

Logistic Regression Model for the Academic Performance of First-Year Medical Students in the Biomedical Area

María Esther Urrutia-Aguilar^{1*}, Ruth Fuentes-García², Vargas Danae Mirel Martínez², Edward Beck³, Silvia Ortiz León¹, Rosalinda Guevara-Guzmán¹

¹Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico

²Faculty of Sciences, National Autonomous University of Mexico, Mexico City, Mexico

³University of Illinois College of Medicine, Chicago, IL, USA

Email: *mariau@unam.mx

How to cite this paper: Urrutia-Aguilar, M. E., Fuentes-García, R., Martínez, V. D. M., Beck, E., León, S. O., & Guevara-Guzmán, R. (2016). Logistic Regression Model for the Academic Performance of First-Year Medical Students in the Biomedical Area. *Creative Education*, 7, 2202-2211.

<http://dx.doi.org/10.4236/ce.2016.715217>

Received: August 13, 2016

Accepted: September 17, 2016

Published: September 20, 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In the medical education field, the prediction of variables that have an impact on the academic performance of students is highly important as supporting programs can be implemented to avoid dropouts or failing scores. Several studies have confirmed the relationship between student performance during the first months at college and the one afterwards; nevertheless, every medical school has its particularities. The objective is to develop a logistic regression model to predict first-year medical students' performance using academic, psychological and vocational variables as well as learning and strategies for self-motivation. The study is observational, transversal and descriptive. The study group consisted of 1205 first-year medical students. Participants completed questionnaires dealing with general knowledge, psychosocial factors, factors associated with career choice, as well as research and autoregulation strategies inventory. Participation was fully voluntary and the results were used under confidential agreement (NDA). The multiple regression model considered pleasure in academic background, percentage of checkmarks in general knowledge, perceived efficiency (categorical with 3 levels), aptitude, interest in biological sciences and health, follow established regulations, drive and the pursue of social prestige as covariables. We conclude that a logistic regression model to predict academic performance, mostly of those medical students under academic risk in the biomedical area, is an efficient tool as it allows valid conclusions for appropriate decision making.

Keywords

Academic Performance, Medical Students, Logistic Regression, Biomedical Area

1. Introduction

The logistic regression model aims to describe the relationship between one or more independent variables that may be continuous, categorical, or binary (Hosmer, Stanley, & Rodney, 2000).

Due to its contribution, the logistic regression model has become a useful tool in business, medicine, epidemiology, sociology and marketing, and so on as it can predict the ailment disease, the granting of a credit to a specific individual, the success or failure of a business, among many others.

In the field of medical education, the prediction of variables that have an impact on the academic performance of students is highly important as supporting programs can be implemented to avoid dropouts or failing scores. A study carried out in the United Kingdom found a 10% rate of dropout in the first-year medical students. When different cohorts were analyzed (Arulampalam, Naylor, & Smith, 2004), they observed an association between the risk of academic failure and the level of knowledge in biology, chemistry and physics, as well as their gender (Oliver, Smith, Winston, Geranmayeh, Behjati, Kingston, & Pollara, 2010).

Research on prediction of academic success in medical students has generated a database of substantial knowledge on the factors predicting school performance (Ferguson, James, & Madeley, 2002; O'Neill, Wallstedt, Eika, & Hartvigsen, 2011). This allows medical schools to offer prompt and tailored support for students at risk of academic failure.

Several studies have confirmed the relationship between student performance during the first months at college and the ones afterwards (Horn & Carroll, 1998; Murtaugh, Burns, & Schuster, 1999; Winston, van der Vleuten, & Scherpbier, 2014); nevertheless, every medical school has its particularities and in the Faculty of Medicine at UNAM, there is a high index of failure in the biomedical area during the first year of studies and thus, an increased number of repeaters in the following academic year.

In a previous study, we showed that emotional variables had the highest correlation with academic performance of the first-year medical students (Urrutia-Aguilar, Ortiz, Fouilloux, Ponce, & Guevara-Guzmán, 2014). However, we consider that this type of educational research should include other types of variables.

To meet this need, in this study we use a logistic regression model to predict medical students' performance in their first year using academic, psychological and vocational variables as well as learning and motivational strategies as variables.

2. Methodology

The population under study included 1205 first-year medical students at Universidad Nacional Autónoma de México (UNAM), 2013-2014. The academic curriculum consists of a six-year training program, with the first two years being considered biomedical studies, the following three year clinical clerkships and the last year social service.

The regressive variables considered in this research work were categorized into four main groups: 1) general knowledge; 2) psychosocial factors; 3) factors associated with

career choice and 4) inventory of research and strategies for self-motivation. The outcome variable for our model was academic performance and it was coded as “0” if the student had failed at least one subject of the biomedical area taken during the first year and “1” in case of successfully completing all the coursework.

The study was observational, cross-sectional, and descriptive and was approved by the Ethics and Research Commissions of the Faculty of Medicine at UNAM (reference number 061/2014).

The type of academic background was considered and classified into Colegio de Ciencias y Humanidades (CCH) and Escuela Nacional Preparatoria (ENP), these two being part of UNAM and any other school system was named as “Other”, including foreign schools.

The scores obtained in the diagnostic exam (designed and validated by the General Directorate of Orientation and Educational Services, UNAM) were gathered. This is a 120 item exam related to general knowledge in the following areas: Physics, Chemistry, Math, Biology, Spanish, Universal History and History of Mexico as well as Geography; all of multiple choice answers.

The instrument of career choice factors (designed and validated by the General Directorate of Orientation and Educational Services, UNAM) was applied. Its objective was to assess 44 influential factors within aptitude, interest and expectations regarding career choice.

In order to assess psychological characteristics, including signs of depression, two different questionnaires were applied in the same place and time for all students after signing an informed consent format: 1) Beck Depression Inventory (Beck, Steer, & Garbin, 1988; Jurado, Villegas, Mendez, Rodriguez et al., 1998); and, 2) Symptom Checklist SCL-90. The Beck Depression Inventory can be self-administered, and has 21 items to assess depression symptoms in adolescents and adults. The total score varies from 0 to 63, with a score ≥ 13 being identified as a probable case of depression. The Symptom Checklist SCL-90 is a screening test to identify symptoms of different psychopathologies, and has 90 items with a Cronbach Alpha of all subscales over 0.713 (Cruz, López, Blas, González et al. 2005). The subscale for depression considered for this study was greater than or equal to 1.514.

Study strategies and self-motivation were categorized using a 91-item multiple choice questionnaire that assessed how students acquire, organize, remember and apply what they learn. The questionnaire also assessed the way they evaluate, plan and control their study strategies with a Cronbach Alpha of 0.97 (Castañeda, Pineda, Gutiérrez, Romero, & Peñalosa, 2010). The student self-assessed their study strategies and self-motivation in four main aspects: 1) data gathering; 2) methods of recovering learned data when facing tasks and exams; 3) methods of processing, which may be convergent or divergent; and, 4) meta cognition and meta motivational self-regulation.

To assess student performance, scores were gathered during the academic year from August 2014 to May 2015 in the following subjects: Cell Biology and Medical Histology, Biochemistry and Molecular Biology and Human Embryology. It is important to high-

light that the departmental exams were applied at the same time to all populations and each contained 50 to 70 items chosen by a group of experts from an items bank (Cronbach ≥ 0.87 Y 0.92 ; level of difficulty = 30 y 70 ; positive discrimination = 70 y 90). The assessment was carried out through departmental exams (50%) and the professor's criterion 50 . Considering that the student has a passing score of 8.5 .

Data was analyzed using the R statistical software. The statistical analysis was descriptive, inferential and multivariate.

Inclusive criteria: In this study data of students who had taken all the surveys and had one grade at least were included. The participation was fully voluntary and the results were used under confidentiality agreement (NDA).

3. Results

The population under study that met the inclusive criteria were 925 students, 37% men and 63% women; 34.7% of the students came from CCH, 52% from ENP and 13.3% from "other". The database contained 72 variables.

Figure 1 shows the CCH students associate failing scores in a positive way while students from "other" show a negative association towards failure.

Two different models with their corresponding co-variables were developed. One included all the variables (i.e. academic background, percentage of check marks in general knowledge, perceived efficiency, aptitude, interest in biological sciences and health, follow established regulations, drive and the pursue of social prestige as covariables) and the second measuring each variable independently.

High-school is the type of academic background at entry.

Perceived efficiency is the factor that considers conviction of oneself that a task can be carried out successfully, the behavior needed to produce results.

The aptitude test is described as the ability the student has towards reasoning (data

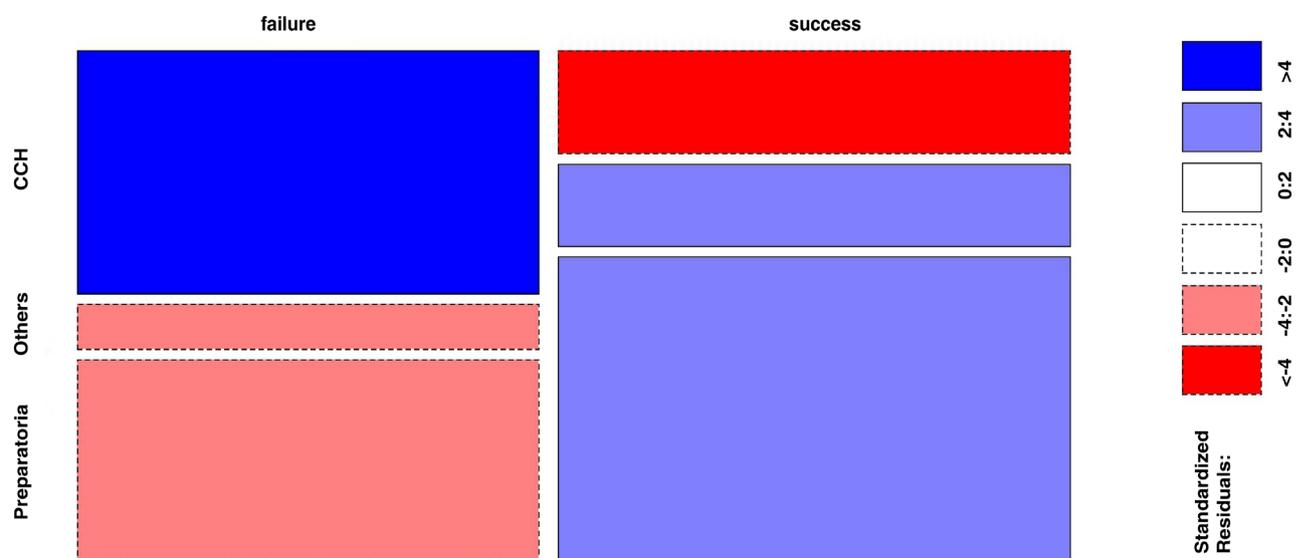


Figure 1. Academic success and bachelor of origin.

analysis and synthesis), mechanical aptitude (identification of physics principles) and assembling of ways (spatial perception).

The motivational factor is understood as having an orientation towards achievement that sets high goals and strive to reach them and as a skill, the interest to achieve a good academic performance.

The social prestige factor assesses the behaviors that the subject show to feel accepted by the referred social group, the sense of belonging.

As our two models are co-related (model one has all covariables of model two), we compared them through the difference of deviances. Model one was the best predictor of academic failure or success as it was statistically significant ($p \leq 0.02$), and thus, we concluded that the model with the most parameters, model one, contains relevant data that is not contained in model two.

Figures 2-6 show the probability of success according to different values of continuous variables, in the model. It can be observed that the probability of success is directly proportional to the score obtained in general knowledge, with the aptitude test, with the interest of the student in biological sciences and health, as well as with the motivational

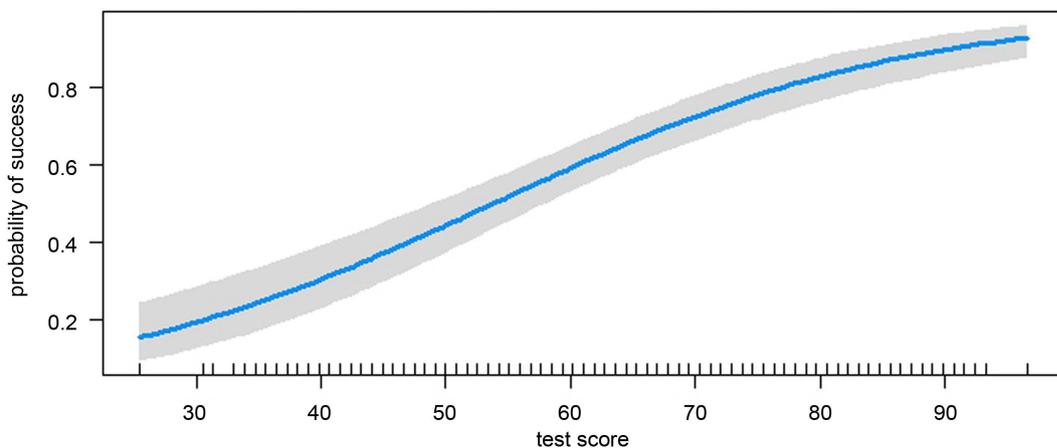


Figure 2. Probability of success explained by the knowledge test.

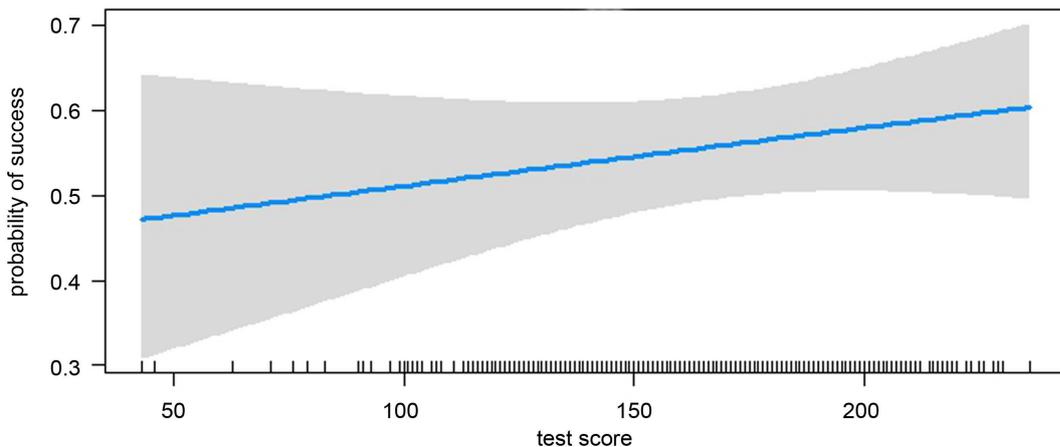


Figure 3. Probability of success explained by the skills test.

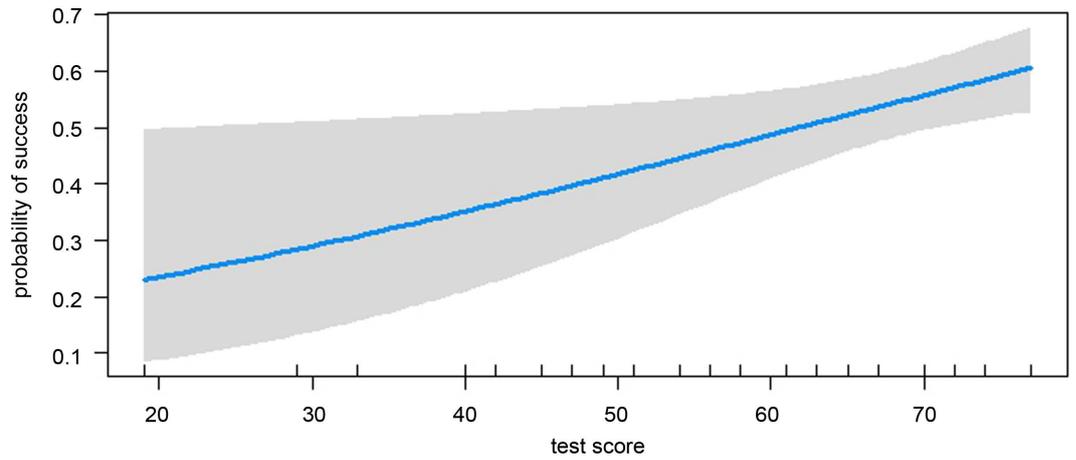


Figure 4. Probability of success explained by the factor biological sciences and health.

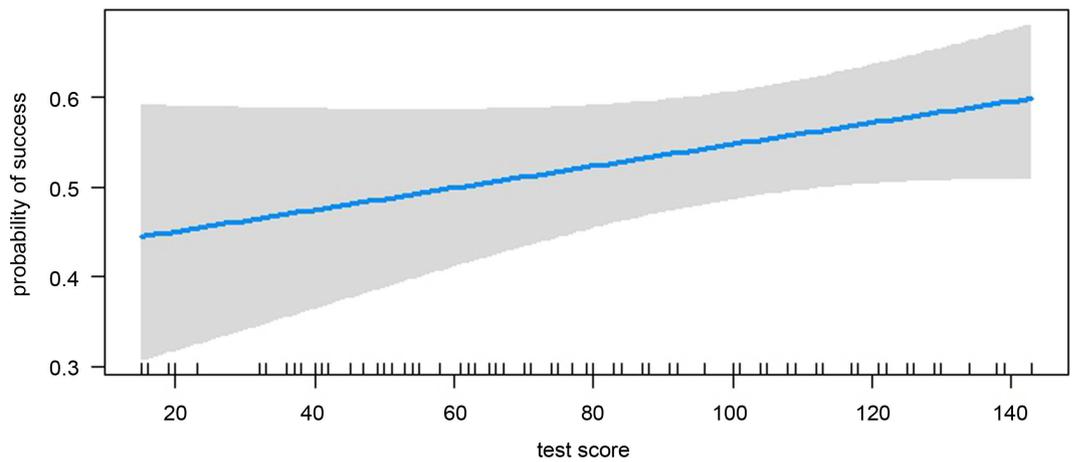


Figure 5. Probability of success explained by motivators.

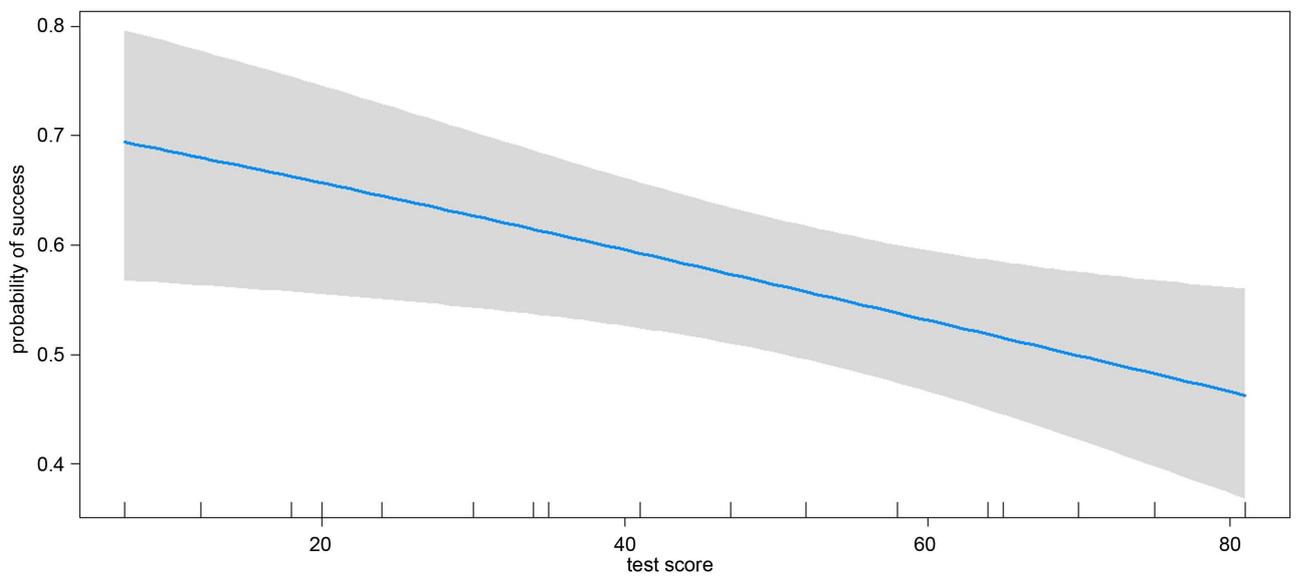


Figure 6. Probability of success explained by the social prestige factor.

factors, which is inversely proportional with the social prestige factor.

Abscissa indicates test score and probability success explained by the knowledge test, skills test, the factor sciences and health and motivators aspect.

We got a $p \leq 0.17$ in the Hosmer Lemeshow test, which shows a good fit of the model to the data.

Regarding the predicting power of the model, if the response variable is “1”, it means that the student received a passing average in the subjects considered in this study and zero if he/she failed one subject. In the best interest of this context, it is of great value to have a higher specificity (zeros well classified as zeros) rather than a higher sensitivity (ones well classified as ones), as our main objective is to predict the population under academic risk (**Table 1**).

$$\text{Sensitivity} = \frac{355}{355+132} = 0.73$$

$$\text{Specificity} = \frac{299}{299+139} = 0.68$$

$$\text{Accuracy} = \frac{355 + 299}{355+132+139+299} = 0.71$$

Considering $\pi_0 = 0.5$ as cutoff point, sensitivity was 0.73 and specificity of 0.68, accuracy 0.71. These values show that the predicting power of the model is considerably good.

When the ROC Curve was created (**Figure 7**) we found that the area under the curve was 0.7736 and thus considered adequate.

4. Discussion

The use of a logistic regression model to predict academic performance among first year medical students, in particular identifying those at academic risk, is an efficient tool as it allows valid conclusions for appropriate decision making.

Literature supports theoretical evidence that link academic performance during first-year of medical school as multifactorial (O’Neill, Wallstedt, Eika, & Hartvigsen, 2011; Ward, Kamien, & Lopez, 2004; Strayhorn, 1999). The fact that academic background has had significant weight in the model can be explained in terms of its self-motivational level. Students whose admission process was through exams have a higher motivation compared to those admitted through another method, as is the case of admission through automatic transfer (Stegers-Jager, Themmen, Cohen-Schotanus, & Steyerberg, 2015; Hulsman, van der Ende, Oorst, Michels, Casteelen, & Griffioen, 2007; Kusurkar, Kruitwagen, ten Cate, & Croiset, 2010).

It was not surprising that the variables with a higher weight of prediction had been

Table 1. Classifying chart.

	$\hat{Y} = 1$	$\hat{Y} = 0$	Total
$Y = 1$	355	132	487
$Y = 0$	139	299	438

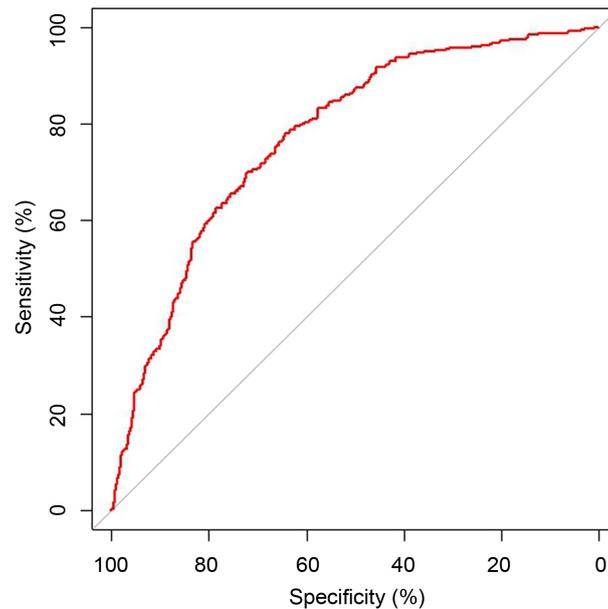


Figure 7. ROC curve.

related to the level of previous knowledge. In former studies, this has been critical for the academic performance, since students with a stronger academic background establish better knowledge interrelations and have more proficient development in clinical clerkships (Ferguson, James, & Madeley, 2002; O'Neill, Wallstedt, Eika, & Hartvigsen, 2011; Winston, van der Vleuten, & Scherpbier, 2014; Mills, Heyworth, Rosenwax, Carr, & Rosenberg, 2009; Yates & James, 2007).

In this study, we showed that student self-perception of his/her own ability to perform the required tasks of a course (perceived efficiency) is a factor that predicts academic performance, which undoubtedly is a cognitive variable that determines learning in the classroom (Castañeda, Pineda, Gutiérrez, Romero, & Peñalosa, 2010). The term of perceived efficiency is linked to the Social Cognitive Theory (Bandura, 1986), which establishes that learning is produced through a reciprocal interaction between environmental, behavioral and cognitive influences.

The self-perception students have of their own abilities influences the choice in tasks, proposed goals, effort, and persistence of actions towards reaching such goals. Therefore, professors must consider the importance of reassuring the student self-esteem when learning new concepts. Self-perception plays a significant role in the academic performance of medical students demonstrated through implementation of one-term geared towards developing their meta cognitive and self-regulation abilities (Winston, van der Vleuten, & Scherpbier, 2014; Winston, Van der Vleuten, & Scherpbier, 2010).

5. Conclusions

The logistic regression model in this study identified variables that predicted academic performance among the first-year medical students. These findings have practical implications, in both the development of in-person or online/distance preparatory courses

designed for high school students matriculating into medical school and the implementation of corrective measures during the first year of medical school. These are based on the need to develop didactic materials that will help correct academic deficiencies. Additionally, teaching training courses could be adopted to deal with topics on teaching strategies to aid students to improve their self-perception.

Nonetheless, we are confident that the model identified in this study can be used across all national medical schools to successfully identify students at risk of academic failure and provide efficient strategies to ensure that students complete their medical training.

A limitation of this study is that departmental exams are mainly structured to assess declarative contents. In the future, research work should undertake validation of logistic regression models.

Acknowledgements

The authors thank Josefina Bolado for the edition and translation of this text into English language.

Competing Interests

The authors declare that they have no competing interests.

References

- Arulampalam, W., Naylor, R. A., & Smith, J. (2004). A Hazard Model of the Probability of Medical School Dropout in the United Kingdom. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *167*, 157-178. <http://dx.doi.org/10.1046/j.0964-1998.2003.00717.x>
- Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Beck, A. T., Steer, R. A., & Garbin, M. C. (1988). Psychometric Properties of the Beck Depression Inventory. Twenty-Five Years of Evaluation. *Clinical Psychology Review*, *8*, 77-100. [http://dx.doi.org/10.1016/0272-7358\(88\)90050-5](http://dx.doi.org/10.1016/0272-7358(88)90050-5)
- Castañeda, F. S., Pineda, G. M. L., Gutiérrez, M. E., Romero, S. N., & Peñalosa, C. E. (2010). Learning and Self-Regulation Strategies and Personal Epistemology Instrument Construction. Construct Validation. *Revista Mexicana de Psicología*, *27*, 77-85.
- Cruz, F. C. S., López, B. L., Blas, G. C., González, M. L. et al. (2005). Data about Validity and Reliability of the Symptom Check List 90 (SCL 90) in a Mexican Population Sample. *Salud Mental*, *28*, 72-81.
- Ferguson, E., James, D., & Madeley, L. (2002). Factors Associated with Success in Medical School: Systematic Review of the Literature. *BMJ*, *324*, 952-957. <http://dx.doi.org/10.1136/bmj.324.7343.952>
- Horn, J. L., & Carroll, C. D. (1998). *Stopouts or Stayouts? Undergraduates Who Leave College in Their First Year*. Washington DC: US Department of Education, National Center for Education Statistics.
- Hosmer, W. D., Stanley, L., & Rodney, X. S. (2000). *Applied Logistic Regression*. New York, NY: Wiley. <http://dx.doi.org/10.1002/0471722146>
- Hulsman, R. L., van der Ende, J. S. J., Oorst, F. J., Michels, R. P. J., Casteelen, G., & Griffioen, F.

- M. M. (2007). Effectiveness of Selection in Medical School Admissions: Evaluation of the Outcomes among Freshman. *Medical Education*, *41*, 369-377. <http://dx.doi.org/10.1111/j.1365-2929.2007.02708.x>
- Jurado, S., Villegas, M. E., Mendez, L., Rodriguez, F. et al. (1998). The Standardization of the Beck Depression Inventory for Residents of Mexico City. *Salud Mental*, *21*, 26-31.
- Kusurkar, R., Kruitwagen, C., ten Cate, O., & Croiset, G. (2010). Effects of Age, Gender and Educational Background on Strength of Motivation for Medical School. *Advances in Health Sciences Education*, *15*, 303-313. <http://dx.doi.org/10.1007/s10459-009-9198-7>
- Mills, C., Heyworth, J., Rosenwax, L., Carr, S., & Rosenberg, M. (2009). Factors Associated with the Academic Success of First Year Health Science Students. *Advances in Health Sciences Education*, *14*, 205-217. <http://dx.doi.org/10.1007/s10459-008-9103-9>
- Murtaugh, P. A., Burns, L. D., & Schuster, J. (1999). Predicting the Retention of University Students. *Research in Higher Education*, *40*, 355-371. <http://dx.doi.org/10.1023/A:1018755201899>
- Oliver, T. A. L., Smith, C., Winston, S. J., Geranmayeh, F., Behjati, S., Kingston, O., & Pollara, G. (2010). Impact of UK Academic Foundation Programmes on Aspirations to Pursue a Career in Academia. *Medical Education*, *44*, 996-1005. <http://dx.doi.org/10.1111/j.1365-2923.2010.03787.x>
- O'Neill, L. D., Wallstedt, B., Eika, B., & Hartvigsen, J. (2011). Factors Associated with Dropout in Medical Education: A Literature Review. *Medical Education*, *45*, 440-454. <http://dx.doi.org/10.1111/j.1365-2923.2010.03898.x>
- Stegers-Jager, K. M., Themmen, A. P. N., Cohen-Schotanus, J., & Steyerberg, E. W. (2015). Predicting Performance: Relative Importance of Students' Background and Past Performance. *Medical Education*, *49*, 933-945. <http://dx.doi.org/10.1111/medu.12779>
- Strayhorn, G. (1999). Participation in a Premedical Summer Programmer for Under-Represented Minority Students as a Predictor of Academic Performance in the First Three Years of Medical School: Two Studies. *Academic Medicine*, *74*, 435-447. <http://dx.doi.org/10.1097/00001888-199904000-00043>
- Urrutia-Aguilar, M. E., Ortiz, L. S., Fouilloux, M. C., Ponce, R. E. R., & Guevara-Guzmán, R. (2014). Academic Performance in First Year Medical Students: An Explanatory Multivariate Model. *Gaceta Médica de México*, *150*, 324-330.
- Ward, A. M., Kamien, M., & Lopez, D. G. (2004). Medical Career Choice and Practice Location: Early Factors Predicting Course Completion, Career Choice and Practice Location. *Medical Education*, *38*, 239-248. <http://dx.doi.org/10.1046/j.1365-2923.2004.01762.x>
- Winston, K. A., van der Vleuten, C. P., & Scherpbier, A. J. (2014). Prediction and Prevention of Failure: An Early Intervention to Assist At-Risk Medical Students. *Medical Teacher*, *36*, 25-31. <http://dx.doi.org/10.3109/0142159X.2013.836270>
- Winston, K. A., Van der Vleuten, C. P. M., & Scherpbier, A. J. J. A. (2010). An Investigation into the Design and Effectiveness of a Mandatory Cognitive Skills Programme for At-Risk Medical Students. *Medical Teacher*, *32*, 236-243. <http://dx.doi.org/10.3109/01421590903197035>
- Yates, J., & James, D. (2007). Risk Factors for Poor Performance on the Undergraduate Medical Course: Cohort Study at Nottingham University. *Medical Education*, *41*, 65-73. <http://dx.doi.org/10.1111/j.1365-2929.2006.02648.x>



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ce@scirp.org