OPEN ACCESS

A step-wise multiple linear regression analysis for identifying predictors of employees' intention to undertake further study

Lazim Abdullah, Ashraf Fahmi Rahim

School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, Malaysia

ABSTRACT: The aim of the current study is to identify predictors of intention to undertake further study among employees. Data regarding employees' intention and related variables were collected from five private companies and eight government agencies at a municipality council in Malaysia. In this study, hundred and twenty employees were fully completed the questionnaire. Ten predictors and one dependent variable (Intention) were defined. A series of six-step analyses of step-wise multiple linear regression was performed. The findings show the linearity assumption about the association between predictors and dependent variable (Intention) is fulfilled. A significant portion of the variability in dependent variable (Intention) is explained by multiple regressions on the predictors. A positive 0.817 regression coefficient suggests that motivation level is the most influential predictor in predicting employees' intention to undertake further study.

Introduction

There is a general recognition over the importance of knowledge in developing career path or to the road of wisdom. Knowledge is and has been at the heart of economic growth and the gradual rise in levels of social well-being since time immemorial (David and Foray, 2002). One of the possible ways in seeking knowledge is undertaking further study. The intention of seeking knowledge is essential, especially in a case of those who entered the job market with lower academic qualification. Learning intention among employees refers to the willingness, readiness or even plan of employees to overcome a gap they are experiencing between their current knowledge, skills and attitudes related to the job and those that are required or assumed to be desirable through a specific educational track or training (Kyndt et al., 2011a). There are a variety of reasons or variables that need to considered before having an intention to undertake further study. The intention to pursue study among different level of person is different depending on roles, functions and disposition. For example, among full time young students, family is the most influencing variable to further their international education (Pampa, 2004). Students are tending to follow their family, friends and peers because most of them are financially supported by family and they have no parties to consult other than peers and family (Chen and Zimitat, 2006). According to Wagner and Fard (2009), the most influenced variable to further study in Malaysia is the cost of education followed by value of education, then contents of the course and then family, friends and peers. Intention to undertake further study by a student is governed by so many variables notwithstanding the roles of parents and peers. Many of these variables, perhaps, not so pertinent for students, where most of them relied on parents to finance their study. However, these multiple variables might be useful to other groups of people who have the similar intention to continue their education at higher level.

Other than full time students where they do not receive any wages, those who have been employed and received wages also assumed to have the similar intention to gain knowledge through further education. It is consistent with current needs where today's jobs are becoming more and more knowledge-intensive (Gvaramadze, 2010). Furthermore, undertake further study is much important especially in a situation where employee sometimes aim to climb their career ladder with the help of paper qualification or advanced skills. It has been suggested that higher qualification could guarantee their employability apart from providing evidence that workers experience a responsibility to show a willingness and capability to learn (Illeris, 2006; Onstenk, 1997). In contrast to typical students' life, employee may consider several other variables when they have intention to undertake further study. Merriam Webster (2014) defines an employee is a person who works for another person or for a company for wages or a salary. Unlike students, employee receives wages or salary but at the same time they also have to commit on other needs. It is evident that employee would consider all relevant variables such as salaries, their skills either interpersonal or analytical skills, occupation sector and motivation as all these variables directly impacted their lives and careers. It seems that intention to undertake further study is determined by a combination of multiple variables of employee including attitude toward education. It is also influenced by a number of

training opportunities offered to them. Baert et al. (2006) summarise these variables in their model of a learning climate. They advocate that characteristics of the learner, characteristics of the learning and training activity and the social context can play an important role in the decision-making process and more specifically affect the learning intention of an employee.

To date, however, identifying the influential variables has placed relatively little focus especially on investigating the assumption whether intention to undertake further study can be associated with all the variables. Therefore, this study will focus on the variables that could be related to the learning intention. Many researches have been conducted to search the association between the intention and the related variables using linear models. One of the linear models that purposely used to establish an association between variables is multiple linear regression. Most of the linear regression analyses in social sciences are interpreted based on hypothesis testing. Rejection or acceptance of the predefined hypothesis is made by comparing the empirical results with the critical *F*-value for chi-square distribution cases or the critical *p*-value for normal distribution cases. Summers et al., (2006), for example, identified predictors of purchase intention using *F*-value with 0.05 probability of rejection in testing hypothesis. In contrast to these approaches, this paper studies the significance of the regression coefficients of multiple linear regressions in investigating the association between the intention to further study and the related variables. Alternatively, identifying predictors can also be implemented using step-wise linear regression where all steps, including ANOVA test, *F*-value, excluded variables, regression coefficient test is considered in the analysis. Therefore, this paper aims to identify predictors of the employees' intention to further their study. The associations between the intentions of employee to undertake further their studies and the relevant predictor variables are investigated using a detailed computation of step-wise multiple linear regression analysis.

A Brief Review of Multiple Linear Regression

Linear regression has been known as an approach for modelling the association between a dependent variable and one or more independent variables. In a case where there is one independent variable, the approach is called simple linear regression. In a case where there is more than one independent variable, the approach is called multiple linear regression (Freedman, 2009). Multiple linear regression analysis is used to identify changes in two or more variables that contribute to changes in the dependent variable. It is a statistical technique used to analyse the association between one dependent variable and a set of predictor variable. Predictor variables are the variables that give effect to changes in the dependent variable. Figure 1 shows the predictors that influence dependent variable.
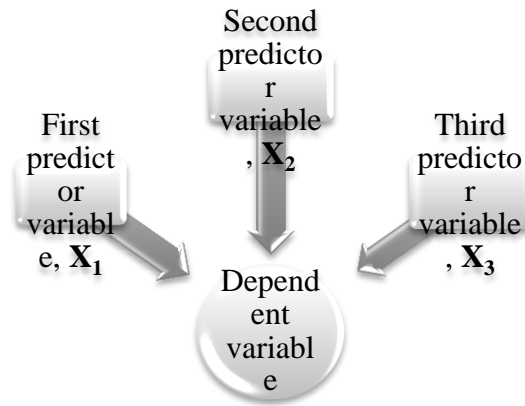

Figure1. Predictors that influence dependent variable.

The number of predictor variables can be more than two, depending on cases.

In multiple linear regression analysis, variable (*X)* named as predictor variable while dependent variable (*Y)* named as criterion variable or response variable. Score of dependent variable (*Y*) is predicted by using *k* predictor variable $(X_1, X_2, X_3, ...$ and $X_k)$, where $k \geq 2$. Multiple linear regression analysis is shown in the following equation.

$Y = b_1 X_1 + b_2 X_2 + ... + b_k X_k + a$,

where *b* is a regression coefficient for each predictor variable, *a* is regression constant in this equation. *Y* is the value for criterion variable or response variable which predicted by predictor variable. Symbol $b_1, b_2, b_3, ..., b_k$ is a regression coefficient for each predictor variable while *a* is a regression constant. According to this equation, for each unit of change in value of each predictor variable, the value of *Y* will change *b* unit when other predictor variables are set unchanged. Multiple linear regression equation would provide evidence related to predictors that could be used to predict the dependent variable.

Methodology
    This section describes in detail how the study was conducted.

Respondents and Data Collection
    The data were collected using the questionnaires distributed to the employees at eight government agencies and five private sectors in the Kuala Terenganu Municipality Council Malaysia. All participants were selected among clerical workers and junior administrative officers. The questionnaire was developed based on literature about intention to further study and its associated factors. Two hundred questionnaires were distributed to employees. However, only a sample of hundred and twenty employees were completed the questionnaire. The number of samples is adequate for a preliminary study where several assumptions of statistical analyses such as linearity assumptions will be tested. The first part of the questionnaire was meant to collect the demographic characteristics of the respondents, including their highest academic qualifications. The description of the participants is presented in Table 1.

Table 1. Demographic characteristics of the sample data

Demographic characteristics	n (%)
Gender	
Male	63 (52.5)
Female	57 (47.5)
Highest Academic Qualification	
Diploma/STPM/Polytechnic Certificate	75 (62.5)
SPM and lower qualifications	45 (37.5)
Terms of employment	
Permanent	113 (94.2)
Contract basis	7 (5.8)

    Almost as many females (52.5 percent, n=63) as males participated (47.5 per cent, n=57) in this study. Regarding to higher academic qualifications, 62.5 per cent (n=75) possessed a diploma or equivalent qualification in various fields of specializations. The rest completed their secondary education with SPM level or lower. The majority of the participants were permanent employees where they were employed until their retirement ages. However, a small percentage of participants worked under contract where they were renewed on a yearly basis.
The second part of the questionnaire composes of several valid and reliable items derived from past research. All items were scored on a five-point Likert scale and its reliability coefficient or internal consistency was 0.608.

Variables
    Predictors in this study are chosen based on general characteristics of learners or specifically in this case study is employees. Any learning individual can be characterised according to their socio-demographic (see Baert et al., 2006 ; Maurer et al., 2003; Sanders et al. 2011b), psychological (see Hazelzet, et al.,) living situation (Kyndt,et al., 2011c) and characteristics regarding learning and education (Baert, 2006). Potential predictors used in this study are identified and selected based on the literature. Ten predictors that employed in this study are gender, marital status, types of a residential area, number of children, salaries, sector, motivation, mode of study, analytical skills and interpersonal skills. These predictors were investigated to observe its effect towards dependent variables, the intention to undertake further study.
    The variables together with the collected data were analyzed using a statistical software.

Data Analysing
    Basically, there are three procedures that can be used in multiple linear regression analysis. The procedures are backward solution procedure, forward solution procedure and step-wise solution procedure. The procedure or method used in this paper is step-wise regression procedure. Step-wise solution procedure is a variation of forward solution procedure. The selection procedure of predictor variable is the same as forward procedure except, after each selection of predictor variable into regression, significant test will be carried out in order to determine the contribution of each predictor variable. Similarly, the next steps are executed where previously inserted variables are accounted into the regression model. The selection procedure is iterated until all variables are inserted into the model. Step-wise regression includes regression models in which the choice of predictive variables is carried out by an automatic procedure (Efroymson, 1960); Draper, and Smith, 1981). According to Diekhoff (1992), step-wise multiple linear regression has advantages compared to other multiple linear regression thanks to its efficiency in inserting predictors and less computational risks. Based on the procedure, only significant predictor variables are inserted into the regression. The second advantage is this step-wise procedure can prevent multi-collinearity problem. Multi-collinearity occurs when there

exists a strong correlation between the predictor variables. This correlation is meaningless and it causes the analysis to become less accurate. These problems can be overcome by using step-wise multiple linear regression because the correlated variables would be excluded into the regression.

Results and Discussion
    The analyses and results are presented according to the following steps.
Step 1: Identify predictors and dependent variable.
Predictors: Gender, marital status, type of residential area, number of children, salaries, employment sector, motivation, method, analytical skills, interpersonal skills.
Dependent variable: Intention.
Step 2: Excluded predictors
All predictors are subjected to be entered and removed from the next analysis. Table 2 shows the excluded predictors variable which later did not include into the model.

Table 2. Excluded Predictors

Multiple Linear Regression Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
	type of residential area	-0.059	-1.131	0.261	-0.116	0.423
	mode of study	0.052	0.765	0.446	0.079	0.252
Predictors	analytical skills	-0.033	-0.947	0.346	-0.097	0.980
	interpersonal skill	-0.030	-0.865	0.389	-0.089	0.979
	marital status	0.139	1.514	0.133	0.154	0.138
	gender	0.034	0.953	0.343	0.098	0.916

    It can be seen that six predictor variables are not significant at $p \leq 0.05$. Moreover, the value of beta in (estimation of beta value when it is inserted into multiple linear regression model) is too small and negligible, so those predictors are excluded from the model by step-wise procedure.
Step 3: Entered predictors
    The remaining predictors are the predictors that significant to the dependent variable when the value of $p \leq 0.05$. Table 3 shows the remaining variables that are significant to the multiple linear regression models.

Table 3. Predictors Entered

Step-wise Model	Predictors Entered	Method
1	motivation level	Step-wise (Criteria: Probability-of-*F*-to-enter $\leq$ 0.050).
2	monthly salaries	Step-wise (Criteria: Probability-of-*F*-to-enter $\leq$ 0.050).
3	occupation sector	Step-wise (Criteria: Probability-of-*F*-to-enter $\leq$ 0.050).
4	number of children	Step-wise (Criteria: Probability-of-*F*-to-enter $\leq$ 0.050).

Step 4: Calculate correlation coefficients (R)
    After establishing the predictors that would be included in the model, then the data of four predictors were analysed further. This analysis is meant to produce results on correlation coefficients for all four predictors. Table 4 shows the correlation between response variable and predictors.
    The combination of all four predictor variables contribute at 88.9 per cents (R = 0.943) variance changes in dependent variable (Intention) to undertake further study. The influential variables are arranged according to the value of R^2. The smallest value of R^2 is the most influenced variable to the dependent variable (Intention). Hence, the value of R^2 in model 1 shows that 81.7 per cents (R = 0.904) changes in dependent variable is due to changes in predictor variables, which is motivation level. This result indicates that motivation level is the main predictor to the intention to undertake further study.
Step 5: ANOVA Test for Regression
    The ANOVA test of significance for regression is performed. It is meant for testing model linearity or fitness of the data toward regression line. The hypotheses are written as

H_0: The variation in dependent variable is not explained by a linear model.
H_a: A significant portion of variation in dependent variable is explained by a linear model.

Table 4 Model summary of correlation coefficients

Step-wise Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	0.904[a]	0.817	0.815	0.214
2	0.935[b]	0.874	0.872	0.178
3	0.940[c]	0.883	0.880	0.173
4	0.943[d]	0.889	0.884	0.169

a. Predictors: (Constant), motivation level
b. Predictors: (Constant), motivation level, monthly salaries
c. Predictors: (Constant), motivation level, monthly salaries, occupation sector
d. Predictors: (Constant), motivation level, monthly salaries, occupation sector, number of children

The results of ANOVA test are summarized in Table 5.

Table 5. ANOVA Test for Regression

Step-wise Model		Sum of Squares	df	Mean Square	F	Sig.
	Regression	21.782	4	5.446	189.663	0.000[e]
4	Residual	2.728	115	0.029		
	Total	24.510	119			

e. Predictors: (Constant), motivation level, monthly salaries, occupation sector, number of children

Table 5 shows the ANOVA test on the predictor variables towards dependent variable. According to the table, all four predictor variables are significant to dependent variable (Intention) to undertake further study. The value of F-distribution can be written as $F(4, 115) = 189.663$, $p \leq 0.05$. Therefore, H_0 is rejected where a significant portion of the variability in dependent variable (Intention) is explained by multiple regression on the predictors.
Step 6: Regression Equation
The step-wise regression analysis is finally analyzed by establishing the regression equation where the relationship between dependent variable and predictors can be interpreted. Table 6 presents regression coefficients between dependent variable and predictors

Table 6. Coefficients of multi linear regression.

Step-wise Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	0.723	0.132		5.473	0.000
	motivation level	0.817	0.101	0.825	8.109	0.000
4	monthly salaries	-0.183	0.028	-0.335	-6.616	0.000
	occupation sector	0.106	0.040	0.093	2.679	0.009
	number of children	-0.072	0.034	-0.193	-2.144	0.035

The significant values ($p \leq 0.05$) show that the model that constructed by dependent variable and predictors can be generalized to a multiple regression equation.
The regression equation can be written as
Intention to undertake further study = 0.723 + 0.817 (motivation level) - 0.183 (monthly salaries) + 0.106 (occupation sector) – 0.072 (number of children).
All four independent variables are predictors to the intention to undertake further study. It can be seen from the equation that motivation level, monthly salaries, occupational sector and number of children are the four variables that can predict the intention to under further study among employees. It is interesting to note that the most influential variable is the motivation level of employees. The result is consistent with most of the past research where they established that students' motivation for learning plays a pivotal role for continuing different levels of further education (see, for example, Gegenfurtner, 2011; Kyndt et al, 2011a; Vansteenkist et al, 2009). Although all the four variables are good predictors,

however, two variables, namely monthly salaries and number of children are predicted negatively toward intention to undertake further study.

Conclusions

Multiple linear regression analysis is a compatible use for determining the most influential variable towards certain cases. The analysis is suitable in human life aspect, especially in economy and social sciences. This paper has shown the capability of the step-wise multiple linear regressions for identifying the most effective predictors that can be used to determine the employees' intention to undertake further study. Hundred and twenty employees were sampled for the case and the analysis was performed using step-wise multiple linear regression analyses. The analyses include identifying entered and excluded variables, calculating correlation coefficient, testing of linearity using the ANOVA test for regression and finally establishing the multiple linear regression equation. It is unveiled that four predictors are statistically significant under probability area less than 0.05. The four predictors are motivation level, monthly salaries, occupation sectors and the number of children. Of the four predictors, motivation level is identified as the most effective predictor. These predictors may help policy makers to predict the likelihood of the intention to undertake further study among employees. However, this predictor is not consistent with the results obtained by Kyndt et al. (2013) due to different definitions of variables used. In their study, self-directedness in career processes and financial satisfaction were positively related to the learning intentions. Perhaps, the differences in the experimental setting and operational definitions could contribute to the inconsistent results. This preliminary investigation has several limitations. Number of employees, related research and scientific research protocols are among the few works that can be improved and could be left for future research.

References

Baert H, De Rick K, Van Valckenborgh K. 2006. Towards the conceptualisation of learning climate, in Vieira de Castro, R., Sancho, A,V, and Guimaraes, P. (Eds), Adult Education: New Routes in a New Landscape, University of Minho, Braga, pp. 87-111.

Chen H, Zimitat C. 2006. Understanding Taiwanese students' decision-making factors regarding Australia international higher education. International Journal of Educational Management, 20(2), 91-100.

David PA, Foray D.2002. An introduction to the economy of the knowledge society, International Social Science Journal, 54(171), 9-23.

Diekhoff GM.1982. Statistics for the social and behaviour sciences. Dubuque: Wm. C. Brown Publishers,

Draper N, Smith H.1981. Applied Regression Analysis, 2d Edition, New York: John Wiley & Sons, Inc.

Efroymson MA. 1960. Multiple regression analysis, Mathematical Methods for Digital Computers, Ralston A. and Wilf,H. S., (eds.), Wiley, New York.

Freedman DA. 2009. Statistical Models: Theory and Practice. Cambridge University Press.

Gegenfurtner A.2011. Motivation and transfer in professional training: A meta-analysis of the moderating effects of knowledge type, instruction, and assessment conditions, Educational Research Review, 6, 153–168.

Gvaramadze I. 2010. Low skilled workers and adult vocational skills upgrading strategies in Denmark and South Korea, Journal of Vocational Education and Training, 62(1), 51-61.

Hazelzet A, Oomens S, Keijzer L.2009. Wat prikkelt laagopgeleide werknemers voor scholing (What stimulates low qualified employees towards education)?, Develop: Kwartaaltijdschrift over Human Resources Development, 5(2), 53-60.

Illeris K. 2006. Lifelong learning and the low skilled, International Journal of Lifelong Education, 25(1), 15-28.

Kyndt E, Dochy F, Cascallar E, Struyven K.2011a. The direct and indirect effect of motivation for learning on students' approaches to learning through the perceptions of workload and task complexity, Higher Education Research & Development, 30, 135–150.

Kyndt E, Govaerts N, Dochy F, Baert H.2011b. The learning intention of low qualified employees: a key for participation in lifelong learning and continuous training, Vocations and Learning, 4(3), 211-229.

Kyndt E, Govaerts N, Keunen L, Dochy F.2013. Examining the learning intentions of low qualified employees: a mixed method study, Journal of Workplace Learning, 25(3), 178 – 197.

Kyndt E, Michielsen M, Van Nooten L, Nijs S, Baert H.2011c. Learning in the second half of the career