Malaysian Journal of Distance Education 9 (1), 91–111 (2007)

Assessing Student's At-Risk Level of Non-Completion in an Open and Distance Learning Course

Rocky Y.K. Fan
School of Science and Technology
The Open University of Hong Kong
Ho Man Tin, Kowloon,
Hong Kong SAR, China
rfan@ouhk.edu.hk

Abstract

Student attrition is a well-documented problem concerning open and distance learning (ODL) institutions. Evidence shows that the non-completion rate on an ODL course can be reduced if the at-risk students are followed up at an early stage. There is a problem in identifying such at-risk students as they may not be obvious at the beginning of their studies. Moreover, it would be difficult to collect at-risk evidence from students during the course presentation for personal assessment. This paper presents a Logistic Regression Model for assessing student's at-risk levels in an ODL course. The model is defined based on the findings in a previous study that ODL experience, academic background and assignment performance are three major variables relating to student attrition. Research results have shown that the model can successfully classify about 80% of students into 'completion' or 'non-completion' after the first assignment score is available. The simple choice of predictors and high classification rate make the model a practical instrument for an early identification of at-risk students.

Abstrak

Keciciran pelajar adalah masalah yang banyak didokumentasikan berkaitan dengan institusi pendidikan terbuka dan jarak jauh. Bukti menunjukkan bahawa kadar ketidaksempurnaan kursus pendidikan terbuka dan jarak jauh boleh dikurangkan jika pelajar yang berkecenderungan untuk tercicir dipantau pada peringkat awal. Terdapat masalah dalam mengenal pasti pelajar yang cenderung untuk tercicir kerana mereka tidak begitu jelas pada peringkat awal pengajian mereka. Tambahan pula ianya agak sukar untuk mengumpul bukti daripada pelajar yang cenderung untuk tercicir ini daripada pembentangan kursus untuk penilaian pelajar. Kertas kerja ini membentangkan Model Regresi Logistik untuk menilai tahap pelajar yang berkecenderungan untuk tercicir dalam kursus pendidikan terbuka dan jarak jauh. Model ini didefinisikan berasaskan kepada dapatan dalam kajian sebelum ini yang menunjukkan bahawa pengalaman pendidikan jarak jauh, latar belakang akademik dan prestasi tugasan adalah tiga pembolehubah yang besar yang berkait dengan keciciran pelajar. Dapatan kajian telah mendapati bahawa model ini boleh berjaya mengklasifikasikan lebih kurang 80% daripada pelajar ke dalam 'penyempurnaan' atau 'tidak penyempurnaan' selepas skor tugasan pertama diperolehi. Pilihan jangkaan mudah dan kadar pengkelasan yang tinggi menjadikan model ini sebagai satu instrumen yang praktikal untuk mengenal pasti pelajar yang berkecenderungan untuk tercicir pada peringkat awal.

Introduction

ODL institutions provide learning opportunities to people who are unable or unwilling to enrol in conventional institutions. Millions of students enroll on one or more ODL courses each year. Many of them, however, do not complete the courses. They either drop out during the course presentation or fail the course assessments. Despite the extensive literature on student attrition (Tinto, 1975; Sweet, 1986; Peters, 1992), the high attrition rate has been and is still a challenging problem concerning ODL institutions. Statistics have shown that the non-completion rate on an ODL course can be as high as 70% (Glatter & Wedell, 1971; Carr & Ledwith, 1988; Fan & Chan, 1999).

Based on various frameworks developed for explaining student attrition, a wide range of support services, from pre-enrolled to post-enrolled and from face-to-face to online, have been designed to assist at-risk students for a better chance of success (Croft, 1991; Metcalfe & Halstead, 1994; Fan, 1999). Nevertheless, even well designed support services could be ineffective for enhancing students' performance (Fan, 1999). For a student support service to be effective, its timing of provision is sometimes an important factor for consideration. Case and Elliott (1997) found that if the at-risk students were encouraged at an early stage, their completion rate could be significantly improved.

In this study, at-risk students are defined as students having a certain high chance of non-completion. To offer early support, the at-risk students need to be identified as early as possible. Some instruments have been developed to identify at-risk students at the beginning of a course presentation (Brindley & Maxim, 1990; Parker, 1995). The instruments assess students mainly based on questionnaire survey and personal interview results collected from the students. Research results have shown that the instruments can predict student completion and non-completion with 70% to 85% accuracy. Based on the prediction results, appropriate actions can be taken to support the at-risk students.

Although the student completion and non-completion is mostly predictable, there are practical problems in collecting evidence for identifying the at-risk students. In a distance learning environment, it is generally difficult to collect personal information from students during a course presentation. Many students, at-risk students in particular, usually do not respond to optional questionnaire surveys and personal interviews (King, 1995; Fan & Chan, 1999). The situation of the non-respondents will then be undetermined. In addition, many at-risk students are not obvious at the beginning of their studies. The information collected at the beginning may not be useful for predicting their problems encountered at a later stage.

This paper proposes an at-risk model for quantifying the at-risk evidence presented in a student record. The model is defined based on the findings in Fan and Chan (1999) that ODL experience and academic background are two important background factors affecting the persistence of a student. In addition, a strong relationship exists between performance on continuous assessment and student attrition. Applied to this study, these results suggest that at-risk evidence for individual students can be defined from their experience at the institution, educational level on entry and assignment scores. All these sources of at-risk evidence are available in the student records either before or during the course presentation.

Student attrition is a multi-causal problem and each student has a different set of factors affecting his or her decision of not completing a course. The at-risk model developed in this study does not attempt to explain why some students are more risky than the others. It aims to provide a basis for an early identification of at-risk students in an ODL course based on available data. The definition and development of the model will be presented in the next section. The model will then be applied to an ODL course offered by the Open University of Hong Kong (OUHK) to facilitate discussion. Finally, further developments and applications of the model will be discussed.

The At-Risk Model

For the purpose of this paper, each student in an ODL course is classified into either 'completion' or 'non-completion' according to his or her final status. A student is classified into completion and referred to as a 'survivor' if he or she passes the course. A 'non-survivor' is a student who either drops out during the presentation or fails the assessment. 'At-risk level' is defined as the chance of a student not completing the course. A student is identified as 'at-risk' if his or her at-risk level exceeds a 'cut-off level'. The cut-off level is usually chosen to minimise cost of misidentification. An identification is considered successful if the result is at-risk for a non-survivor and non-risky for a survivor.

This study aims to define a model that summarises the at-risk evidence observed in a student record and presents the result as an at-risk level. In mathematical terms, the at-risk level of a student is a zero-to-one number defined as a function of some characteristics of the student. The choice of characteristics will be considered later. Using the characteristics as predictors, logistic regression will be applied to construct the model. Logistic regression is a useful statistical technique for situations where individuals are to be classified into one of two distinct populations (Afifi & Clark, 1990). A logistic regression model having predictors assumes the following mathematical form:

The chance of being in the first population

$$= \frac{\exp\{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k\}} \qquad (G_k)$$

In the general model G_k, $_i$s are parameters to be estimated in model construction and x_is are predictors to be supplied by practitioners for estimating the parameters and applying the model. Note that for any given

$_i$s and x_is, the value of the right-hand expression in G_k always lies between 0 and 1. In practice, the parameters are estimated from available records each having known population class and predictors. The model can then be applied to any person having known predictors for calculating his or her chance of being in the first population. The chance can be compared with a cut-off level for classifying the person into one of the two distinct populations.

In this study, the two distinct populations for the logistic regression are defined to be 'non-completion' and 'completion' with 'non-completion' being the first population. It follows that the chance of being in the first population is the at-risk level. The predictors are some characteristics of the students to be chosen. An objective of this study is to define the predictors such that they can be derived from available data and do not involve additional information to be collected from questionnaire surveys or personal interviews. In other words, the usable data for defining the predictors will only include personal particulars provided by students at registration and official data generated in their studies.

At the OUHK, personal particulars including some optional information of each student are collected at registration and stored in his or her student record. The record will also contain the study history and up-to-date results of the student on individual courses studying at the university. It is not expected that the information available in the record can fully explain why a student does or does not complete a course. Some of the background characteristics and study results, however, can be very useful indicators of the problems (or potential problems) and persistence that affect one's final status (Kember, 1989; Tinto, 1982). The model predictors for this study will be defined using such indicators.

To determine the model predictors, some results in Fan and Chan (1999) would be useful. In that study, the characteristics of those students who did not complete each of two foundation courses were investigated. The characteristics available for investigation included date of birth, sex, marital status, occupation, educational level on entry, current study load and number of years studying at the OUHK. It was found that experience at the university and academic background were two significant and common background factors affecting one's persistence of studying. In the same study, it was also found that the performance of a student on assignments was significantly related to his or her final status.

The OUHK adopts an open entry policy and provides higher education for adults principally through distance learning. ODL enhances study opportunity and flexibility; however, it also presents challenges to students who are not well prepared. Two major causes of student attrition are that students do not have adequate academic preparation and they cannot adapt to the distance learning environment (Fan & Chan, 1999). Their relationship to experience at the university and educational level on entry is obvious. It is therefore sensible to use these two background characteristics for at-risk assessment especially at the beginning of the course presentation.

Assessment from background characteristics only provides an average assessment of students having a similar background. It does not take into account personal factors that may have affected one's chance of success. Moreover, the results in Sweet (1986) suggest that background characteristic explain only a small amount of variance in persistence. For more accurate run-time assessment during a course presentation, up-to-date personal factors should be considered. The strong relationship between assignment scores and final status suggests that assignment scores have reflected most of the impact of personal factors on final status and therefore a desirable at-risk indicator for run-time assessment.

Experience at the university, educational level on entry and assignment scores are three valuable sources of information for at-risk assessment. In particular, the assignment scores provide up-to-date information for run-time assessment during a course presentation. These variables will form a basis for defining predictors in the at-risk model. The example to be presented in the next section will demonstrate that, even using only such simple variables, the model can provide a very accurate prediction on the completion and non-completion of individual students. Before going into details of the example, the at-risk model and its construction are now presented based on the above discussion.

For an ODL course having n assignments, there will be n + 2 pieces of information available for model construction, that is, OUHK experience, academic background and n assignment scores. Supplying the information in previous student records to any statistical package, an at-risk model in the form of G_k can be easily fitted. Precisely, there is not one but a sequence of n at-risk models constructed for each course. In addition to the background factors, the first model uses the first assignment score, the second model uses the first two, and so on. To assess a student at any particular time during a course presentation, the model involving the background factors and all available assignment scores can be used.

The model G_k defines a general form of the at-risk model having k predictors. In general, the kth at-risk model has k + 2 predictors including two background factors and k assignment scores. There are various ways to define predictors from the background factors. The results in Fan and Chan (1999) show that it would be sufficient to classify students according to whether they are new or old to the university and whether they have met the academic requirements for registration at a local conventional university. To apply the model, the parameters in the model must be estimated. The variables involved in the estimation are defined as follows:

- *New* An indicator coded 1 for a new OUHK student and 0 for an old student.
- *Low* An indicator coded 1 if the student has not matriculated and 0 otherwise.
- As_i The ith assignment score, i = 1, …, k.
- *Fail* An indicator coded 1 for a non-survivor and 0 for a survivor.
- ARL_k The kth at-risk level defined by the kth at-risk model.

New, *Low* and $AS_i s$ are predictors in the at-risk model. Fail identifies the population class and is defined from the final status. The model parameters are estimated from previous student records. Each record defines a data point including the predictors and *fail*. For ease of comparison, the assignment scores AS_is are normalised to have values between 0 and 1. The at-risk level ARL_k is not involved in the estimation. It is calculated using the fitted model. As mentioned before, a total of n at-risk models will be constructed for a course having n assignments. The kth model using *New*, *Low*, AS_1, …, AS_k as predictors has the following form and will be referred to as the basic at-risk model B_k.

$$ARL_k = \frac{\exp\{\beta_0 + \beta_1 New + \beta_2 Low + \beta_3 AS_1 + \cdots + \beta_{k+2} AS_k\}}{1 + \exp\{\beta_0 + \beta_1 New + \beta_2 Low + \beta_3 AS_1 + \cdots + \beta_{k+2} AS_k\}} \qquad (B_k)$$

To summarise, this section has developed the basic at-risk model B_k using available data. The model can be applied for analysing the data as well as assessing current students. In general, n basic models can be constructed for a course having n assignments. The kth model has k + 2 predictors *New*, *Low*, AS_1, …, AS_k and k + 3 parameters β_0, β_1, …, β_{k+2}. The parameters are estimated using the k + 3 dimensional data points (*New*, *Low*, AS_1, …, AS_k, *Fail*) derived from previous student records. The estimation process is repeated for k = 1, … n. In practice, the estimation of parameters requires the assistance of a statistical package. For this study, the package SPSS will be used.

An Example

This section presents an example to facilitate discussion on the implementation, application and evaluation of the basic at-risk model. In practice, applying the model for at-risk assessment involves two steps and at least two course presentations. The first step is to estimate the model parameters using previous records. The second step is to apply the fitted model for assessing current students. In this example, a single presentation will be used and its students will be divided into two equal halves, one for parameter estimation and the other for at-risk assessment. This arrangement aims to provide the same environment for the two steps and therefore a better basis for model evaluation.

In this example, the basic at-risk model will be implemented for the course MT210. MT210 was a middle level ODL course in computing offered every year at the OUHK. The course was designed for students pursuing degrees in computing and related disciplines. Students enrolling on this course were required to submit four tutor-marked assignments during the course presentation for continuous assessment and attend a final examination at the end of the presentation. Each assignment is worth 25% of the overall continuous assessment score. A student is required to obtain at least 40% in each of the continuous assessment and the final examination for passing the course.

The following discussion will be based on the student records of the April 2000 presentation of MT210 ending in February 2001. There were 828 students enrolled on the course. Using SPSS, the students are randomly divided into two main groups of size 414. Group 1 plays the role as 'previous students' for parameter estimation and Group 2 as 'current students' for at-risk assessment. A model evaluation will be done based on the estimation and assessment results. As *New* and *Low* are two major factors affecting student performance, the evaluation will also be done for each *New-Low* subgroup. The number of students in each main group and *New-Low* subgroup is given in Table 1.

Table 1 Numbers of MT210 students

New	*Low*	Group 1	Group 2
Yes	Yes	32	32
	No	88	80
No	Yes	118	125
	No	176	177
Total		414	414

MT210 has four tutor-marked assignments and therefore four basic at-risk models will be constructed. The first step of the construction is to derive the values of *New*, *Low*, AS_1, …, AS_4 and *Fail* for each student in Group 1 from his or her OUHK record. Using the SPSS option for logistic regression with *New*, *Low* and AS_1 as predictors and *Fail* as the independent variable, the four parameters β_0 to β_3 in the first basic at-risk model B_1 are estimated. In a similar manner, the parameters in the basic models B_2, B_3 and B_4 are estimated. Table 2 summarises the estimates of the model parameters, where β_0 is the intercept, β_1 to β_6 are the parameters for *New*, *Low*, AS_1, …, AS_4 respectively.

Table 2 Estimates of parameters (Model B_k, Group 1)

	β_0	β_1	β_2	β_3	β_4	β_5	β_6
B_1	3.71**	0.84**	0.48*	–6.28**	N/A	N/A	N/A
B_2	3.51**	0.86**	0.58*	–1.19	–6.40**	N/A	N/A
B_3	3.27**	1.11**	0.79**	–0.75	–3.33**	–5.66**	N/A
B_4	3.42**	1.04*	0.74	–0.49	–2.27*	–3.02**	–7.77**

* significant at 0.10 level, ** significant at 0.05 level

The SPSS results show that all four basic models are significant. (In this study, a result is considered significant if the result is statistically significant at the 0.05 level.) The signs of the parameters agree with the common sense that new students, students having lower educational background and students performing less satisfactory on assignments are generally more risky than their counterparts. The results also show that the two background factors tend to be more significant at the beginning of a course presentation. In addition, if two assignment scores are involved in the same model, the second score is always more significant and has a larger contribution to the at-risk level than the first one.

With the estimates in Table 2, the four basic at-risk models defined for MT210 are ready for applications. An application is to define and analyse the contribution of a predictor. For instance, the contribution of *Low* can be defined as the difference between the at-risk level at *Low* = 1 and the level at *Low* = 0. To evaluate the model fitting, one may compare the predicted non-completion rates (average at-risk levels) with the actual non-completion rates (*NCR*) for Group 1. Table 3 shows that the predicted rates match closely with the actual rates for the main group and all *New-Low* subgroups, even for the first basic model involving only two background factors and one assignment score.

Table 3 Predicted and actual non-completion rates (Model B_k, Group 1)

New	*Low*	$\overline{ARL_1}$	$\overline{ARL_2}$	$\overline{ARL_3}$	$\overline{ARL_4}$	*NCR*
Yes	Yes	0.64	0.66	0.65	0.69	0.66
	No	0.45	0.44	0.45	0.43	0.44
No	Yes	0.46	0.46	0.46	0.45	0.46
	No	0.35	0.35	0.35	0.36	0.35
Overall		0.43	0.43	0.43	0.43	0.43

The close matching has two important implications. Firstly, the predictors are not merely statistically significant, they are practically useful in defining student's at-risk levels. Secondly, the relationship between the predictors and the average at-risk level (within the main group or each *New-Low* subgroup) is well defined by the basic at-risk model. The model is useful by itself as a summary of available data in many aspects, such as identifying unusual contributions of individual predictors for further investigations of attrition-related problems. A more direct and practical application of the model, of course, is to use it for at-risk assessment of current students.

At-risk levels of current students can be calculated by substituting their predictor values into the fitted basic models. In reality, the final status of these students are unknown when they are being assessed. In this example, Group 2 serves as a group of current students but their final status have been available. To evaluate the fitted models as at-risk assessors, the at-risk levels of each student in Group 2 are calculated and compared with his or her final status. The results are evaluated in terms of predicted non-completion rates and identification rates of completion and non-completion. The predicted and actual non-completion rates for Group 2 are given in Table 4.

Table 4 Predicted and actual non-completion rates (Model B_k, Group 2)

New	*Low*	$\overline{ARL_1}$	$\overline{ARL_2}$	$\overline{ARL_3}$	$\overline{ARL_4}$	*NCR*
Yes	Yes	0.63	0.58	0.66	0.61	0.56
	No	0.48	0.48	0.47	0.43	0.46
No	Yes	0.45	0.52	0.51	0.49	0.50
	No	0.31	0.36	0.35	0.35	0.37
Overall		0.41	0.45	0.45	0.43	0.44

The above results show that the basic models can provide accurate predictions of the non-completion rates. The differences between the predicted and actual non-completion rates are generally insignificant except for the first predicted rate of the last subgroup (old students having high educational background). For the last subgroup, the first predicted rate is 0.31 with standard deviation 0.02 and the actual rate is 0.37. It follows that the non-completion rate of this subgroup may not be accurately predicted using the first basic model. The result suggests that the student attrition of this subgroup is generally less related to its background and initial performance.

To address the student attrition problem in MT210, it would be helpful to identify the at-risk students in the course and then focus on them for more efficient support. The at-risk levels defined by the fitted basic models can be used to identify the at-risk students. The idea is to classify a student as 'at-risk' or 'non-risky' by comparing his or her at-risk level with a pre-defined cut-off level. This example takes 0.5 as the cut-off level, that is, a student having an at-risk level higher than 0.5 will be referred to as an at-risk student. The identification rates of classifying the non-survivors in MT210 as at-risk and the survivors as non-risky are summarised in Table 5.

Table 5 Identification rates (Model B_k, Group 2)

New	*Low*	B_1	B_2	B_3	B_4
Yes	Yes	0.81	0.84	0.94	0.94
	No	0.78	0.85	0.94	0.91
No	Yes	0.74	0.87	0.88	0.90
	No	0.79	0.89	0.90	0.95
Overall		0.78	0.87	0.91	0.93

The high identification rates in Table 5 indicate that the fitted models have provided very accurate at-risk assessments for individual students. The performances of the models are similar for Group 2 and the *New-Low* subgroups, that is, the identification rates start at about 80% and go up to 90+%. The rates include successful identifications of both non-survivors and survivors. For practical reasons such as defining an optimal cut-off level for identifying at-risk students, it is usually important to understand also the performances of the fitted models for non-survivors and for survivors. The identification rates of the fitted models for the non-survivors in MT210 are given in Table 6.

Table 6 Identification rates for non-survivors (Model B_k, Group 2)

New	*Low*	B_1	B_2	B_3	B_4
Yes	Yes	0.83	0.83	1.00	1.00
	No	0.59	0.73	0.92	0.86
No	Yes	0.58	0.87	0.85	0.94
	No	0.47	0.79	0.83	0.88
Overall		0.57	0.81	0.87	0.91

The results in Table 6 suggest that the fitted basic models generally perform better in identifying survivors than non-survivors; nevertheless, they manage to identify 57% of non-survivors at the beginning and the identification rate improves quickly to exceeding 80%. The performances of the models are similar for all *New-Low* subgroups except for the first model for the first subgroup. The outstanding result of the first model for new students having low educational background suggests that most of the student attrition in this subgroup, if not caused by the inadequacy in any of OUHK experience, academic background and initial performance, is at least closely related to and therefore predicted by these factors.

Discussion

The basic at-risk model B_k aims to provide a simple and accurate at-risk assessment based on available data for addressing the student attrition problem. The MT210 example demonstrates that the aim has been reasonably achieved. It should note that the basic model is not a complete solution to the attrition problem. The at-risk assessment is defined based on some background factors and academic results; however, the model does not explain why a student has a certain chance of non-completion. For instance, low assignment scores lead to high at-risk levels but the academic results do not imply that academic problems exist. It may only assume that problems having negative impact on academic performance exist.

The basic model has many attractive features. It assesses students based on available data only and the assessment can be updated during the course presentation when the next assignment score is available. The MT210 example shows that the model has defined a reliable relationship between the at-risk level and the predictors *New*, *Low*, AS_1, …, AS_k, similar results have been obtained for courses having different continuous assessment components. The relationship can be used for course evaluation as well as at-risk assessment. Using a statistical package, the model can be constructed for almost any ODL courses. Moreover, it can be easily amended with additional predictors for further enhancement.

In the basic model, *New*, AS_1, …, AS_k are always obtainable and *Low* needs to be collected from students. Personal particulars such as academic background, though useful for at-risk assessment, are usually collected on a voluntary basis and their availability is not guaranteed. In the case where personal particulars are not available, the assessment needs to rely on *New*, AS_1, …, AS_k. To handle this case, a simplified at-risk model S_k can be defined by removing *Low* from the basic model B_k. For demonstration, the simplified model S_1 is constructed for the MT210 example. The predicted non-completion rates in Table 7 suggest that the performance of the simplified model can be close to that of the basic model.

$$ARL_1 = \frac{\exp\{3.96 + 0.76New - 6.34AS_1\}}{1 + \exp\{3.96 + 0.76New - 6.34AS_1\}} \qquad (S_1)$$

Table 7 Predicted and actual non-completion rates (Model S_1)

	Group 1		Group 2	
New	$\overline{ARL_1}$	*NCR*	$\overline{ARL_1}$	*NCR*
Yes	0.50	0.50	0.52	0.49
No	0.39	0.39	0.37	0.42
Overall	0.43	0.43	0.41	0.44

The basic model uses assignment scores for at-risk assessment and cannot be applied until students submit their assignments. For a preliminary assessment, a pre-enrolled model B_0 can be defined using only the background factors. The at-risk level ARL_0 can be interpreted as the predicted non-completion rate of students having a similar *New* and *Low* background. For demonstration, B_0 is constructed for the MT210 example and both predictors are found statistically significant. The identification rates (ID%) in Table 8 show that B_0 may not be very helpful for personal assessment. The predicted non-completion rates, however, can provide a reasonable overall picture for course's planning and pre-enrolled counselling.

$$ARL_0 = \frac{\exp\{-0.66 + 0.51New + 0.55Low\}}{1 + \exp\{-0.66 + 0.51New + 0.55Low\}} \qquad (B_0)$$

Table 8 Predicted non-completion rates and identification rates (Model B_0, Group 2)

New	*Low*	ARL_0	ID%
Yes	Yes	0.60	0.56
	No	0.46	0.54
No	Yes	0.47	0.50
	No	0.34	0.63
Overall		0.42	0.57

It has been demonstrated that the basic model is a simple and reliable instrument for at-risk assessment. In addition, the choice of predictors allows it to be applied to almost any ODL courses. Different institutions generally have different information in their student records and different courses may take different predictors for better-performed models. In situations where better predictors are identified and available, the basic model can be easily amended to include the new predictors for enhancing its performance. New predictors need not involve new information. They can be rearranged from existing predictors for simplicity or other purposes. Here below is an alternative of the basic model.

The basic model B_k has k + 2 predictors, where k is the number of assignment scores involved in the model. In general, the number of scores varies from model to model and can be rather large. For practical reasons, at-risk models having a small and fixed number of predictors are sometimes preferred. If an understanding of the contributions of individual scores is not necessary, a refinement can be to replace all scores in the basic model with a single predictor representing the overall contribution of the scores. There are many choices for such a single predictor. For instance, the following refined at-risk model R_k is defined by replacing all assignment scores in the basic model with a weighted average of the scores.

$$ARL_k = \frac{\exp\{\beta_0 + \beta_1 New + \beta_2 Low + \beta_3 WAS_k\}}{1 + \exp\{\beta_0 + \beta_1 New + \beta_2 Low + \beta_3 WAS_k\}}, \qquad (R_k)$$

where WAS_k is the weighted average , $\frac{\sum_{i=1}^{k} W_i AS_i}{\sum_{i=1}^{k} W_i}$

W_i is the weight of the *i*th assignment in the course.

Note that the first refined model and the first basic model are the same by definition. For demonstration, four refined models are constructed for the MT210 example. Using SPSS, the four parameters in each refined model are estimated using the Group 1 records (Table 9). The results show that all refined models and all parameters except the one for the predictor *Low* are significant. The at-risk levels of the Group 2 students are recalculated using the refined models. The predicted non-completion rates and identification rates are given in Tables 10 and 11 respectively. The results show that the performance of the refined models is comparable to that of the basic models.

Table 9 Estimates of parameters (Model R_k, Group 1)

	β_0	β_1	β_2	β_3
R_1	3.71**	0.84**	0.48*	–6.28**
R_2	4.97**	0.90**	0.56*	–9.09**
R_3	5.12**	1.04**	0.68*	–11.03**
R_4	5.48**	1.09**	0.66	–13.37**

*significant at 0.10 level, **significant at 0.05 level

Table 10 Predicted non-completion rates (Model R_k, Group 2)

New	*Low*	β_3	$\overline{ARL_2}$	$\overline{ARL_3}$	$\overline{ARL_4}$
Yes	Yes	0.63	0.59	0.64	0.64
	No	0.48	0.48	0.47	0.45
No	Yes	0.45	0.51	0.51	0.50
	No	0.31	0.35	0.35	0.35
Overall		0.41	0.44	0.44	0.44

Table 11 Identification rates (Model R_k, Group 2)

New	*Low*	R_1	R_2	R_3	R_4
Yes	Yes	0.81	0.84	0.94	0.97
	No	0.78	0.83	0.91	0.89
No	Yes	0.74	0.88	0.89	0.94
	No	0.79	0.90	0.91	0.94
Overall		0.78	0.87	0.91	0.93

Both B_k and R_k are full at-risk models constructed for students having different *New* and *Low* background. To study the behaviour of a particular *New*-*Low* subgroup, a sub-model can be obtained by fixing *New* and *Low* in a full model. For instance, setting both *New* and *Low* in the refined model R_k to 1 gives a sub-model for new students having low educational background for analysis. In general, there are k predictors (assignment scores) in a basic sub-model and one predictor (weighted average of assignment scores) in a refined sub-model. The analysis of at-risk models is mostly algebraic. It can be graphical for refined sub-models as WAS_k will be the only predictor. Here below demonstrates the graphical approach.

Consider the *New-Low* sub-models derived from R_1. Figure 1 shows that the at-risk level curves generally decrease with respect to WAS_1. Reading between WAS_1 and ARL_1 is simple. Drawing a vertical line at WAS_1 = 0.4, it can be seen that students passing the first assignment marginally possess very high at-risk levels, from 0.77 for old students having high educational background to 0.93 for new students having low educational background. Drawing a horizontal line at the 0.5 cut-off level, it is found that the corresponding cut-off score ranges between 59 for old students having high educational background and 80 for new students having low educational background.

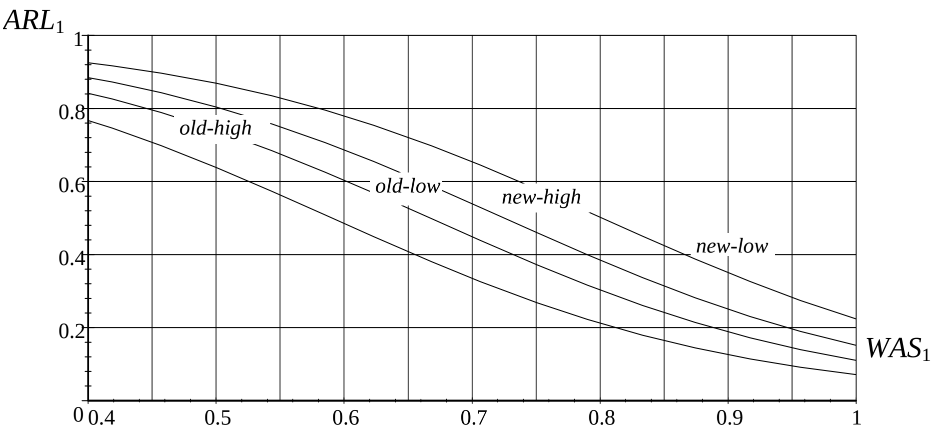

Figure 1 At-risk levels (Sub-models from R_1)

Finally, although the at-risk models defined in this study provide objective assessments, they can give unexpected results. The above cut-off scores are just an example. No one would consider students scoring 80 in the first assignment as at-risk in general, but the results indicate that some of these students should be followed up. This is the power of the at-risk models. The models are constructed based on available records, and experience has shown that 50% of new students having low educational background and scoring 80 in the first assignment did not complete the course. The models define the at-risk levels of a student from experience. They do not wait until problems actually exist.

Conclusion

This paper has defined an objective instrument for measuring student's at-risk levels during the presentation of an ODL course. The instrument is sequence of at-risk models developed based on available data. Enhancements of the instrument with better predictors are simple. In principle, the instrument can be applied to almost any courses using a statistical package. The instrument can serve as a summary of available data or an at-risk assessor. The latter provides a practical basis for identifying at-risk students. The instrument by itself is not a complete solution to the student attrition problem and does not provide prescriptive advice. Follow-up actions are required to identify and solve the real problems of individual students.

Evidence shows that many at-risk students do not pay attention to their situations and are not aware of the potential problems facing them until it is too late. In a pilot study, Fan (2004) assessed an experimental group of students in a foundation course using the refined models. Based on the assessment results and information available in the student records, assessment reports were generated after each assignment and sent to the students by email. As a result, the non-completion rate of the experimental group dropped significantly. Further studies are being done to verify the result, but the pilot results have demonstrated how the at-risk models may be applied to develop an effective solution to the student attrition problem.

References

Afifi, A.A. & Clark, V. (1990). Logistic Regression, Computer-aided Multivariate Analysis, New York: Van Nostrand Reinhold, pp. 317–343.

Brindley, J. & Maxim, J. (1990). Student support services: The case for a proactive approach. *Journal of Distance Education*, *5*(1), pp. 66–70.

Carr, R. & Ledwith, F. (1988). Helping disadvantaged students. *Teaching at a Distance*, *18*, pp. 77–85.

Case, P.S. & Elliott B. (1997). Attrition and retention in distance learning programs: Problems, strategies solutions. *Open Praxis*, *1*, pp. 30–33.

Croft, M. (1991). Student support services: An overview, Report on Round Table on Student Support Services. Canada: Commonwealth of Learning, pp. 3–30.

Fan, R. (1999). Use of course-related support services and the relationship to academic performance. In Carr, R. et al. (eds). *The Asian Distance Learner*. Hong Kong: Open University of Hong Kong Press, pp. 130–149.

Fan, R. (2004). A run-time assessment of student's at-risk levels in an ODL course. Proceedings of the 21st ICDE World Conference, Hong Kong, CD-ROM.

Fan, R. & Chan, M. (1999). A study of student attrition in foundation courses. Open Education Research, *4*, pp. 21–66.

Glatter, R. & Wedell, E.G. (1971). Study by correspondence. London: Longman.

Kember, D. (1989). An illustration, with case studies, of a linear-process model of drop-out from distance education. *Distance Education*, *10*(2), pp. 196–211.

King, T.J. (1995). The identification of high dropout risk distance education students by the analysis of student records data. In Sewart, D. (ed.). One World Many Voices: Quality in Open and Distance Learning. Milton Keynes: The Open University, *2*, pp. 98–101.

Metcalfe, J.A. & Halstead, A. (1994). Into science: Open learning, drop-out and accreditation. *Adults Learning*, *5*(10), pp. 261–263.

Parker, A. (1995). Predicting dropout from distance education, in Sewart, D. (ed.). One World Many Voices: Quality in Open and Distance Learning. Milton Keynes: The Open University, *2*, pp. 176–179.

Peters, O. (1992). Some observations on dropping out in distance education. *Distance Education*, *13*(2), pp. 234–269.

Sweet, R. (1986). Student dropout in distance education: An application of Tinto's model. *Distance Education*, *7*(2), pp. 201–213.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), pp. 89–125.

Tinto, V. (1982). Limits of theory and practice in student attrition. *Journal of Higher Education*, *53*(6), pp. 687–700.