JOURNAL OF THE ASIA PACIFIC ECONOMY
https://doi.org/10.1080/13547860.2024.2306673

Routledge
Taylor & Francis Group

Check for updates

Statistical identification of predictors of dropout in secondary education: evidence from Malaysia

Rosna Awang Hashim[a], Hock-Eam Lim[b], Mohd Fairuz Jafar[a], S. Kanageswari Suppiah Shanmugam[a] and Nurliyana Bukhari[a]

[a]School of Education, Universiti Utara Malaysia, Sintok, Malaysia; [b]School of Economics, Finance and Banking, Universiti Utara Malaysia, Sintok, Malaysia

ABSTRACT

The issue of school dropout has been long rooted in Malaysia. To address this issue effectively, further insight into the predictors of the school dropout and the early statistical identification of dropouts are crucial. Nevertheless, uncertainty has been persistent on these two aspects. This paper aims to identify the predictors and assess the feasibility of early identification of school dropouts. Using a school administrative database that comprised data on more than 100,000 students, the school dropout predictors, such as gender, academic achievement, and household income were identified. The estimated model correctly identified early on the 20.83% of the school dropouts. This is substantially higher than the overall sample mean of school dropouts (5%). Thus, it is feasible to identify early on the school dropout statistically. The findings have provided insightful inputs that could strengthen our intervention strategies and policies to help alleviate the problem of school dropout in Malaysia.

ARTICLE HISTORY
Received 30 May 2023
Accepted 14 January 2024

KEYWORDS
School dropout; predictors; statistical early identification; secondary education in malaysia

Introduction

Education is an investment as on its hand rests the nation's economic growth and sustainable development. We believe that educating the youths, who form the backbone of the country, is universally viewed as investing in the future of the country. Youths with higher future earnings are the key agents of social change, economic development and technological innovations. The youth of a country, being the country's significant human resource, naturally warrants that governments worldwide have a vested interest in the growth and livelihood of their youths by educating them.

One of the United Nations most recent global development goals, the Sustainable Development Goals (SDG) 4 for quality education stipulates in Goal 4.1, that provisions need to be made for all girls and boys to complete free, equitable and quality primary and secondary education by 2030. Universal Primary Education, which is aligned to the SDG 4, gives particular emphasis on the importance of inclusive and equitable quality education, and has underscored the requirement for children everywhere, boys

CONTACT Hock-Eam Lim lheam@uum.edu.my
© 2024 Informa UK Limited, trading as Taylor & Francis Group

and girls alike to complete a full course of primary schooling by 2015. On the other hand, Universal Secondary Education prioritises compulsory secondary education for all students between 10 years six months and eighteen years (Sachs et al. 2022). Thus, the importance of youth in completing their primary and secondary education has long been recognized and various efforts have been made to achieve this.

However, the school dropout problem is impeding efforts to achieve the noble objectives of SGD 4, especially the goal of secondary education for all. Since 2008, the downward trend of the global school dropout rate has slowed down. Then, in 2018 it stagnated at around 35%, 16% and 8% for upper, lower secondary and primary schools respectively (UIS 2019a). The two-digit rates translate into millions of dropouts and highlight the seriousness of the secondary school dropout problem.

By levels of income, there are substantial gaps found in the school dropout rate. The high-income countries, i.e. the developed countries have recorded the lowest dropout rate, at 2.5% and 7.8% for lower and upper secondary schools respectively. The dropout rate increased substantially as the levels of income became lower, i.e. in developing countries. The upper, lower-middle, and low-income countries were reported to have upper and lower secondary school dropout rates of 19.6% (7.2%), 43.6% (17%) and 60.8% (38.5%) respectively (UIS 2019a). The high school dropout rates were projected, the projection horizon was from 2018 to 2030, to remain stagnant (UIS 2019b). Thus, the high/secondary school dropout rates are seen as likely to be unchanged in the future among the developing countries.

Malaysia, as one of the developing countries in Asia, is facing the problem of secondary school dropout. In its 2018 Annual Report the Malaysia Education Blueprint 2013–2025, the Malaysian government has expressed its concern on the issue of dropouts and decreasing enrolment in secondary schools (Ministry of Education 2018a). Statistically, Malaysia has only 82.46% of secondary school enrolment in 2020, which is substantially lower than the targeted objective of 100% (UIS 2022). The percentage of secondary school enrolment in Malaysia has been decreasing over the last few years (2014 to 2020), from the peak of around 85% (2014) to 82.46% (2020).

Malaysian society has also raised its concerns over the rise in the number of secondary school dropouts. For instance, there were calls to study the root causes of the problem of secondary school dropout (Mokhtar 2023). This was especially in calls for government intervention to carry out studies on the problem of school dropout among students from the low-income households (for example, the B40 households, i.e. the households in the bottom 40th income percentile rank, see Appendix A for details). Indeed, the B40 students have been singled out as the most vulnerable to the risk of becoming school dropouts (Krishnan 2021).

The problem of the secondary school dropout is not new to Malaysia. The first comprehensive school dropout survey, the 1973 Drop-out Study was conducted by the Ministry of Education, Malaysia. The study found that the dropout out rates were around 27%, 40%, and 50% for English, Malay, and Chinese medium secondary schools respectively (Lee 1984). Then, two decades ago, as reported in Siti Eshah, Wong, and Haniz (2016), the average dropout rate in Malaysian secondary schools was reduced, but still recorded a two-digit high, 13.9% and 10.2% for the years between 2001–2004 and between 2005–2009, respectively. Undeniably, the issue of

school dropout has been an old problem and indeed, various attempts have been made by the Malaysian school authorities to address this persistent predicament. However, the school dropout problem still prevails today. Thousands of students have been dropping out from secondary schools every year in Malaysia (Patel 2014). The dropout problem in secondary school has not been solved adequately.

The school dropout problem could cause considerable negative impacts on the students themselves, their families and society. For instance, it will lead to poor literacy, a decrease in income, an increase in crime activities and add to the government's social welfare burdens. These negative impacts were found to be predominant amongst disadvantaged children, i.e. students from poor families. Like what has been described by Thomas Jefferson, as quoted in Sharifah et al. (2011), "*There is nothing more unequal than the equal treatment of unequal people*". The children of poor families are therefore, most likely to be school dropouts. This will hamper the Malaysian government's efforts to achieve the vision of an inclusive and equitable education for all.

In this regard, the problem of school dropout in Malaysia needs to be addressed, and it is more than just identifying its predictors. Indeed, previous studies about the problem of school dropout in Malaysia have already identified some of the predictors of students dropping out of schools. However, due to the constraint of data availability, some predictors might have been overlooked, more specifically the income of B40 households. There also remains uncertainty on the feasibility of using the identified predictors to develop a statistical prediction model for the early identification of the school dropout. Thus, this paper aims to fill these gaps, by using a more reliable and larger sample obtained from the administrative records of schools.

There are some crucial contributions of the present research reported here. First, to our knowledge, this is the first school dropout study in Malaysia that employs a large sample that comprised actual school dropouts, and this was used to estimate the predictors, develop a statistical prediction model, and evaluate its early identification feasibility. Second, the same students in the study sample were observed over two time periods, i.e. in 2018 and 2019. This has enabled the use of independent variables with values that happened in 2018, before the occurrence of the dropping out of school in 2019. This minimizes the potential for endogeneity bias in estimation. Third, the B40 income classification by the Malaysian government was taken into consideration by piece-wise specification. This provided a further insight on the effect of income on the school dropout rate, and was consistent with the Malaysian government policy to help B40 households. Malaysian society has expressed their concerns about the high school dropout rate among B40 households. However, it appears that there has been no information about the status of the dropout rate among B40 families in Malaysia. Thus, the findings of the present study will be critical in enhancing the design of an effective intervention policy for the problem of school dropout, which has yet to be solved adequately in Malaysia.

Literature review

Definition of school dropout

In the literature, there have been various definitions of school dropout (see Mohamed Irshana, Sivapalan, and Junaenah 2017). The dropout has been defined as someone

"…who have not completed a cycle of basic education, which depending on the compulsory age of enrolment…" (Hunt 2008, 1). The dropout also has been defined as "…a child who has enrolled in any year who fails to enrol in an immediately succeeding year before completing the prescribed cycle…" (Ministry of Education Malaysia 1967, 157). The former has defined the school dropout from a population proportion perspective, whereas the latter from a student cohort perspective. The present study has adopted the student cohort perspective, where the dropouts in secondary education are the students who have exited from the secondary education system before completing their compulsory secondary education, and excludes those who are in the age of compulsory secondary education enrolment, but have never been enrolled for secondary school education. This is consistent with the definition used by the Ministry of Education, Malaysia.

Predictors of school dropout

In order to address the school dropout problem, a thorough understanding of the nature and predictors of school dropout are necessary. According to Hunt (2008), to drop out is a process instead of a consequence of one single identified event, and thus, it has more than one predictor. Previous studies have identified the various school dropout predictors. Overall, the predictors could be categorized into the following four groups: student, family, school, and community-related predictors (see the reviews by Rumberger and Lim (2008), De Witte et al. (2013), Gubbels, van der Put, and Assink (2019) and Mohammad Faisal et al. (2023)).

Among the student-related predictors, one of the prominent predictors is academic achievement. Students who are not able to meet the minimum standards in reading, writing and counting are more likely to drop out due to school disengagement. Gubbels et al. (2019) performed a meta-analysis on the findings of previous studies and synthesized 23 dropout risk predictors. The academic related predictors, such as academic achievement, grade retention and learning difficulties were found to have the largest effect on students dropping out of school. Rumberger and Lim (2008) reviewed a total of 203 empirical studies in the U.S. high school dropout context, and most of these studies found that academic achievement was a significant predictor. A similar finding is also corroborated in recent reviews of studies by Farwis (2020), Ngamba et al. (2021), and Mohammad Faisal et al. (2023).

In Malaysia, poor academic achievement has been found to be the main reason that parents have cited as the cause of their children's school dropout problem (Patel 2014). Another Malaysian study by Siti Eshah et al. (2016) also found that academic achievement was a crucial predictor. Intuitively, as described by Nath et al. (2017), academic achievement was one of the adverse situations that could arise in schools which would then push students out from the school. Thus, academic achievement proxies a student's integration with the school academic environment. It is therefore, not a surprise that academic achievement is a prominent predictor of school dropout.

On the other hand, the other student-related predictors are found to be academic aspirations, grade retention, school attendance or absenteeism, problematic behaviour

(discipline problems), and some demographic characteristic such as gender and ethnic groups (De Witte et al. 2013; Mohammad Faisal et al. 2023).

Among the family-related predictors, socio-economic status is the frequently identified predictor for school dropout. Using a systematic literature review approach, Mohammad Faisal et al. (2023) pointed out that the family's economic status such as income, was mostly found to be a predictor for school dropout. The school dropout phenomenon occurs disproportionately among students from low-income families. In Malaysia, poverty has been consistently found to be a predictor of school dropout among the indigenous students (see Sheera and Azizan 2017), and the majority of dropout students are found to be from low-income families (Patel 2014). This hampers the Malaysian government's effort to increase inter-generation upward social mobility for the poor, and highlights the fact that more effort is needed (such as early identification) to solve the dropout problem effectively. The other identified family-related predictors of school dropout are the parents' educational level, employment, and family size (see Mohammad Faisal et al. 2023).

With respect to the school-related predictors, factors such as distance of school from home, free lunch status, expulsions, school size and type, variety of courses, school resources, school equipment, and number of staff, are commonly found to be reliable predictors. For example, Haepp and Lyu (2018) found a negative effect of school removal on student's school completion rate (i.e. dropout) in China and the effect was worsened if the distance of new schools was far away from students' home. The type of school, such as whether it was private or public funded, religious, grammar, vocational or technical, could be a proxy of the school's heterogeneity in resource and academic environment, and was found to be closely related to the school dropout rate (De Witte et al. 2013; Nath et al. 2017; Rumberger and Lim 2008). Moreover, these school-related predictors were found to be more predictive of secondary school dropout than primary school dropout (Bella and Mputu 2004).

With regard to the community-related predictors, students who live in a neighbourhood that has a lack of conducive facilities, such as playgrounds and green areas, mostly low-income, minority and poor areas, are more susceptible to the school dropout phenomenon (De Witte et al. 2013; Mohammad Faisal et al. 2023). In addition, community peers, traditional and cultural values (such as endorsement of early marriage) have a significant influence on the school dropout rate (Nath et al. 2017). Some of the community-related predictors could be closely linked to the student-related predictors, such as ethnicity and gender. For instance, the findings of Kumar et al. (2023) showed that girls were more likely to dropout from school in India; whereas in Malaysia, it was the case that the girls who were less likely to dropout (Nagaraj et al. 2016). This could be due to the different cultural values on gender between these two countries. Ethnicity could be connected to low-income and minority communities.

In a nutshell, there are multiple predictors of school dropout which comprise of student, family, school, and community-related factors. This is consistent with the ecological system perspective where the school dropout is influenced by external factors that surround the students, such as their family, school, and the community, as

well as internal factors that are related to the students themselves (López, Cárdenas, and González 2021).

Early identification of school dropout

While there is an abundance of evidence in the literature which has provided a rich description of predictors associated with "idle" youths who are neither in school nor employed, very little is done on studying early detection models especially among marginalised students, i.e. students from low-income families in Malaysia. There are a few Malaysian studies, for example, Lee (1984), Patel (2014), and Sheera and Azizan, (2017) that have examined the effects of family income on school dropout. However, their income scales were not similar to the ones measured in the B40, M40 and T20 classifications. Moreover, no attempts were made by these studies to develop an early detection model.

As pointed out by Singh and Alhulail (2022), the school dropout has been a major concern for education administrators and authorities; yet the ability to early identify the school dropout remains a major challenge. The early identification of school dropout is not an easy task. Outside of Malaysia, there have been numerous studies on the early identification of school dropout, especially in the developed countries. For example, Heppen and Therriault (2008) develop an early warning system tool for school dropout using the predictors of absenteeism rate and academic achievements (long course, GPA, course failed, and credit earned), and Henry, Knight, and Thornberry (2012) construct the school disengagement warning index based on the predictors of standardized test scores, attendance, failing subjects, school suspensions, and grade retention. Using variables from school administrative data, Adelman et al. (2018) have illustrated the feasibility of school dropout prediction statistically.

In Malaysia, Wong et al. (2022) have proposed utilizing the machine learning algorithms method to early identify dropouts of primary and secondary school students. Since the method can derive the patterns of dropout and make decisions by analyzing the large dataset without substantial human intervention, it is claimed to be a feasible method that requires the least utilization of school resources (Wong et al. 2022). Indeed, due to the large population of students, a statistical prediction model can be used as a quick screening tool to narrow down the potential dropouts to a small number that becomes manageable in light of the school's limited resources. However, Wong et al. (2022) do not estimate or evaluate the feasibility of a statistical prediction model using actual school dropout data.

In other words, it is crucial to statistically screen out the potential dropout students from a large pool of students and bring them to the close attention of school authorities, and in a number that can be well managed by the school authorities. Thus, early identification could enhance the effectiveness of current Malaysian government policies in addressing the persistent secondary school dropout problem.

The use of a statistical prediction model to early identify those of high risk from a large population has been widely implemented in other fields of study. For example, in developing an active labour market policy to target those who are more likely to be long-term unemployed (from a large pool of registered unemployed people), the

statistical prediction model is utilized as a tool to early identify those who are likely to be long-term unemployed, i.e. likely to claim the maximum amount of unemployment benefit (Black et al. 2003; Lechner and Smith 2007; Lim 2008; Nativel 2004; Payne and Payne 2000). This early identification enables the authority to allocate their limited places of re-training programs to those who need it the most. This reduces the number of long-term unemployed.

In Malaysia, the Ministry of Education has developed an early identification model named "*Murid Berisiko Cicir di Sekolah*"[1] (Ministry of Education 2018b). The model is based on the following seven predictors: student-related (attendance, academic achievement, and number of disciplinary offences committed), and family-related (distance from school, household income, parental status, and disability status). Certainly, this is a good attempt to identify early the potential school dropout. However, to our knowledge, there are no studies or published reports on the effectiveness of the ministry's early identification model, probably due to data unavailability. Indirectly, the effectiveness could be evaluated from two perspectives. First, evaluate the relative importance of the seven predictors that have been included in the model. Are they the most influential predictors? Second, the seven predictors are assigned equal weights in predicting the probability of dropout, i.e. assuming that their marginal effects are similar. Is this assumption valid statistically?

Previous studies have clearly shown that the marginal effects of the school dropout predictors are differential (see Singh and Alhulail 2022). Hence, the equal weight assumption could be restrictive and leads to the issue of questionable predictive power. To relax the equal weight assumption, the estimated effects of the school dropout predictors could serve as weights for the predictors. This requires an empirical study of the actual school dropout, i.e. a statistical prediction model where the estimated marginal effects (effects on the school dropout predictors) can be used as weight in performing the early identification of school dropout.

Research gaps

The present study has identified a few research gaps. First, there are some studies which have examined the school dropout phenomenon in Malaysia, but utilize a relatively small sample or aggregate data. These include for instance, school dropout of 150 (Patel 2014) and 169 respondents (Sheera and Azizan 2017), district (Siti Eshah et al. 2016) and national aggregate data (Nagaraj et al. 2016). The relatively small sample has hampered a more contemporary and insightful understanding of the dropout issue. Statistically, a study with a large sample is needed.

Second, the effects of household income by percentile rank of the bottom 40% (B40), middle 40% (M40) and top 20% (T20), have yet to be explored. In Malaysia, the B40 income classification will lead to important policy implications. The development plan and funding of the government are developed by focusing on the B40 household (Nor Samsiah et al. 2020). Thus, it is imperative to examine the effect of the B40, M40 and T20 household income on school dropout.

Third, uncertainty still remains on the feasibility of early identification of school dropout in Malaysia. The effectiveness of the Education ministry's early identification model

is yet to be evaluated. A more comprehensive early identification model and its feasibility for implementation, that which is based on the school administrative data, is needed.

To fill the gaps, we have identified the predictors of actual secondary school dropout in Malaysia, by including the B40 household income as one of predictors, using a large school administrative data. Then, the identified predictors were used to build a statistical prediction model for school dropout. Finally, there is the clear need to evaluate its feasibility.

Data and methods

Data

The school administrative data were obtained from one of the State Education Departments in Malaysia.[2] The data consists of socio-demographics, academic achievements, and disciplinary records of secondary school students from Form 1 to 4 (levels of school, see Appendix A for details) for 2018. The school dropout[3] status was obtained in 2019. The data comprised four student cohorts: Form 1 (2018) to 2 (2019), Form 2 (2018) to 3 (2019), Form 3 (2018) to 4 (2019), and Form 4 (2018) to 5 (2019). Table 1 presents the four cohorts and the number of school dropouts.

Combining the four student cohorts, we obtained a cross-sectional sample of 109,671 students, with a total dropout numbering 5,486 (5%). The large sample size, rich with a variety of information has enabled us to do a thorough investigation on the school dropout predictors, thus, growing the body of literature on school dropout in Malaysia.

The predictors made available to the present study were however, contingent on the availability of data. The school administrative data provides student-related variables, such as gender, religions, orphan and disability status, academic performance, and disciplinary offences; family-related variables, such as household income, and number of dependents; and school-related variables, such as levels of schooling, and change in the school's average attendance. In the literature, these variables have been found by previous studies to be the significant predictors of school dropout, either in Malaysia or other countries.[4] Appendix A presents the definition and measurement of variables. The sample characteristics are presented in Appendix A.

Method: predictors of school dropout

The risk of being dropped out of school is a latent variable, i.e. unobservable. The increasing risk of being a dropout at a certain level, will trigger the occurrence of school dropout. The triggered school dropout is observable. Using the latent variable framework, the

Table 1. School dropout.

			School dropout	
Cohort	2018	2019	Freq.	%
1	Form 1	Form 2	1048	3.88
2	Form 2	Form 3	628	2.30
3	Form 3	Form 4	2553	9.11
4	Form 4	Form 5	1257	4.59
Combined			5486	5.00

unobservable risk of dropout can be linked to the observed school dropout. To estimate and predict the risk of being a dropout, due to the binary category outcomes of school dropout, the probability model is employed, i.e. logistic regression (Long 1997).

Assuming that there is an unobservable variable (Y^*) which represents a student's risk being a dropout. The Y^* is associated linearly with student, family, and school-related variables (Xs). Then, we obtain the following Equation (1):

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \beta_8 X_{8i} + \beta_9 X_{9i} + \beta_{10} X_{10i} + \beta_{11} X_{11i} + \beta_{12} X_{12i} + \beta_{13} X_{13i} + \beta_{14} X_{14i} + \beta_{15} X_{15i} + e_i$$

$$i = 1, 2, \ldots, \ 109671 \tag{1}$$

From Equation (1), the predictors (independent variables) are: student-related variables (X_1 = gender, X_2 = religions, X_3 = Disable status, X_4 = Orphan status, X_5 = number of disciplinary offences committed, X_6 = academic performance of English language (compared to Malay), X_7= academic performance of Mathematics (compared to Malay), X_8= academic performance of Sciences (compared to Malay)); family-related variables (X_9 = household income of B40, X_{10} = household income of M40, X_{11} = household income of T20, X_{12} = number of dependents in family); and school-related variables (X_{13} = levels of schooling, X_{14} = types of school, X_{15} = change in school's average attendance).

We can present Equation (1) in the matrix notation as below:

$$Y^* = X'\beta + e \tag{2}$$

where,
$Y* =$ unobservable risk of being a dropout (a column vector of Y^*)
$X =$ independent variables (a matrix of $X_1 - X_{15}$)
$\beta =$ coefficients (a column vector of $\beta_0 - \beta_{15}$)
$e =$ error term (a column vector of e)

The observed outcome of dropout, i.e. the actual occurrence of dropout (Y) takes the value of 1 if the student is "dropout", and 0 if the student "does not dropout". The observed outcome of dropout (Y) is related to the Y^* (which is unobservable) as below:

$$Y_i = \begin{cases} 1 \ \ if \ \ Y_i^* > 0 \\ 0 \ \ if \ \ Y_i^* \leq 0 \end{cases} \tag{3}$$

Assuming that the error term in Equation (2) is logistic distributed, the probability that the student being dropout from school is given as below:

$$Pr(Y = 1 \,|\, \mathrm{X}) = Pr(e < \mathrm{X}' \beta \,|\, \mathrm{X}) = F(X' \beta) \tag{4}$$

where F is the logistic cumulative density function. The maximum likelihood estimation is used to estimate the parameters (β) and predict the probability of dropout from school, using the robust variance estimates (Huber/White/sandwich estimator of variance).

Method: early identification of school dropout

The estimated parameters ($\hat{\beta}$) from Equation (4) could be employed to construct the prediction model of risk of being a dropout from school. Given the values of *Xs* of a student and $\hat{\beta}$, we can calculate the Equation (4), i.e. the predicted probability of school dropout for the student (risk of being a school dropout). Then, the predicted probability can be percentile ranked among the students in a school. Students who are ranked the top, for example, top 1%, can be early identified as the potential school dropout statistically. The out-of-sample (or test sample) predictive power can be used to evaluate its feasibility. A high predictive power indicates a high feasibility.

Results and analysis: predictors

We estimate two logit models (Model 1, without quadratic effects; Model 2, with quadratic effects) for the school dropout and evaluate the goodness of fit of the estimated models. Table 2 presents the results of goodness of fit of the estimated models. Overall, the estimated models were found to have a good fit with the sample: Pseudo R^2 was more than 0.05,[5] p-value of Wald test was almost equal to zero, and percentage correctly predicted was more than 96%. There was also no serious multicollinearity found.[6] Statistically, there was no evidence of specification error in the estimated models.

Table 3 presents the estimated logit models (Model 1 and 2). Model 2 with quadratic effects (number of dependents and disciplinary offences committed) was found to be superior to Model 1. The quadratic effects were found to be significant. The variables that were significant in Model 1, were also found to be significant in Model 2. Furthermore, there were no substantial differences found in the estimated signs (between Model 1 and 2). Thus, the interpretation was based on the estimated Model 2.

Predictors of school dropout

From Table 3, the predictors of school dropout were identified. The significant predictors were found to be gender, religion, orphaned status, disabled status, number of disciplinary offences committed, and academic achievement (student-related); income in B40, number of dependents (family-related); levels of schooling, types of school and change in the school's average attendance (school-related).

Table 2. Results of goodness of fit tests.

	Model 1	Model 2
Pseudo R-squared	0.0670	0.0679
Wald test (*p*-value)	0.0000	0.0000
% Correctly predicted	96.43	96.43
General specification test	0.413	0.168
VIF: mean	1.47	3.30
range	1.00 to 3.34[1]	1.00 to 18.70[2]

Note. 1. Goodness of fit test (Full model verse intercept only model).
2. Excluding the squared number of dependents, the range is 1.00 to 3.34.

Table 3. The estimated logit model.

	Model 1		Model 2	
	Coef.	*p*-value	Coef.	*p*-value
Student-related predictors:				
Dmale	0.448***	0.000	0.450***	0.000
DIslam	0.500***	0.000	0.503***	0.000
Orphan	0.166**	0.024	0.134*	0.068
Disable	0.823***	0.000	0.814***	0.000
Disciplinary offences committed	−0.034	0.606	−0.150***	0.060
BIminusBM[3]	0.006***	0.000	0.0058***	0.000
M3minusBM[3]	0.007***	0.000	0.007***	0.000
SNminusBM[3]	0.008***	0.000	0.008***	0.000
Family-related predictors:				
Number of dependents	0.009	0.450	−0.172***	0.000
IncomeB40 ('00)[2]	−0.006***	0.000	−0.006***	0.000
IncomeM40 ('00)[2]	0.002	0.637	0.001	0.687
IncomeT20 ('00)[2]	−0.010	0.161	−0.010	0.162
Number of dependents	0.009	0.450	−0.172***	0.000
School-related predictors:				
Form 1 to Form 2[1a]	−0.887***	0.000	−0.872***	0.000
Form 2 to Form 3[1a]	−1.012***	0.000	−0.997***	0.000
Form 4 to Form 5[1a]	−1.648***	0.000	−1.650***	0.000
SMK2018[1b]	0.492***	0.000	0.496***	0.000
SBP2018[1b]	−0.654***	0.000	−0.652***	0.000
Change in school's average attendance[4]	−5.485***	0.000	−5.436***	0.000
Quadratic specification:				
Squared of number of dependents			0.016***	0.000
Squared of disciplinary offences committed			0.009***	0.007
Constant:				
_cons	−3.344***	0.000	−2.901***	0.000

Note. 1. Comparison group.
a. Form 3 to Form 4.
b.Other types of school consists of the following: SABK (Government-funded Religious School), SMKA (National Islamic Secondary School), SMPK (Special Education National Secondary School), SMT (Technical Secondary School), and S. Model Khas (Special Model School).
2. Household income where 1 unit equals RM100. Spline specification of the following income classification for Kedah state: B40 (below RM4850), M40 (RM4850 to 10959), T20 (RM10960 and above).
3. The marks of the Malay language as a benchmark. BIminusBM = English language marks minus Malay language marks; M3minusBM = Mathematics marks minus Malay language marks; SNminusBM = Sciences marks minus Malay language marks.
4. Changes in the percentage of students attending school (aggregate at school level).
5. ***, **, and * represent the 1%, 5% and 10% level of significance, respectively.

Student-related predictors

Males were found to be more likely to dropout from school than females. In Malaysia, over time, the secondary school completion rate of females has been found to be higher than that of the males (Siti Eshah et al. 2016). A similar result was also found by Nagaraj et al. (2016), a study using a cross-sectional data census sample (over the years of 1970, 1980, 1991 and 2000). The gender differential might be due to the higher ability of females to deal with problems more effectively, and better coping skills of females to meet learning challenges in the context of the Malaysian school education system (Nagaraj et al. 2016). Thus, gender could be a proxy for some underlying and unobserved variables important to the school dropout.

Academic achievement of the Malay language was used as a benchmark for achievement in the English language, Mathematics and Sciences.[7] The estimated positive impacts of academic achievements (see Table 3) show the importance of academic achievement in the Malay language. The low achievement in the Malay

language will increase the student's chances of becoming a dropout.[8] As pointed out by Patel (2014), Malay is the language of instruction in the secondary school, its proficiency facilitates the student's ability to cope with teaching and learning in school. Indeed, academic achievement which represents the student's academic integration with school, has been consistently found to be a significant predictor, either in or outside Malaysia (see reviews by Rumberger and Lim 2008; Farwis 2020; Ngamba et al. 2021; Mohammad Faisal et al. 2023).

We also found that the number of disciplinary offences committed in school had a U-shaped effect on the probability of becoming a dropout. The other student-related predictors were religion (compared to non-Islam, Islam students were more likely to drop out), orphaned status (compared to non-orphaned, orphaned students were more likely to drop out), and disabled status (compared to non-disabled, disabled students were more likely to drop out).

Family-related predictors

The B40 household income has been found to have a significant effect on whether a student dropped out of school (while the M40 and T20 household income were found to be insignificant). For students from B40 families, the increase of household income will decrease their likelihood of dropout. This is consistent with the findings of previous studies that school dropout was largely disproportionate for low-income families (Patel 2014; Sheera and Azizan 2017). Nevertheless, this favourable effect of income was only limited to income in the B40 category. The increase of income beyond the B40 (i.e. increase to M40 and T20 income) has insignificant impacts. This finding supports the Malaysian government policies that have focused on B40 families to eradicate poverty and solve the dropout problem. The number of dependents in a family was found to have a significant U-shaped impact on the likelihood of dropout.

School-related predictors

Comparing students who were in transition from Form 3 to 4, students who were in transition from Form 1 to 2, Form 2 to 3, and Form 4 to 5, were found to be less likely to drop out. Thus, Form 3 to 4 was seen as the key transition phase of students dropping out from secondary school in Malaysia (lower to upper-secondary). This finding is consistent with that in Patel (2014). Completion of lower-secondary education could be seen as a critical point to decide either to continue schooling or dropout to work.

The other significant school-related predictors were the types of school, for example, students from a school with full boarding facility were found to be less likely to dropout; and as for a school's average aggregate attendance rate, it has been shown that students from schools experiencing an increase of the average attendance rate were less likely to drop out.

The estimated probability of dropout

To gain further insight into the quadratic effects of the number of disciplinary offences committed and the number of dependents in a family, we estimated the

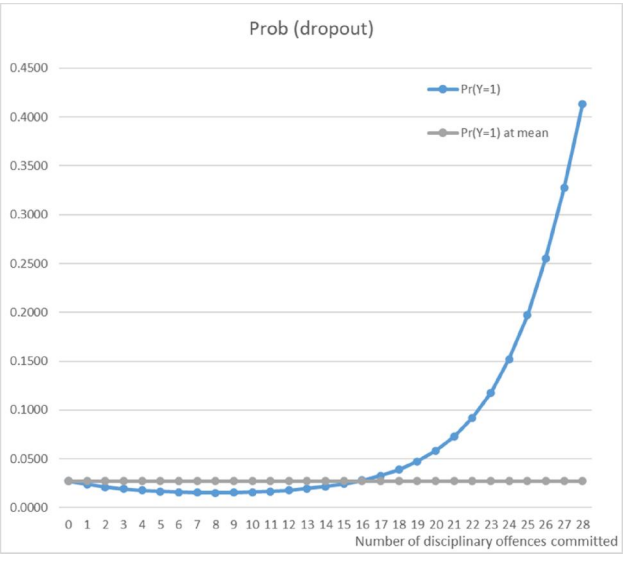


Figure 1. Probability of dropout and disciplinary offences committed.

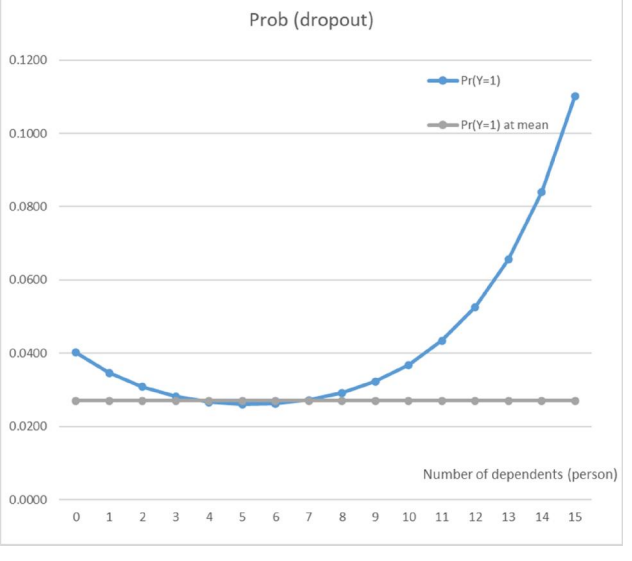


Figure 2. Probability of dropout and number of dependents.

probability of dropout over the values of sample range, holding the values of other variables at mean respectively.[9] Figures 1 and 2 depict the results of the estimated probability of dropout.

Figure 1 shows the effect of the number of disciplinary offences committed on the probability of dropout. The effects were found to be reversed L-shape than inverted U-shape. Thus, the negative effects (at the low numbers of offences) were substantially smaller than the positive effects (at the high numbers). This is clearly shown by the slope's steepness of the reversed L-shape curve (see Figure 1). Quantitatively,

from one to seven offences, the probability of dropout was decreasing at almost a flat rate. However, as the offences increase from eight offences and onwards, the probability of dropout increased at an increasing rate. At the highest number of offences (28), the probability of dropout was estimated to be 41.35%. This implies that students who committed disciplinary offences, their probability of dropout is initially not substantially affected. Nevertheless, if the number of offences committed is increasing continuously, it will eventually trigger the dropout. This finding suggests that students who were approaching a total of eight disciplinary offences committed should be given extra additional attention to prevent them from becoming dropouts.

On the other hand, the number of dependents showed a U-shape effect on the probability of dropout (Figure 2). Initially, at a low number of dependents, the increase was beneficial (reduces school dropout). However, when the number of dependents was more than five, the effect became disadvantageous. Quantitatively, when the number of dependents was five, the probability of dropout was at around 2.6% (almost close to the dropout of a sample representative student of 2.7%). However, if the number of dependents increases further, the probability of dropout will increase. For example, when the number of dependents was 15 (the highest), the probability of dropout was 11.01%. This suggests that students who were from families with a high number of dependents (more than five), were at a high risk of becoming dropouts.

In summary, the results have revealed that the high risk students are identified as follows: students who were male, Muslim, orphaned, disabled, in Form 3 (in transit to Form 4), and from the SMK (National Secondary School); committed more than eight disciplinary offences, from a low household income family, there were more than five dependents in the household, showed low academic achievement, especially in the Malay language, and from school experiencing drops in the aggregate student attendance.

The relative importance of predictors

We performed a dominance analysis to evaluate the relative importance of each predictor, i.e. to assess the amount of prediction error that was reduced by each predictor (Azen and Traxel 2009; Luchman 2021). Table 4 presents the results of the dominance analysis. It was found that the level of schooling was the most important predictor, as it accounted for 60.06% of the prediction error reduction. This was followed by the types of school (12.09%), and gender (10.03%). The other predictors (academic achievement, income B40, religion, and others) contributed the remaining percentage. Thus, the top three predictors (level of schooling, types of school, and gender) contributed around 82% of the prediction error reduction. However, the three predictors have not been included in the Education ministry's early identification model. Thus, the model's effectiveness could be at risk due to this potential misspecification.

Results and analysis: early identification

In-sample and out-of-sample

We use the estimated logistic regression of Model 2 (see Table 3) as a prediction model for school dropout.[10] We randomly selected the 1%, 5%, 10%, 20% and 30% of

Table 4. The relative importance of predictors.

Predictors	Standardized Dominance statistics	Ranking
Levels of schooling[a]	0.6006	1
Types of school[b]	0.1209	2
Gender (Dmale)	0.1003	3
Scores Sciences (minus Malay language) (SNminusBM)	0.0496	4
Mathematics (minus to Malay) (M3minusBM)	0.0330	5
Religion (DIslam)	0.0278	6
English (compared to Malay) (BIminusBM)	0.0129	7
Household income B40 (incomeB40)	0.0119	8
Change in school's average attendance	0.0114	9
Number of dependents (persons)[c]	0.0112	10
Disabled (OKU2018)	0.0091	11
Disciplinary offences committed[d]	0.0050	12
Orphaned (Yatim2018)	0.0028	13
Household income M40 (incomeM40)	0.0019	14
Household income T40 (incomeT20)	0.0017	15

Note. The following categorical predictors that are represented by more than two dummy variables are grouped as one in evaluating the relative importance:
a. Levels of school (school transitions): D2018T1, D2018T2, D2018T4.
b. Types of school: SMK2018, SBP2018.
c. Number of dependents: TANGGUNGANPENJAGA2018, sqTANGGUNGANPENJAGA2018.
d. Disciplinary offences: TotalSalah18_19M, sqTotalSalah18_19M.

the students for an out-of-sample evaluation. The remaining observations at 99%, 95%, 90%, 80% and 70% were considered in-sample and were used to estimate the prediction model. Table 5 presents the in-sample and out-of-sample information. In the out-of-sample group, the students' actual dropout outcomes were known, and the students were not used in the estimation of the prediction model. Thus, it resembled a real-life prediction of a student's becoming a dropout. Hence, the predictive power shows the feasibility of the estimated prediction model.

Early identification

The prediction model (Model 2) is estimated based on the sample Set 1 to 5 respectively (see Table 5). The estimated prediction models were used to predict the probability of dropout for all students. Then, by in-sample and out-of-sample, the students were percentile ranked based on their predicted probability of dropout. The students who were ranked top (high probability of dropout) were identified early as high risk dropout, i.e. potential school dropouts. On the other hand, students who were ranked bottom (low probability of dropout) were identified early as low risk dropout, i.e. not potential school dropouts.

Due to the large population of students and limited resources of the school, the statistical early identification could be utilized as a quick screening tool to identify early the students who are potential school dropouts, from the large pool of students. This enables an efficient use of school resources. Assume that the school focuses on 1% of its students for dropout prevention,[11] the top 1% students (in percentile rank of predicted probability of dropout) are identified early as potential school dropouts and the bottom 1% students are identified early as not a potential school dropout.

Table 5. In-sample and out-of-sample.

		Sample		Dropout	
Set	Type	%	Size	Freq	%
1 (1%)	Out-of-sample	1	947	31	3.27
	In-sample	99	93790	3350	3.57
2 (5%)	Out-of-sample	5	4737	168	3.55
	In-sample	95	90000	3213	3.57
3 (10%)	Out-of-sample	10	9474	332	3.50
	In-sample	90	85263	3049	3.58
4 (20%)	Out-of-sample	20	18947	656	3.46
	In-sample	80	75790	2725	3.60
5 (30%)	Out-of-sample	30	28421	1837	3.65
	In-sample	70	66316	2344	3.53

Predictive power evaluation

To evaluate the predictive power of the model, we used the hit-miss table (also known as Confusion matrix) that compares the actual dropout status of students to the predicted dropouts (i.e. the early identification). The higher the percentage that the matrix predicted was then matched correctly with the actual dropout status (percentage correctly predicted), the higher the predictive power means that the feasibility of early identification is higher. The predictive power was evaluated for in-sample and out-of-sample. The results are as summarized in Table 6.

From Table 6, the overall percentage correctly predicted was found to be around 57% (ranging from 55% to 59.4%). For example, overall, the Set 2 (out-of-sample) correctly predicted the actual dropout status of 57 (out of the 96 students). Thus, overall, 59.4% of school dropouts were correctly predicted.

In terms of non-school dropout, the prediction models were found to have extremely high predictive power. Their percentage correctly predicted was found to be in the range from 97.9% to 100%. For instance, for Set 1 (in-sample), among the 1876 students, a total of 938 students were identified early as not a potential school dropout (Bottom 1% of the predicted probability of dropout), their actual dropout status showed that 934 of them were not a school dropout. Thus, there was a 99.6% correctly predicted result for the non-school dropout. This high percentage correctly predicted (no dropout) is not surprising. Most of the students are not potential school dropouts. If we use the naïve prediction model, i.e. the model that uses the simple sample mean (95%) of not school dropout to predict the no dropout, we can obtain a 95% correctly predicted. In this context, the estimated prediction model has improved the predictive power by around 5 percentage points.

In terms of predicting school dropout, we found that the prediction power (ranging from 10% to 20.8%) was reduced as compared to not dropout (ranging from 97.9% to 100%). The school dropout is a rare event outcome, i.e. the occurrence of school dropout is substantially rarer than not school dropout (event of dropout occurring infrequently). The school dropout has a low sample percentage. Thus, one cannot correctly predict it with a high percentage statistically. For instance, for Set 2 (out-of-sample), out of the total of 96 students, 48 of them were predicted as potential dropouts; however, it was only 10 of them who actually dropped out from school. Thus, this leads to 20.83% correctly predicted. One might be concerned about the

Table 6. Predictive ability (model 2).

% Correctly predicted:	Set 1 (1%)	Set 2 (5%)	Set 3 (10%)	Set 4 (20%)	Set 5 (30%)
	In-sample				
Overall	0.574	0.574	0.575	**0.577**	0.569
Dropout	0.152	0.150	0.154	**0.154**	0.143
No dropout	0.996	0.998	0.996	**0.999**	0.995
% Correctly predicted:	Out-of-sample				
Overall	0.550	**0.594**	0.568	0.563	0.572
Dropout	0.100	**0.208**	0.147	0.137	0.147
No dropout	**1.000**	0.979	0.989	0.989	0.996

low correctness percentage (20.83%), i.e. on average, the early prediction model could help the school authority to identify only one fifth of the actual dropout.

It is important to note that as dropout is a rare event outcome, statistically, it is impossible to have a high correctness percentage. Compared to the naive model,[12] i.e. the simplest prediction model that uses the sample mean of 5% to predict the school dropout, the correctness percentage will be 5% (it is around 3.5% if we use out-of-sample mean). This implies that by using the naïve model, it is only one twentieth (or around one thirtieth) of actual school dropouts who will be identified. Thus, the percentage correctly predicted (20.83%) was around four times higher than that by the naïve model. Therefore, relatively this is a substantial improvement.

Implications and recommendations

Using a large sample of secondary school students, the estimated logit model has suggested that there could be various school dropout predictors, with differential effects. The top three predictors, namely the levels of schooling, types of school and gender which could contribute to most of the predictive error reductions, had not been included in the early identification model of the Ministry of Education, Malaysia. The early identification model could therefore, have been a case of model misspecification. In addition, its equal weight assumption might be invalid. Thus, it is suggested that the Malaysia government need to re-specify and revise its school dropout prediction model and use the predictors' estimated marginal effects as weight.

The present study has unveiled a more complete set of predictors of secondary school dropout and provided a better understanding the school dropout phenomenon in Malaysia. The predictive power of the proposed estimated model has been found to be around four times higher the naïve model (that uses the sample mean to predict) for the early identification of school dropout. Moreover, the information that is needed to estimate the prediction model can be obtained from the school's administrative record, thus, no additional data is needed. Upon incorporating the estimated parameters into an ICT application (software application), the school authority can predict the risk of its students becoming a dropout at its finger tips.

As proposed by Wong et al. (2022), the ability to carry out early identification will reduce the number of potential school dropouts to a smaller figure that could then be better managed by the school in terms of its dropout preventive measures (such as the 1% of students in the school who are ranked top in the list with respect to their predicted probability of becoming dropouts), and therefore, reduce the need for

further human intervention. Thus, it is strongly recommended that the various vested parties use the proposed prediction model and therefore, more specifically, the government seriously consider using it as a quick screening tool to early identify the potential school dropout from a large pool of students. This will complement the existing policies in reducing the number of school dropouts in the country.

Based on the findings in our analysis, we discuss the implications and recommendations of the result on educational policies, especially in Malaysian context. According to the finding, the low achievement of students in the Malay language among students is a cause for concern in Malaysian education policy. With the majority of the population being Malay, the mastery of the Malay language is crucial in promoting national unity and communication. Poor performance in the Malay language may lead to students feeling disengaged and unmotivated, leading to a higher risk of dropout (Kuldas, Hashim, and Ismail 2015). In response to this issue, the Malaysian government has implemented various policies to improve the standard of education in the country. The Education Blueprint 2013–2025 is one such initiative, it is aimed at transforming the education system to meet the demands of the twenty first century (James, Talin, and Bikar 2022). One of the strategies outlined in the blueprint was to enhance the quality of language proficiency, including the Malay language, by improving the teaching and learning process (Adams and Tan 2022).

The Ministry of Education has also introduced programs such as the "Literacy and Numeracy Screening" (LINUS) program, which was launched in 2010 to identify and address early literacy and numeracy problems among primary school students (Lee et al. 2020). The program aims to provide remedial assistance to students who are struggling with the Malay language, as well as other subjects. However, the program was discontinued in 2016, and the Ministry of Education has since replaced it with a new program called the Special Recovery Program (SRP). The SRP program aims to provide more targeted intervention and assistance to students who are struggling with their studies. The discontinuation of the LINUS program highlights the need for the continuous evaluation and adaptation of education policies in Malaysia. It is suggested that to mitigate the dropout problem, the SRP program has to re-focus its intervention on the literacy of the Malay language and ensure that all students have the functional skills in the language.

Next, the present paper has found that the transition from Form 3 to Form 4 in Malaysian secondary schools is a critical period for students, and it has been identified as a high-risk time that leads to the dropout phenomenon. The transition from a lower level to a higher one is marked by changes in curriculum, teaching style, and student expectations, which can be overwhelming for some students. Malaysian education policy needs to emphasize the importance of providing students with quality and equitable education that prepares them for higher education and the workforce. This can be achieved through initiatives such as investing in teacher training programs, providing better teaching resources, and increasing support for schools in low-income areas. With equitable quality education, students will be better prepared for the academic rigours of Form 4 and beyond. Consequently, there needs to be a shift towards more personalized education. Students have different learning styles and needs, and this can affect their ability to cope with the upper secondary demands of Form 4.

Education policy needs to consider the diverse needs of students and provide personalized education that caters to their individual strengths and weaknesses. This can be achieved through initiatives such as the development of student-centred teaching methods, the provision of learning support programs, and the implementation of individualized education plans. Therefore, there needs to be an emphasis on student well-being. The transition from Form 3 to Form 4 can be a stressful time for students, and this can contribute to the dropout rates. Education policy needs to recognize the importance of student well-being and provide support for students to help them cope with the challenges of secondary school (Graham, Powell, and Truscott 2016). This can be achieved by the government through revising its initiatives, such as counselling services, mental health programs, and unwavering support for students with disabilities or special needs.

In the Malaysian context, the government has always recognized the importance of attendance and has implemented policies to improve school attendance rates. However, there is still room for improvement in addressing the root causes of poor attendance and reducing dropout rates in the nation's schools. One of the key factors that contribute to poor attendance and dropout rates is poverty. Students from low-income families may face financial constraints that make it difficult for them to attend school regularly. The Malaysian government has implemented various policies to address poverty, such as the 1Malaysia People's Aid (BR1M, introduced in 2012), and in the state of Selangor, the *Inisiatif Peduli Rakyat* (Caring for the People Initiative) launched in 2008. The latter was aimed at providing financial assistance to low-income families. Besides that, there were also changes at the policy level and one such policy is the Compulsory Education Act 1961, which requires all children between the ages of 6 and 16 to attend school. However, there is still a constant need for more targeted interventions to address the specific needs of students from low-income families and to ensure that they have access to equitable quality education and to mitigate the problem of school dropout.

The relationship between disciplinary offences and the probability of dropout is a complex issue that requires a multi-faceted approach. In Malaysia, the government has implemented several policies aimed at reducing disciplinary offences in schools and ensuring that students who commit offences are provided with the necessary support to prevent them from dropping out. One such policy is the School Improvement Programme (SIP), which was introduced in 2010. Its aim was to improve the quality of education in schools and reduce disciplinary offences. This programme has provided training for teachers and school administrators on strategies for promoting positive behavior and addressing disciplinary issues in schools.

Disciplinary issues in schools are a common problem. They can have a negative impact on the learning environment and student outcomes. One way to address this issue is through teacher mentoring programs that focus on providing teachers with the necessary skills and knowledge to manage and control disciplinary issues in the classroom. Teacher mentoring programs that focus on discipline management can provide teachers with a range of strategies and techniques to manage student behavior in the classroom (Jafar et al. 2021). These strategies can include positive behavior management, effective communication, and conflict resolution techniques. By providing teachers with these skills, they can create a positive learning environment that

encourages student positive behavior and thus, minimizes the likelihood of disciplinary issues. In Malaysia, the government has implemented several teacher mentoring programs aimed at addressing disciplinary issues in schools. These programs include the National Professional Qualification for Education Leaders (NPQEL), which was introduced in 2007. NPQEL was designed to provide teachers with the necessary skills and knowledge to manage disciplinary issues in the classroom (Jafar et al. 2022). It is suggested that there should be more programs which take into consideration the non-linear impacts and that more resources could be allocated to help those students who have committed more than eight disciplinary offences.

Conclusion

Using a large sample of secondary school students, this study has identified a comprehensive set of predictors, namely gender, religion, orphaned status, disability status, household income, number of dependents, academic achievement, years of schooling, student attendance rate at school, and the number of disciplinary offences committed, to study the phenomenon of secondary school dropout in Malaysia. The present study has revealed a range of factors beyond those identified by the early identification model proposed by the Ministry of Education. More specifically, the present study has shown that the high risk students were identified as students who were male, Muslims, orphaned, disabled, in transition from Form 3 to Form 4, were from a school experiencing a low student attendance rate, had low academic achievement especially in the Malay language, were from the family of the B40 income group, and the number of dependents would be more than five, and had committed more than eight disciplinary offences in school.

The findings have highlighted the need for a more complex and nuanced approach in the early identification of dropouts and the prevention of the phenomenon, taking into account the varying effects of the up to date reliable predictors. The study proposes a prediction model that has four times a higher predictive power than a naïve prediction model and can be implemented through an ICT application that requires minimal human intervention.

The findings of the present paper have strengthened our understanding of the school dropout phenomenon and in turn, has even asserted its claim to have provided a feasible early identification model. These advances in the field could provide insightful input that can enhance the strategies to intervene in the complex problem of school dropout in Malaysia. By identifying high-risk students and implementing targeted interventions, education authorities can reduce dropout rates and ensure that all students have access to equitable quality education. However, to be sure other ongoing evaluation and adaptation of education policies and programs are also necessary to ensure that they are effective and responsive to the changing needs of students and society.

There are a few caveats about the findings. First, some school dropout predictors are excluded due to the issue of data unavailability. Second, the early identification model needs to be revised once more up to date data is made available. Third, the model employed a cross-sectional data set and thus, a dynamic perspective of the dropout phenomenon, such as the effects of time-varying variables, were not evaluated. Fourth,

the school dropout phenomenon is a rare event outcome. Statistically, it is impossible to predict it with a high degree of certainty, in other words achieving a high percentage score. It is suitable to use as a quick screen tool and supplement this with the other non-statistical prediction (such as the judgement of a school counsellor). Future studies are suggested to investigate more thoroughly these contentions.

Notes

1. literally, it can be translated as "Students at risk of becoming a school dropout".
2. Due to data confidentiality and sensitivity, the name of the state could not be revealed.
3. According to the State Education Department, the school dropout is defined as "the students who leave school due to reasons such as the inability to continue schooling, are expelled from school, are not interested to go to school, have to work, and are beset with family problems".
4. Some predictors of school dropout such as student motivation, learning support, and community-related factors (Singh and Alhamad 2022; Singh and Alhulail 2022), are not included due to data unavailability.
5. Pseudo R^2 is normally found to be low for categorical dependent variable models that use individual level data. For examples, see Di Tella, MacCulloch, and Oswald (2003), and Killingsworth (2021).
6. Model 2 is found to have high multicollinearity between the number of dependents and its squared. This is expected due to the construction of nonlinear (quadratic) effects. Moreover, since the number of dependents and its squared are found to be significant (see Table 3), the bias of multicollinearity should be at its minimum.
7. Since the Malay language is the national language (school's administrative and communication language as well), the Malay language is used as a benchmark. The impacts of academic achievement without benchmarking show that academic achievement (Malay, English, Mathematics and Sciences) has significant and negative impacts on school dropout.
8. For the impact of each subject separately, by the statistical profiling of high (top 10% in terms of estimated probability of dropout) and low risk (bottom 10%) of dropout students reveal that the average marks of high (low) risk students are 35.99 (67.62), 29.21 (53.43), 29.72 (46.26), and 32.02 (30.09), for Malay, English, Mathematics and Sciences, respectively.
9. If we hold all the variables at mean values (i.e. resemble a sample representative student), the estimated probability was found to be 2.7% (based on estimated Model 2).
10. Theoretically, since the school dropout is measured in binary outcome (0, 1), the logistic regression is an appropriate choice of model (Long 1997). Empirically, the logistic regression is found to be as good as the other predictive models (Nayeem Hasan 2019)
11. The choice of 1% is arbitrary. It could be 10% or other numbers, depending on the school's resources and student population.
12. A naïve model is the simplest model, i.e. using the sample mean to predict the school dropout. In the present study, we could just use the sample mean (5%) to predict school dropout, and the percentage correctly predicted will be 5%. This could serve as a benchmark to evaluate the predictive power of the estimated model. This is like the interpretation of R^2 that use the sample mean as a benchmark (i.e. the sum of squares total, see Gujarati 2004).

Acknowledgments

We would like to thank the Yayasan Sime Darby (YSD) for its financial support in completing this paper (Research Grant Code S/O: 21037). Thanks also due to Universiti Utara Malaysia for its support to this paper. Thanks to Ms. Nor'Ain Mohamad Hadis and others, who have provided their valuable support and comments on the early draft of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Yayasan Sime Darby (YSD) under Grant Code S/O 21037.

Notes on contributors

Rosna Awang-Hashim is a Professor in Educational Psychology at the School of Education, Universiti Utara Malaysia (UUM) and the former Editor-in-Chief of the Malaysian Journal of Learning and Instruction (MJLI). Her research interests include motivation psychology, learning and teaching in higher education, and socio-psychological processes in learning and teaching interactions. Currently, she's the Chair of the Cluster of Education and Human Development of the Academy of Professors Malaysia and Chair of Service Learning Malaysia-University for Society (SULAM), MOHE Malaysia.

Hock-Eam Lim is a professor from the School of Economics, Finance and Banking, Universiti Utara Malaysia. He has obtained his Master degree from University College London and Doctoral degree from Monash University. His research interests are in applied micro-econometrics, economics of happiness and labour economics. He teaches courses in econometrics, sampling methods, research methodology and forecasting, both at undergraduate and graduate level. He has published in various journals either local or international and actively participates in conferences.

Mohd Fairuz Jafar is a senior lecturer at School of Education, Universiti Utara Malaysia. He has obtained his Master of Education and Doctoral degree in Educational Psychology from Universiti Sains Malaysia. Master of Education (M.Ed) specialization in Educational Psychology at Universiti Sains Malaysia (USM). His research interests are in educational psychology, teacher's education, adolescent psychology, psychological well-being, deep learning and quantitative research.

S. Kanageswari Suppiah Shanmugam, served at SEAMEO RECSAM as a Mathematics Education Specialist before joining the School of Education, Universiti Utara Malaysia. She has a doctoral degree in Psychometrics and Educational Evaluation. She is an active member of the National Council on Measurement in Education, the International Community of Emergent Researchers in Mathematics Education (University of Hamburg, Germany). She is also the Associate Member for the Network on Education Quality Monitoring (Asia-Pacific UNESCO Bangkok) and is on the advisory board for the Philippines Basic Education Math Teacher's Society. Her specialisation is cross-cultural assessments, particularly mathematics.

Nurliyana Bukhari received her Ph.D. in educational research, measurement, and evaluation from the University of North Carolina at Greensboro (UNCG) and her Master's in research, measurement, and evaluation from the University of Miami. Her research focuses on issues of fairness in testing; technology integration in assessment; understanding learners; assessment design; validation; and psychometrics estimation.

Data availability statement

Due to confidentiality and non-disclosure agreement that signed between the researchers and school authority, the data could not be shared.

References

Adams, D., and K. L. Tan. 2022. "Malaysian Education System: Progress, Standards and Aspirations." In *Education in Malaysia: Developments, Reforms and Prospects*, edited by D. Adams, 1–18. Oxfordshire: Routledge.

Adelman, M., F. Haimovich, A. Ham, and E. Vazquez. 2018. "Predicting School Dropout with Administrative Data: New Evidence from Guatemala and Honduras." *Education Economics* 26 (4): 356–372. https://doi.org/10.1080/09645292.2018.1433127

Azen, R., and N. Traxel. 2009. "Using Dominance Analysis to Determine Predictor Importance in Logistic Regression." *Journal of Educational and Behavioral Statistics* 34 (3): 319–347. https://doi.org/10.3102/1076998609332754

Bella, N., and H. Mputu. 2004. "Dropout in Primary and Secondary School: A Global Issue and an Obstacle to the Achievement of the Education for All Goals." *International Journal on School Disaffection* 2 (2): 14–30. https://doi.org/10.18546/IJSD.02.2.04

Black, A. D., A. J. Smith, M. Plesca, and S. Shannon. 2003. *Profiling UI Claimants to Allocate Reemployment Services: Evidence and Recommendations for States.* USA: ETA Occasional Papers, Department of Labor, U.S.

De Witte, K., S. Cabus, G. Thyssen, W. Groot, and H. M. van den Brink. 2013. "A Critical Review of the Literature on School Dropout." *Educational Research Review* 10: 13–28. https://doi.org/10.1016/j.edurev.2013.05.002

Di Tella, R., R. J. MacCulloch, and A. J. Oswald. 2003. "The Macroeconomics of Happiness." *Review of Economics and Statistics* 85 (4): 809–827. https://doi.org/10.1162/003465303772815745

Farwis, M. 2020. "Analysis of School Dropout among Secondary School Students: Case of Sammanthrai Education Zone." *International Journal of Academic Multidisciplinary Research* 4 (12): 79–85.

Graham, A., M. A. Powell, and J. Truscott. 2016. "Facilitating Student Well-Being: Relationships Do Matter." *Educational Research* 58 (4): 366–383. https://doi.org/10.1080/00131881.2016.1228841

Gubbels, J., E. C. van der Put, and M. Assink. 2019. "Risk Factors for School Absenteeism and Dropout: A Meta-Analytic Review." *Journal of Youth and Adolescence* 48 (9): 1637–1667. https://doi.org/10.1007/s10964-019-01072-5

Gujarati, D. N. 2004. *Basic Econometrics* (4th Ed.). New York: McGraw-Hill Companies.

Haepp, T., and L. Lyu. 2018. "The Impact of Primary School Investment Reallocation on Educational Attainment in Rural China." *Journal of the Asia Pacific Economy* 23 (4): 606–627. https://doi.org/10.1080/13547860.2018.1515004

Henry, L. K., E. K. Knight, and P. T. Thornberry. 2012. "School Disengagement as a Predictor of Dropout, Delinquency and Problem Substance Use during Adolescence and Early Adulthood." *Journal of Youth and Adolescence* 41 (2): 156–166. https://doi.org/10.1007/s10964-011-9665-3

Heppen, B. J., and B. S. Therriault. 2008. Developing Early Warning System to Identify Potential High School Dropouts. National High School Centre, Betterhighschool.org. https://files.eric.ed.gov/fulltext/ED521558.pdf.

Hunt, F. 2008. "Dropping out from School: A Cross Country Review of Literature." Consortium for Research on Educational Access, Transitions and Equity, Research Monograph 16 https://www.researchgate.net/publication/234629293_Dropping_Out_from_School_A_Cross_Country_Review_of_Literature.

Jafar, M. F., M. F. M. Yaakob, H. Awang, D. M. Zain, and M. Kasim. 2022. "Disentangling the Toing and Froing of Professional Learning Community Implementation by Reconnecting Educational Policy with School Culture." *International Journal of Instruction* 15 (2): 307–328. https://doi.org/10.29333/iji.2022.15218a

Jafar, M. F., M. F. M. Yaakob, R. Mustapha, M. N. A. Aziz, M. R. Yusof, and H. Awang. 2021. "Quality of Mentoring of Mentor Teachers: Perspective of the Trainee Teachers."

International Journal of Evaluation and Research in Education (IJERE) 10 (2): 632–640. https://doi.org/10.11591/ijere.v10i2.21035

James, J., R. Talin, and S. S. Bikar. 2022. "Defining Student's Achievement Based on the Understanding of National Education Philosophy and Malaysia's Education Blueprint (2013–2025)." *Malaysian Journal of Social Sciences and Humanities (MJSSH)* 7 (4): E001401. https://doi.org/10.47405/mjssh.v7i4.1401

Killingsworth, M. A. 2021. "Experienced Well-Being Rises with Income, Even above $75,000 per Year." *Proceedings of the National Academy of Sciences* 118 (4): E2016976118. https://doi.org/10.1073/pnas.2016976118

Krishnan, D. B. 2021, December 10. Sharp rise in dropouts unless govt intervenes. *NST online*. https://www.nst.com.my/news/nation/2021/12/753002/sharp-rise-dropouts-unless-govt-intervenes.

Kuldas, S., S. Hashim, and H. N. Ismail. 2015. "Malaysian Adolescent Students' Needs for Enhancing Thinking Skills, Counteracting Risk Factors and Demonstrating Academic Resilience." *International Journal of Adolescence and Youth* 20 (1): 32–47. https://doi.org/10.1080/02673843.2014.973890

Kumar, P., S. K. Patel, S. Debbarma, and N. Saggurti. 2023. "Determinants of School Dropouts among Adolescents: Evidence from a Longitudinal Study in India." *PLoS One* 18 (3): E0282468. https://doi.org/10.1371/journal.pone.0282468

Lechner, M., and A. J. Smith. 2007. "What is the Value Added by Caseworkers?" *Labour Economics* 14 (2): 135–151. https://doi.org/10.1016/j.labeco.2004.12.002

Lee, M. F. 1984. "Peninsular Malaysia." In *The Drop-out Problem in Primary Education: Some Case Studies*, edited by UNESCO, 118–161. Bangkok: UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000062375.

Lee, J. A., S. Lee, N. F. M. Yusoff, P. H. Ong, Z. S. Nordin, and H. Winskel. 2020. "An Early Reading Assessment Battery for Multilingual Learners in Malaysia." *Frontiers in Psychology* 11: 1700. https://doi.org/10.3389/fpsyg.2020.01700

Lim, H. E. 2008. "Feasibility of Early Identification of Low Employability Graduate in Malaysia." *Malaysian Journal of Economic Studies* 45 (2): 95–112.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications, Inc.

López, V., K. Cárdenas, and L. González. 2021. "The Effect of School Psychologists and Social Workers on School Achievement and Failure: A National Multilevel Study in Chile." *Frontiers in Psychology* 12: 639089. https://doi.org/10.3389/fpsyg.2021.639089

Luchman, J. N. 2021. "Determining Relative Importance in Stata Using Dominance Analysis: Domin and Dome." *The Stata Journal* 21 (2): 510–538. https://doi.org/10.1177/1536867X211025837

Ministry of Education Malaysia. 1967. "Education Statistics of Malaysia 1938 to 1967." The Educational Planning and Research Division, Ministry of Education Malaysia, Putrajaya. https://www.fmm.org.my/upload/educational%20statistics%20of%20malaysia%201938-1967.pdf.

Ministry of Education Malaysia. 2018a. Malaysia Education Blueprint 2013–2025: Annual Report 2018. Ministry of Education Malaysia. Putrajaya. https://www.padu.edu.my/wp-content/uploads/2019/07/AR2018-Eng.pdf.

Ministry of Education Malaysia. 2018b. *Garis panduan mengurus murid berisiko cicir di Sekolah [Guidelines for managing risk of students' dropout in schools]*. Ministry of Education Malaysia, Putrajaya. https://www.moe.gov.my/pekeliling/1935-surat-pekeliling-ikhtisas-bilangan-4-tahun-2018-garis-panduan-mengurus-murid-berisiko-cicir-di-sekolah/file.

Mohammad Faisal, N. N., P. R. Megat Abdul Rahim, N. F. K. Mokh Hatta, and A. H. Abu Bakar. 2023. "Understanding School Dropouts Phenomenon among Youths through Behavioural Trends." *Environment-Behaviour Proceedings Journal* 8 (23): 3–11. https://doi.org/10.21834/ebpj.v8i23.4474

Mohamed Irshana, R., S. Sivapalan, and S. Junaenah. 2017. "Social Environmental Determinants of Student Dropout in the Plantation Settlement." *Malaysian Journal of Society and Space* 13 (2): 54–64.

Mokhtar, M. 2023. "Are we Learning from School Dropout Rates?" Free Malaysia Today, Online News. https://www.freemalaysiatoday.com/category/opinion/2023/03/26/are-we-learning-from-school-dropout-rates/.

Nagaraj, S., K. H. Lee, K. L. Goh, and N. P. Tey. 2016. "Malaysian Adolescents Not in School: The Nexus of Education, Work and Gender." *Malaysian Journal of Economic Studies* 53 (1): 87–113.

Nath, S. R., D. Ferris, M. M. Kabir, T. Chowdhury, and A. Hossain. 2017. Transition Land Dropout in Lower Income Countries: Case Studies of Secondary Education in Bangladesh and Uganda. World Innovation Summit for Education (WISE), Qatar Foundation. https://www.wise-qatar.org/app/uploads/2019/04/rr.3.2017_brac.pdf.

Natível, C. 2004. *Economics Transition, Unemployment and Active Labour Market Policy: Lessons and Perspectives from the East German Bundeslander*. Birmingham: University of Birmingham Press.

Nayeem Hasan, M. 2019. A Comparison of Logistic Regression and Linear Discriminant Analysis in Predicting the Female Students Attrition from School in Bangladesh. IEEE Xplore. https://ieeexplore.ieee.org/document/9068776.

Ngamba, K. H., L. S. Lombo, G. Lombo, and N. Ekira Viviar. 2021. "Causes and Consequences of School Dropout in Kinshasa: Students' Perspectives before and after Dropping out." *Journal of African Education* 2 (3): 175–192.

Nor Samsiah, S., A. Fikri, Z. Ali, M. Zakree, and K. Nadiyah. 2020. "Drop-out Prediction in Higher Education among B40 Students." *International Journal of Advanced Computer Science and Applications* 11 (11): 550–559. https://doi.org/10.14569/IJACSA.2020.0111169

Patel, T. 2014. Dropping *O*ut of *S*chool in Malaysia: What *We Know* and *What Needs* to be *D*one. Institute for Democracy and Economic Affairs (IDEAS), Policy IDEAS No. 14, Kuala Lumpur. https://www.ideas.org.my/wp-content/uploads/2021/04/Policy-No-14-1.pdf.

Payne, C., and J. Payne. 2000. Early Identification of the Long-Term Unemployed. London: PSI Research Discussion Paper 4.

Rumberger, R. W., and S. A. Lim. 2008. *Why Student Drop out of School: A Review of 25 Years of Research*. Santa Barbara: California Dropotu Research Project, University of California.

Sachs, J. D., G. Lafortune, C. Kroll, G. Fuller, and F. Woelm. 2022. Sustainable Development Report 2022. Cambridge University Press. https://s3.amazonaws.com/sustainabledevelopment.report/2022/2022-sustainable-development-report.pdf.

Sharifah, M. N., R. Samsilah, M. Aminuddin, H. A. H. Kamaruddin, M. A. Mohamad Azhar, and A. M. Jaimah. 2011. "Dropout Prevention Initiatives for Malaysian Indigenous Orang Asli Children." *International Journal on School Disaffection* 8 (1): 42–56.

Sheera, N. S., and Z. Azizan. 2017. "Exploring the School Dropout Factors among Indigenous Students in Melaka." *Journal of Administrative Science* 14 (3): 1–13.

Singh, H. P., and I. A. Alhamad. 2022. "Influence of National Culture on Perspectives and Factors Affecting Student Dropout: A Comparative Study of Australia, Saudi Arabia, and Ethiopia." *Archives of Business Research* 10 (11): 287–300. https://doi.org/10.14738/abr.1011.13508

Singh, H. P., and H. N. Alhulail. 2022. "Predicting Student-Teachers Dropout Risk and Early Identification: A Four-Step Logistic Regression Approach." *IEEE Access*. 10: 6470–6482. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9676659. https://doi.org/10.1109/ACCESS.2022.3141992

Siti Eshah, M., T. K. Wong, and I. Haniz. 2016. "Trends and Factors for Dropout among Secondary School Students in Perak." *Journal of Research, Policy & Practice of Teachers & Teacher Education* 6 (1): 5–15.

UIS. 2019a. New Methodology Shows That 258 Million Children, Adolescents and Youth Are out of School. Fact Sheet no.56, UNESCO Institute for Statistics (UIS). https://uis.unesco.

org/sites/default/files/documents/new-methodology-shows-258-million-children-adolescents-and-youth-are-out-school.pdf.
UIS. 2019b. Meeting Commitments: Are Countries on Track to Achieve SDG 4? UNESCO Institute for Statistics (UIS). https://unesdoc.unesco.org/ark:/48223/pf0000369009.
UIS. 2022. School Enrolment, Secondary (% Gross) – Malaysia. UNESCO Institute for Statistics (UIS). https://data.worldbank.org/indicator/SE.SEC.ENRR?locations=MY.
Wong, E. L. M., A. J. Nasir, M. S. Narishah, and I. Izian. 2022. "Dropout Early Warning System (DEWS) in Malaysia's Primary and Secondary Education: A Conceptual Paper." In *International Conference on Information Systems and Intelligent Applications (ICISIA) 2022, Lecture Notes in Networks and Systems*, edited by M. Al-Emran, M. A. Al-Sharafi, and K. Shaalan (vol. 550, pp. 427–434). Berlin: Springer. https://doi.org/10.1007/978-3-031-16865-9_33

Appendix A

Definition and Measurement of Variables

Variable	Definition	Measurement
Dropout	Exit from secondary education before completing it	=1 if dropout, =0 if otherwise
Dmale	Gender	=1 if male, =0 if female
DIslam	Religions	=1 if Muslim, =0 if otherwise
Orphaned	Status of Orphan	=1 if orphaned, =0 if otherwise
Disabled	Status of disable	=1 if disabled, =0 if otherwise
Levels of school	Form 1 to Form 5 (Form 1, 2, 3, 4 and 5 reflects 1st, 2nd, 3rd, 4th and 5th years of study of secondary school. Form 1 to 3 is lower-secondary; Form 4 to 5 is upper-secondary)	Ordinal number
Income group	Household monthly income group: B40: less than RM4850 M40: RM4850-RM10959 T20: more than or equal to RM10960 (According to the Department of Statistics, Malaysia, the household monthly income percentile rank are as follows: the bottom 40% percentile rank is less than RM4850; the middle 40% is between RM4850 to RM10959; and the top 20% is more than RM10960)	Discrete number
Number of dependents	The number of persons in the household that depends on the household head for their livelihood	Discrete number
Malay language	Marks obtained	Percentage (0–100%)
English language	Marks obtained	Percentage (0–100%)
Mathematics	Marks obtained	Percentage (0–100%)
Sciences	Marks obtained	Percentage (0–100%)
Disciplinary offences	The number of disciplinary offences committed as recorded in the school record system (Schools are	Discrete number

(*continued*)

Continued.

Variable	Definition	Measurement
	required to keep a record of the number of disciplinary offences committed by the students in the school)	
Types of school	SABK (Government-funded Religious School), SBP (Fully Residential School), SMK (National Secondary School), SMKA (National Islamic Secondary School), SMPK (Special Education National Secondary School), SMT (Technical Secondary School), S. Model Khas (Special Model School)	Nominal number
Change in school's average attendance	The change in a school's average attendance as compared to the previous year: zero (no change), positive (increase), negative (decrease)	Continuous number

Appendix B
Sample Characteristics

Categorical variables		Freq	%
Form	1 to 2	26,979	24.60
	2 to 3	27,261	24.86
	3 to 4	28,037	25.56
	4 to 5	27,394	24.98
Types of school	SABK (Government-funded Religious School)	6,915	6.44
	SBP (Fully Residential School)	2,237	2.08
	SMK (National Secondary School)	94,592	88.08
	SMKA (National Islamic Secondary School)	2,425	2.26
	SMPK (Special Education National Secondary School)	65	0.06
	SMT (Technical Secondary School - SPM)	297	0.28
	S. Model Khas (Special Model School)	862	0.8
Gender	Female	56,284	51.32
	Male	53,386	48.68
Religions	Others	17,722	16.16
	Islam	91,948	83.84
Orphaned?	No	103,449	94.33
	Yes	6,222	5.67
Disabled?	No	109,193	99.56
	Yes	478	0.44

Discrete/continuous variables	Obs	Mean	Std Dev.	Median
Marks obtained:				
Malay language	115188	46.11	25.83	50
English language	115188	36.45	24.07	37
Mathematics	115188	33.36	22.85	31
Sciences	115188	35.72	22.59	37
Household income (RM per month)	106013	2029.53	2048.34	1300
Number of dependents (persons)	106730	5.13	1.61	5
Disciplinary offences committed	115188	0.06	0.45	0
Change in the school's average attendance*	102743	−0.0034	0.0137	−0.0025

Note *Changes in the percentage of students attending school (aggregate at school level).