IIAI Open Conference Publication Series
IIAI Letters on Institutional Research
Vol.004, LIR287
DOI: https://doi.org/10.52731/lir.v004.287

Identifying the High Risk Duration to the Semester to Drop Out of College Using Dropout Probability

Naruhiko Shiratori *

Abstract

This study investigates dropout probability, which quantifies the risk of a student discontinuing their studies, to determine the length of the high-risk phase prior to dropout. The departure of a student from college poses significant adverse impacts on both the individual and the institution. Universities have implemented various dropout prevention measures; however, their effectiveness hinges on timely execution. Through an analysis of dropout probabilities by semester for 173 students who eventually dropped out, it was found that these students were at an increased risk of dropping out an average of 2.97 semesters, or approximately 18 months, before actually leaving the university.

Keywords: Dropout Prediction, Dropout Probability, Dropout Risk

1 Introduction

In fiscal year 2022, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) reported a nationwide university and junior college dropout rate of 1.94% (52,459 students) [1]. Contrastingly, a 2019 Yomiuri Shimbun survey indicated that dropout rates by individual universities and departments frequently exceed 10% [2]. The survey also highlighted that private universities generally report higher dropout rates, averaging around 8%, compared to national and public universities. Furthermore, a 2023 MEXT survey focusing on six-year pharmacy school programs revealed dropout rates exceeding 30% at many institutions [3]. Despite an overall low national average dropout rate of approximately 2%, there is significant variation across different universities and departments, with many reporting rates far from low.

Given that dropout rates significantly impact both students and faculty, universities are implementing various measures to prevent dropouts. As Seidman has indicated, effective dropout prevention relies on the early identification of high-risk students coupled with prompt, continuous, and intensive interventions [4]. While early detection is crucial, it raises the question: how far in advance can we predict a student's likelihood of dropping out? Understanding the timeline before a student drops out is essential for efficiently allocating resources to prevention strategies. To address this, the present study introduces a method that uses semester-by-semester dropout probabilities to pinpoint when the risk of dropping out begins to escalate.

* Tokyo City University, Tokyo, Japan

2 Related Studies

2.1 Dropout Prediction Research

Numerous studies have focused on predicting college dropout rates, primarily in Europe and the U.S. Bigham and colleagues demonstrated that logistic regression models can identify how retention rates from the first to the second year vary based on factors such as parental educational background and ethnicity [5]. Guarin and colleagues utilized data on enrollment and subsequent academic performance, analyzing variables like entrance examination format, gender, and high school type to classify students at risk of dropping out due to poor performance [6]. In Japan, Kondo and colleagues have successfully predicted dropouts at the beginning of the third year using variables including gender, department, and entrance examination category [7]. These studies highlight the diverse methodologies employed to predict university dropout rates, particularly in Europe and the United States.

2.2 Research on understanding dropout risk

Research into dropout prediction has led to significant studies, such as the one by Shiratori and colleagues, who employed a logistic regression model to assess student dropout risk on an individual and semester-by-semester basis [8]. This model helped identify the state transitions preceding a dropout and outlined typical dropout patterns. Furthermore, using the same dropout probability data, Shiratori clarified the risk levels for students who did not drop out, categorizing five distinct patterns from enrollment to graduation [9].

As highlighted, studies on dropout prediction and risk have developed methods to forecast student dropout and to decode student status via dropout prediction models. Chapter 1 discusses the necessity of early intervention to prevent student dropouts. However, existing research does not specify how long students remain at high risk before actually dropping out. Calculating this duration is crucial for implementing timely and effective dropout prevention strategies.

3 Data and Variables

3.1 Data

This study employs a method based on the dropout probability analysis conducted by Shiratori et al. The dataset comprises information from students enrolled at A University of Social Sciences in Tokyo between 2012 and 2014. It excludes transfer students, re-enrollment students, early graduates, graduate students, and those lacking pre-enrollment or post-enrollment data. The analysis includes data from 849 students, out of which 173 dropped out.

3.2 Variables

The calculation of dropout probability utilized the following objective and explanatory variables:

3.2.1 Objective variable: y

The objective variable, , indicates whether a respondent dropped out (y=1) or not (y=0).

3.2.2 Explanatory Variables 1: Variables available by time of admission

Variables available at the time of enrollment included:

- Gender (male = 1, female = 0)
- International students (international = 1, non-international = 0)
- Number of days absent from high school
- Type of high school (full-time = 1, other = 0)
- Correspondence high school (correspondence = 1, other = 0)

3.2.3 Explanatory Variables 1: Variables obtainable after enrollment

Variables obtainable after enrollment included:

- Total number of credits earned up to the semester in question
- Number of credits earned during the semester in question
- GPA for the semester in question.

4 Methods

4.1 Calculation of Dropout Probability

The dropout probability for each student m in each semester s is calculated using the following logistic regression formula. In this model, $\alpha_{s,0}$ represents the constant term and $\alpha_{s,i}$ denotes the coefficient for the variable x_i which is available at the time of admission. The coefficient $\beta_{s,k,j}$ is for the variable available after admission. These coefficients are estimated using the maximum likelihood method to determine the dropout probability $p_s[m]$ for student m during semester s. The logistic regression model ensures that the dropout probability values are constrained between 0 and 1.

$$\ln\left(\frac{p_s[m]}{1-p_s[m]}\right) = \alpha_{s,0} + \sum_{i=1}^{6} \alpha_{s,i} x_i[m] + \sum_{k=1}^{s} \sum_{j=1}^{3} \beta_{s,k,j} x_{k,j}[m]$$

4.2 Identifying High Risk Using Dropout Probability

The dropout probability calculated for each student per semester is used to assess the level of risk prior to dropping out. In this study, a dropout probability of 0.5 or higher is considered indicative of high risk. This criterion allows us to determine whether a student is at high risk for dropping out in any given semester, from the time of enrollment through to the semester they potentially drop out.

4.3 Calculation of the Duration from High-Risk Identification to Dropout

After assessing whether a student is at high risk, as previously described, we calculate for each student m the duration period from the semester in which he/she was first identified as high risk (denoted as $s_{hr}[m]$) to the semester he/she dropped out of college (denoted as $s[m]$). This duration is determined by subtracting $s_{hr}[m]$, the semester of initial high-risk identification, from $s[m]$, the semester of dropout.

5 Experimental Results and Discussion

5.1 Duration from High Risk to Dropout Time

We calculated the duration from high risk to dropout time as described in Chapter 4. The average duration is 2.97 semesters, with a standard deviation of 1.94 semesters; since one semester approximately equals six months, this implies that on average, students were at high risk about 18 months prior and have remained in college for an additional 1.5 years after becoming high risk. Figure 1 below presents a histogram of the duration to dropout. This histogram reveals that the period of high risk is most pronounced during the first and second semesters, decreases with the passage of semesters, and peaks again in the sixth semester. It indicates that it typically takes four semesters to reach the dropout point after many students become high risk.

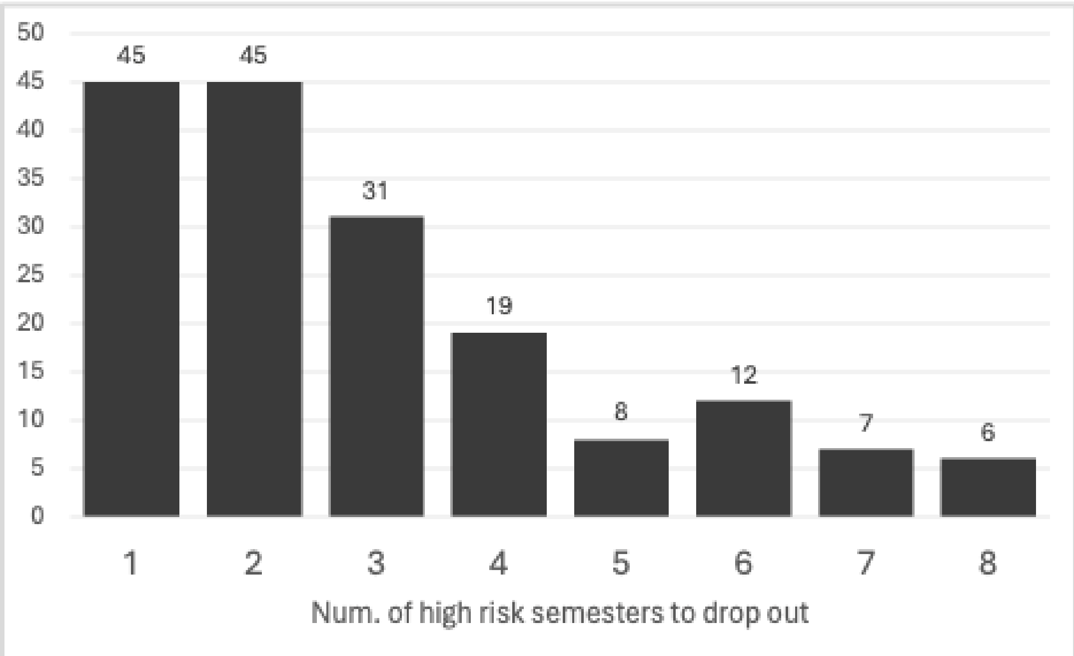

Figure 1: The histogram of the time to dropout period.

5.2 Comparison by semester dropped out

The next step involves comparing the duration from when students were identified as high risk to when they dropped out, segmented by the semester of dropout. The first row of Table 1 below indicates the semester each student dropped out; the second row lists the number of students; and the third row displays the average duration in semesters from becoming high risk to dropout. Semesters are coded with '1' representing the spring semester of the first year, '8' as the fall semester of the fourth year, and '9' as the fall semester of the fourth year or later. Although the majority of students dropped out in their fourth semester, before completing their sophomore year, these students were identified as high risk at least one semester earlier. For those who dropped out during their sophomore year (in the third or fourth semester), the high-risk period typically began from the spring semester of their freshman year. Additionally, it is evident that students who dropped out after their junior year (post-fifth semester) were, on average, identified as high risk from their first year.

Table 1: Average of high-risk semester periods per semester dropped out

Semester in which they dropped out	1	2	3	4	5	6	7	8	9
Num. of Student	2	30	38	36	14	15	11	17	10
Average number of high risk semesters to drop out	1	1.10	1.92	2.78	3.21	3.80	4.55	5.71	5.60

In addition, Table 2 displays the percentages arranged with the semester of dropout on the vertical axis and the duration from when the student was identified as high risk to the semester of dropout on the horizontal axis. For instance, among students who dropped out in semester 3, 13.2% were identified as high risk before the third semester, 65.8% before the second semester, and 21.1% before the first semester. Importantly, this table also shows that students who dropped out during their junior or senior years (from the fourth semester onward) were consistently identified as high risk from as early as their freshman year. Specifically, the percentages of students who were high risk from their first year are as follows: 78.9% in the third semester, 66.7% in the fourth semester, 50.0% in the fifth semester, 40.0% in the sixth semester, and 54.5% in the seventh semester. These rates continue at 41.2% for students who dropped out in the eighth semester and 40.0% for those in the ninth semester. These high percentages underscore that the risk of dropping out is established early and remains significant even for students who drop out later in their academic careers.

Table 2: Crosstabulation table of semesters dropped out and high-risk periods

	number of high risk semesters to drop out								
Semester in which they dropped out	1	2	3	4	5	6	7	8	9
1	100.0%								
2	90.0%	10.0%							
3	21.1%	65.8%	13.2%						
4	2.8%	30.6%	52.8%	13.9%					
5	14.3%	7.1%	28.6%	42.9%	7.1%				
6	13.3%	6.7%	13.3%	26.7%	33.3%	6.7%			
7	9.1%	18.2%	0.0%	9.1%	9.1%	54.5%	0.0%		
8	0.0%	5.9%	5.9%	17.6%	5.9%	23.5%	29.4%	11.8%	
9	20.0%	10.0%	0.0%	0.0%	0.0%	10.0%	20.0%	40.0%	0%

6 Conclusion

This study has introduced a framework for assessing the probability of dropout, aiming to predict the likelihood of a student's withdrawal from school with significant lead time. Our analysis of dropout probabilities across different semesters for the students who withdrew (173 in total) has revealed that, on average, these students were already identified as high risk approximately 18 months before their actual dropout. Moreover, the investigation into the semester-wise characteristics of dropout has elucidated that over half of the students who withdrew after their second year were already at an increased risk from their first year.

Looking ahead, our research will be dedicated to identifying the optimal times for intervention within the pre-dropout timeline to foster more effective dropout prevention measures.

Acknowledgement

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research C 23K02668.

References

[1] Ministry of Education, Culture, Sports, Science and Technology. [Survey on the status of student enrollment (drop-outs and leaves of absence) (as of the end of the 2022 academic year)] "Gakusei no shugakujokyo (Chutai, Kyugaku) tou ni kansuru Chosa no Kekka ni Tsuite (in Japanese)". 2023.

[2] Yomiuri Shimbun Kyoiku Network, [University Competencies 2019] "Daigaku no Jitsuryoku 2019 (in Japanese)",Chuo Koron Shinsho, 2018..

[3] Ministry of Education, Culture, Sports, Science and Technology, [Withdrawal Status of Six-Year Programs of School of Pharmacy - Results of the 2023 Survey] "Yakugakubu no 6nensei Katai ni okeru Taigaku Jokyo tou – 2023nen (Reiwa 5nen) do Chosa Kekka" , 2024; https://www.mext.go.jp/content/202309011-mxt _igaku-000027071_01.pdf

[4] A. Seidman, "Taking Action: A Retention Formula and Model for Student Success," College student retention: Formula for student success, pp. 267–284, Jan. 2012.

[5] M. A. Bingham and N. W. Solverson, "Using Enrollment Data to Predict Retention Rate," J. Stud. Aff. Res. Pract., vol. 53, no. 1, pp. 51–64, Feb. 2016.

[6] C. E. Lopez Guarin, E. L. Guzman, and F. A. Gonzalez, "A Model to Predict Low Academic Performance at a Specific Enrollment Using Data Mining," IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, vol. 10, no. 3, pp. 119–125, Oct. 2015.

[7] Nobuhiko Kondo and Toshiharu Hatanaka., [Modeling of Learning Process based on Bayesian Network] "Bayesian Network niyoru Shugakujoutai Suitei Model no Koutiku (in Japanese)". Nihon Kyoikukougakkai Ronbunshi (Japan journal of educational technology), vol. 41, no. 3, pp. 271–81, 2018,

[8] Naruhiko Shiratori, Tetsuya Oishi, Shintaro Tajiri, Masao Mori, Masao Murota, [Making Dropout Patterns Using Transition of Dropout Probability] ”Chutaikakuritsu no seni wo motiita chutaigakusei no ruikeika (in Japanese)”. Nihon Kyoikukougakkai Ronbunshi (Japan journal of educational technology), vol 44, no.1, pp. 11-22, 2020.

[9] Naruhiko Shiratori, “Typology of students graduating from college using dropout probabilities,” IIAI Letters on Institutional Research, vol. 3, p. 1, 2023.