrs	16+hrs

Score Sh

results from Wright et al (op cit) suggested that the most vulnerable students would

probably be those least likely to return the questionnaire.

In the event the data available for analysis at new student registration in the UKOU is only

Sex

Age

Previous education qualification

Course choice

Socio-economic status (inferred from occupational status)

Other factors are collected such as special needs and financial award status but often at a later stage, som

Some of these factors made only relatively small difference in student progress. For example the difference in pass rates between men and women is only about 10 percentage points. But when all the factors are taken together the differences predicted can be much larger. In the first analysis undertaken using this method the predicted probability of success varied from 84% (for a well-qualified woman studying arts course with other positive characteristics) to 9% (for an unqualified man studying technology courses and other negative factors). The output data from the analysis was in the form of a spreadsheet that showed a predicted probability of success (expressed as %) for any particular student. An extract from that output is shown as Table 1 (the full spreadsheet contained predictions for nearly 3500 new students taking about 5300 courses between them). The initial letter of the course code denotes the faculty so M- are maths courses, S- are science courses , T- are technology courses and A- are arts courses.

Student	Sex	Course code	Predicted probability of success % (pps)
1	M	S281	9.4
2	M	M358	13.1
3	M	T171	13.6

- which ran through to

Student	Sex	Course code	Predicted probability of success %
5321	F	A103	82.1
5339	F	A103	84.4
5340	F	A103	84.4

Table 1 Extract from the spreadsheet of predicted probability of success.

The distribution for the predicted probability of success (pps) for the 5300 students-courses involved in this prediction is shown in Fig 3.

number of students in each 10%pps band
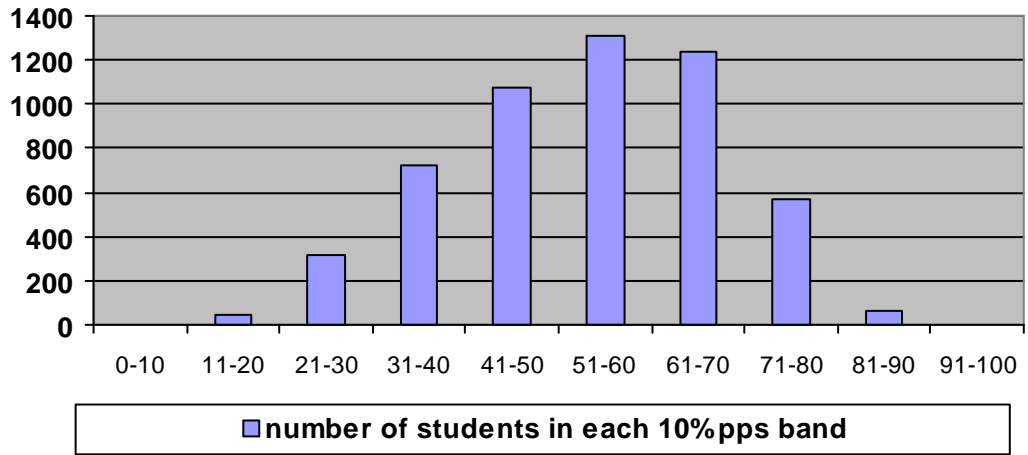
F

65-70	67
70-75	72
75-80	76
80-85	80

Table 2 Comparison of predicted pass rates and actual pass rates in 5% bands (n=5287)

There is good agreement between predicted and actual values although for the lowest and highest bands the number of students involved is very small – there is only one student in the 5-10% pps band out of 3,500 for example.

Of course for any one student the critical factor is not their probability of passing but their actual result which is pass or fail. The overall pass rate is around 44% and if the pps distribution is divided at that point the prediction of pass/fail is correct in 65% of cases – i.e. for any one student with a pps above 56% a prediction of a pass will be correct in 65% of cases.

The courses analysed above are largely conventional distance education courses. But when it comes to online learning the same factors appear to apply with the addition of familiarity with computers. For example Dupin-Bryant (2004) in a study of pre-entry variables in online learning found that prior education and IT skills were the most important determinants of student retention.

Using the Predictive model
The stimulus for using the predictive model in the UKOU was the recognition that in order to increase its retention rates the university needed to undertake proactive contact with its new students rather than waiting for new students to contact it. It was realised that such contact had to be individual and targeted so that it was almost inevitably made through individual phone calls. However the cost of proactively contacting all the 35,000 new students each year was thought to be prohibitive so a way was sought of targeting new students so that interventions would produce the greatest effect.

It was assumed that proactive contact with students with a high predicted probability of success was unlikely to increase that pps by any substantial amount and that better effects would be produced by concentrating contact on students with lower pps's. Thus proactive contact was concentrated on students in the below 56% pps category. This

In a review reported by Simpson (2004) there was an increase in retention of around 4% for the contacted students compared with a control group of identical approach. This is similar to that reported by Kroger (2003) of a 5% increase in student retention using similar methods at Ohio State University with full-time students.

However the assumption that contact with higher pps students would have a lesser retention effect was not tested and it remains to be shown at what levels contact may be most effective. The evidence to date suggests that contact at very low pps levels does not make as much difference as contact in the middle of the range but this is based on very small student numbers. It does however raise various ethical issues.

Ethical issues in using the predictive model.
There are a number of potentially sensitive issues to do with acting on the data supplied by the predictive model.

a   Limited accuracy. Using that data to target resources on students is open to the
         criticism that the data is limited in accuracy. This is particularly true as a substantial
         proportion of new students do not give full personal data. It is occasionally the case
         that an apparently vulnerable new student was phoned only for the adviser to
         discover that – for example – the student has a high level q ee

However the University is shortly to introduce a Customer Relationship 13 Management (CRM) system and it is likely that a pps will form part of a student's record on that system. It is much more likely that students will then become aware of that data and ask to see it.

Being upfront with students. Indeed there are arguments that say the data should be made available to students in some way. Should a university allow a student with apparently only a 20% or less chance of success to embark on a course without some kind of strong warning that that is the case? On the other hand just to contact such a student with that information could be irritating for someone who had just omitted pertinent data which vitiated the prediction or very demoralising for someone who through sheer motivation would have overcome their disadvantages and succeeded.

One possible route through this ethical maze would be to provide the data but mediated through an appropriate person such as the tutor in the Napier model. This is likely to be difficult in a mass education system such as the OU for cost reasons but it may be possible to encourage students to undertake a 'self-assessment' using a questionnaire similar to the Napier but designed to be self-assessed by the student. An example developed by the author is shown below. However this has not yet been developmentally tested.

FTCHV

HOW GOOD ARE YOUR CHANCES OF PASSING?

Everyone who starts with the OU has a chance of succeeding. Of course you'll need commitment, time and energy. And a sense of humour will help!

There are also factors in your background which we know may affect your performance in your first year. This questionnaire is designed to help you
- become aware of the factors which may affect your performance
- to identify factors which might apply to you particularly
- to point to actions which you might be able to take on some of the factors to improve your chances of success.

Start with a score of 60 points. Answer each question in turn and add or subtract a point score as you go along.

Initial Score :   60   points

1. Are you male or female?
   Male : subtract 5 Female: No change   Revised Score: points
2. How old are you?   Revised Score: points
   Under 30 : Subtract 13 Age 30 or above : No change
3. What level is this course?
   Level 1: Add 23 Level 2 :

4.   What Faculty is this course?   Revised Score: points
    A : Add 16   D or L: Add 8   E or K: Add 7   M : Add 6
    S : Subtract 3   T: Add 1   Other: No change
5.   What is the credit rating of this course?   Revised Score: points
    15pts : Subtract 23   30pts : Subtract 9   60pts : No change
6.   How many courses are you taking in total this year?   Revised Score: points
    1 course : Add 5   2 or more courses : No change
7.   What are your current highest educational qualifications?
    Degree or equivalent : Add 17
    A-level or equivalent : Add 12
  O level, GCSE or

The predictive modelling of student success can be sufficiently accurate to be worth using 15 for targeting support onto vulnerable students. There is some limited evidence of the effectiveness of this approach particularly if statistical methods rather than questionnaires are used. However the model also raises ethical issues about the use of the data notably whether and how such predictions can be shared with students. But the model may also be useful for setting benchmarks in the evaluation of courses and student support.

Wright, N. and Tanner, M.S. (2002) Medical student compliance with simple administrative tasks and success in final exams – a retrospective cohort study, *British Medical Journal* 7353, 29 June 2002 pp1554-1555.