A STRUCTURAL EQUATION MODEL FOR ANALYSIS OF FACTORS ASSOCIATED WITH THE CHOICE OF ENGINEERING DEGREES IN A TECHNICAL UNIVERSITY

Antonio Hervás, Joan Guàrdia Olmos, Maribel Peró Cebollero, Roberto Capilla Lladró, Pedro Pablo Soriano Jiménez
Este trabajo se encuentra en fase de desarrollo. Si desea recibir una copia actualizada por favor diríjase al autor de contacto. Rogamos disculpen las molestias.

A STRUCTURAL EQUATION MODEL FOR ANALYSIS OF FACTORS ASSOCIATED WITH THE CHOICE OF ENGINEERING DEGREES IN A TECHNICAL UNIVERSITY.

Antonio Hervás[1,4] Joan Guàrdia Olmos[2,3],Maribel Peró Cebollero[2,3],Roberto Capilla Lladró[1],Pedro Pablo Soriano Jiménez[1]

[1] Universitat Politècnica de València, [2] Facultat de Psicologia. Universitat de Barcelona, [3] Institut de Recerca en Cervell, Cognició i Conducta (IR3C), [4]Intituto Universitario de Matemática Multidisciplinar (IMM)

Autor de contacto:

Antonio Hervàs

Departamento de Matemàtica Aplicada

Camino de Vera, s/n

46022 Valencia

Teléfono: +34 963877660

Fax: +34 963877669

Email: ahervas@imm.upv.es

Abstract:

Many different factors are taken into account by students when choosing a degree and university. Some of these are general considerations, such as the quality of the degree course (ratio available places/places in first choice, cut-off mark, etc.) while others are subjective factors (e.g.: friends doing the same course). This paper presents a partial multivariate model that considers the weight of the different variables linked to this decision, as identified in the bibliography. We analyzed four samples of first-year students (total n = 1790) from different engineering degree courses at the Universitat Politecnica de Valencia (UPV) in the 2010-2011-2012 academic years. All the students involved in the study had chosen this university and their courses as their first option. The overall effect shows that the structural model adjusts reasonably well to the different engineering courses analyzed. Similarly, the individual models for each engineering degree manage to identify the different effects involved. In the case of the engineering degree based on new technologies (ICT) the statistical effects are much greater and more statistically significant than in the other three branches of engineering considered. Social and individual factors were seen to have more impact on the choice of degrees at the UPV.

Key words:

Technological College Selection, Structural Equation Models, University Access, Higher Education Management.

Introduction

The question of which factors determine a student's selection of university and degree course has been the subject of debate for some time. Many specialists have opined on the subject with more or less structured inputs and empirical contributions to justify their positions.

It is easy to obtain questionable but often-repeated evidence on the criteria used in selecting a university and study course. Arguments such as "my friends are also going", "It's close to home" or the well-known "the cut-off is low" are often cited. Studies such as those published by the Universidad Antonio de Nebrija (2001) (www.nebrija.com/servicios/publicaciones.php) or the Universitat Oberta de Catalunya (2001) (http://elcrps.uoc.edu/ojs/index.php/elcrps) show how most students use this type of argument at the expense of those that academic specialists would prefer them to use.

This issue has also been raised in Europe, where the universities (and even more so the universities of our immediate environment) have found that new students are highly motivated but are somewhat lacking in the necessary capacities. The European Access Network (EAN) (www.ean-edu.org) for example has been very active in this field. In the USA the tradition is very similar with the "slight" difference that recruitment systems are supposedly competitive, but are ultimately based on the financial resources of prospective students, very different from our policy of scholarships and student grants. In our system grants play a social role, while in the USA there is a policy of recruiting talent, as has been shown by Murphy and McGarrity [1] in a descriptive study of 350 American colleges. Their conclusions have been verified by other authors, who also found this vision of college selection as the key to the future of students. Supporting evidence for this position can be found in the Hispanic Association of Colleges and Universities (www.hacu.net), whose aim is to incorporate Hispanic minorities in the American university system, based on the quality of the system and the 'trust' in the university. The situation described by Murphy and McGarrity is similar to the present situation in Spain and, as we have already said, can be clearly seen in some proven schemes, as in the case of Engineering [2] or that of women in certain degrees [3] or in exclusion-prone groups such as ethnic minorities [4].

Most authors agree on the relationship of the factors that determine the selection of a college or university and these are clearly identified in some of the works cited, as in [4]. In fact, they are consistent with many of those proposed some time ago in various forums and have even been included in the general documents of the European Students Union (www.esib.org) to establish unsystematic models, somewhat more phenomenological than demonstrable. Many of the proposals are similar in type. First, it is considered that the demand for a degree and a university has a direct relationship with a factor generated from the subjective perception of this degree by students and their environment. The ratio between places and the demand for first choice is established from indirect indicators, as shown by the consolidation of the faculties and degrees most in demand. In more specific models, a certain longitudinal effect is admitted, so that the ratio between supply and demand in previous years is taken into account in the subjective perception of the present situation.

The Spanish public university admittance system is a relatively simple one. Prospective students range from those who have obtained the highest to those with the lowest acceptable grades. The condition is that there is a minimum requirement to access the system, so that only those who have successfully passed the pre-university (high school or Baccalaureate) stage are eligible for consideration in the place allocation process. Those who have successfully passed high school and also a common university entrance exam, known as the *Prueba de Acceso a la Universidad* (PAU), can apply for a place at any university in the public system. The access grades of students considered for admittance consist of a weighted linear combination of the average grade of all high school subjects and the average marks of subjects taken in the PAU.

The system of access grades means that all students seeking a place at a public university can be ranked from the highest rating to the lowest and that places can be systematically assigned according to the preferences of the students. This ensures that no student gets a place in the public system with a lower

entrance grade than another student who would not have been admitted to the same degree course. It sometimes happens that when the time comes to assign a place to a student and due to his/her access grade there are no vacancies for the specified course or university. In such cases students they are assigned their second option, or third if the second is not feasible due to lack of capacity. However, recent data from the Spanish universities admittance system shows that more than 92% are assigned their first choice and 98% one of their first three options. Finally, the access grade of the last student admitted to a given degree course establishes what is known as the *cut-off mark*, i.e. the qualification below which a place cannot be obtained in the public system for that degree and university.

Obviously, the number of places available for the different degrees and universities and the applications received are the mechanisms that define the cut-off mark, so that a high demand coupled with limited places carries high cut-off marks, and likewise the degrees widely on offer with a low demand imply low cut-off marks. In no case does the level of the cut-off mark reflect the perceived difficulty of the degree or any other circumstance that is not strictly the relationship between supply and demand. Similarly, the number of first choice applications for a certain degree and university is a good indicator of these universities' capacity to attract students.

Other studies, such as Capilla, [5], Raymond, [6], Huang and Fang, [7] and Veenstra, Dey and Herrin, [8], also see the demand for a university course as a very non-specific assessment of the "social value" given to the degree and also to the university. This more qualitative or more subjective perception is constructed from certain parameters such as the absolute value of the cut-off mark in the application process, since it is used socially as an indicator of quality in sought-after degree courses, but not in degrees with a low demand and cut-off mark. No one doubts the difficulty of a college course in mathematics, but few consider its cut-off mark. Only when the available places are scarce is the cut-off mark used as an indicator of quality. The relationship between both factors is clear and some authors, like the aforementioned Ford, [4], attribute most of the variance to it, focusing on aspects such as the perceived utility of the degree. Such a proposal coincides with some contextual data. For example, the utility data offered by the Quality Agency of the University System in Catalonia (www.aqu.cat) indicate that most students and parents attribute utility to many degrees that are not in fact in great demand by employers (such as computer science, for example). These subjective perceptions, in the opinion of these authors and also in ours, are decisive when selecting a university and course of studies. Similarly, some unsystematic data that has appeared in forums and seminars points to the existence of personal factors in this process, such as geographical issues, transport, gender, vocation and interest in the subject, etc. These personal aspects, as could be expected, sometimes coincide with aspects of the student's pre-university academic history.

All these factors have certain similarities and links, occasionally spurious, which must be carefully analyzed, and which give rise to a complex network of effects which should be able to explain the variability observed in the university and degree course selection process. If we know the weight of each factor, its effect, intensity and direction, we can plan much more effectively university access guidance campaigns, generate strategies for the advancement of under-represented groups and, finally, make the transition between pre-university and university much smoother. There are many who consider that the inadequate transition and lack of proper planning are largely responsible for drop-outs, absenteeism and academic failure, however, this question is outside the scope of the present paper. Finally, we should point out that an analysis of the phenomenon dealt with here only acquires a realistic dimension in longitudinal terms and with a large sample of subjects that not only considers demographic variations due, for example, to immigration, but also takes into account the impact of new degrees in Spain and the perhaps excessively wide range of courses available. In view of the above, the main objective of this work is to verify a multivariate model based on the principles of the structural equations model (SEM), which analyzes the impact of the different variables and factors identified by the bibliography as linked to the decision process on university degree courses, in the Universitat Politecnica de Valencia (UPV) as paradigmatic institution in this environment..

The study was carried out on various cohorts of freshmen students enrolled for the first time in the Spanish university system in the academic years 2010-2011 and 2011 - 2012 and formed part of a larger project which analyzed the decision process in other areas of knowledge.

Path Diagram

The structural relationships used in the analysis were based on the structural model (SEM) proposed by Guàrdia et al in [9], which obtained good fits of the structural model shown in Figure 1 as applied to a psychology degree course. A further development of the model was applied to several other degrees and universities [10].
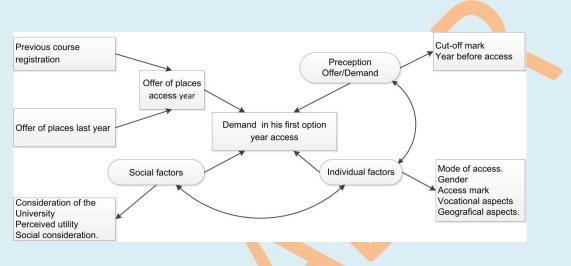

Figure 1: Diagram of the proposed structural model

As can be seen, this model involves the simultaneous use of directly observable variables free of error (represented by rectangles in Figure 1) and latent variables (represented by ovals). This created certain difficulties of notation on translating the proposal into statistical terms specific to structural models, as shown in Figure 2.

Figure 2: Specification of proposed structural model with identification of the free parameters to be estimated.

The structural equations that can be specified from Figure 2 are as follows:

$$Y_1 = \beta_{11}X_1 + \beta_{12}X_2 + \zeta_1$$
$$Y_2 = \beta_{21}Y_1 + \gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \gamma_{23}\xi_3 + \zeta_2.$$

Finally, to conform to the general precepts and assumptions of structural equation models, we considered the following statistical assumptions for quantitative variables $E(X_i) = E(Y_i) = E(\xi_i) = 0$ and $Var(X_i) = Var(Y_i) = Var(\xi_i) = 1$. Consequently, all quantitative variables were transformed by reduction and standardization, and similarly $E(\varepsilon_i\varepsilon_j) = E(\delta_i\delta_j) = E(\xi\delta) = E(\eta\varepsilon) = E(\zeta_i\zeta_j) = 0$;; assuming initially that the errors of measurement were uncorrelated with each other, as in the case of the observable and latent variables. The categorical observable variables (type of baccalaureate and gender) were considered separately and subjected to an own estimation process described below.

For the sake of brevity, the structures of exogenous measurement models (Λ_x) are not included here. We adopted the correlations between exogenous variables (both observable and latent) that had been shown to be significant in previous pilot studies. In all cases, the exogenous measurement models specified in the model comply with the conditions for applying the usual order conditions. In addition, the proposed

model meets the identification condition, since it presents positive degrees of freedom (degree of freedom df = 321).

Method

Participants

Four accidental samples were obtained (n = 2244) composed of undergraduate students doing different degree courses at the Universitat Politécnica de Valencia. The UPV is one of four polytechnic universities in Spain and offers a wide range of courses. Table 1 shows some of its main features

Universitat Politécnica de Valencia (UPV)	Autonomous Community: Valencia	Located in Valencia, on the east coast of Spain, it is highly focused on technology transfer and specialized engineering degrees together with others in the field of social sciences and also fine arts. However, it has the characteristics of a medium-sized university as regards the number of students. Data 2012 Students: 36.855 Degrees: 74 Faculties and Schools: 13 Lecturers: 2.764 Staff: 2.617

Table 1. General characteristics of the UPV

The UPV sample consisted of 265 students enrolled in Social Science degree programs (12%), 189 in experimental sciences and health (8%), and the rest (80%) were engaged in engineering studies, the fundamental faculty at the UPV. This paper is focused exclusively on engineering students (n = 1790), whose basic variables are shown in Table 2.

Some subjects, Aeronautical Engineering for example, were not included in the sample due to insufficient numbers. Ages ranged from 18 to 21 in all samples, so that there was wide homogeneity in the distribution (M = 18.77; SD = 0.38). The students tended to come from the technical areas of the Baccalaureate (98% from Science and Technology and the residual 2% from other branches). The cut-off mark ranged from 5 (minimum access) to 12.96 (maximum possible 14 points), reaching M = 8.12 and SD = 1.52, but with a very skewed distribution towards the tail on the right.

Gender		Degree Groups			Cohorts	
	Agronomics	Construction and Civil Engineering	Information and Communication Technology	Industrial Engineering	2010-2011	2011-2012
376 (21%) female 1414 (79%) male	395 22%	420 23%	440 25%	535 37%	805 45%	985 55%

Table 2. Samples of UPV engineering students in 2010/2011 and 2011/2012.

Questionnaire

Each student was administered the questionnaire proposed by Guardia et al. (2012), [10], which showed good values for reliability and validity. The questionnaire variables related to access was divided into two second order factors (social and individual) defined by six primary factors: Consideration of the University; Perceived Utility and Social Considerations as social factors. The primary factors were:

Vocational Aspects; Influence of Geographical Location for the individual factor, including the Access Grade, Method of Access and Gender. In the initial study, Cronbach's α values ranged between .84 and .95 for all factors. The factorial validity analyses carried out with confirmatory factor analysis also showed a good fit that confirmed the structure of the factors described (χ^2 = 1234.74; p =. 18). The questionnaire consisted of a total of 25 items defined in a scale of ordinal response from 1 (maximum disagreement) to 7 (maximum agreement) leaving a complementary value for cases in which the proposed item did not apply; for example, if students had not attended the university's open-day or did not have access to the services of an educational counselor, etc.

In addition to the variables included in the questionnaire, data were obtained from institutional variables, i.e. cut-off marks, places offered in the academic years considered, numbers of students enrolled in each degree course, etc. While some of this information was gathered from the students themselves, data was also obtained from official sources, i.e., the universities involved and the official statistical data from the Education Ministry (www.educacion.es). The data from the students themselves was not analyzed since it responded to secondary objectives of the overall project and was designed to estimate the real knowledge of the students accessing the University System.

Procedure
The questionnaire was administered by computer in such a way that the final total sample emerged from an accidental sampling. The results and data for each student were then processed according to the planned statistical analysis that was performed using SPSS v 21.0 and Amos v 19.0.

Statistical analysis
To conform to the general precepts and assumptions of structural equation models, the following statistical assumptions were made for quantitative variables $E(X_i) = E(Y_i) = E(\xi_i) = 0$ and $Var(X_i) = Var(Y_i) = Var(\xi_i) = 1$. As a result, all quantitative variables were transformed through reduction and standardization and, likewise, $E(\varepsilon_i\varepsilon_j) = E(\delta_i\delta_j) = E(\xi\delta) = E(\eta\varepsilon) = E(\zeta_i\zeta_j) = 0$; assuming initially that the errors of measurement were uncorrelated to each other, as was the case of the observable and latent variables. The categorical observable variables (type of Baccalaureate and gender) were considered as such and subjected to an estimation process described below, [11, 12].

The structures of the exogenous measurement models (Λ_x) are not described for reasons of space. We should just point out that we measured the correlations between exogenous variables (both observable and latent) that previous studies have shown to be relevant. The exogenous measurement models specified in the proposed model meet the usual application conditions of order [13, 14]. In addition, the proposed model verifies the identification condition since it presents positive degrees of freedom (degree of freedom df = 321). Further information on specific model data can be found in [9].

Results

First, parametric statistical tests were carried out to assess whether various samples considered showed any significant differences in relation to the gender and access marks of the students. No significant difference was obtained, so that the modifying effects of marginal distributions could be ruled out of the later analyses. The only exception was the distribution of the gender variable, in which the proportion of women compared to an equiprobable population distribution was significantly lower than that of men (χ^2 = 89.43;) (p <. 001, V =.82), as the UPV students are mostly men. We analyzed the replies to the questionnaire according to the factors described, comparing results by degree courses and did not find any statistically significant differences between courses.

Thus, for each of the subsamples (engineering branches) the matrix of Pearson correlation between all the variables involved in the analysis was obtained, taking into account that in pairs in which the gender variable was considered it was estimated by biserial correlations and the type of Baccalaureate were estimated by polychoric correlations. The different parameters of the model were estimated using the

AFD technique to estimate free distribution, since many of the variables involved presented high values of asymmetry. Although these did not affect the estimation of the correlation, given the large sample size, they could have affected the value of the estimates of the structural parameters. All the analyses were carried out using SPSS and Amos software. Table 3 shows the overall fit in the UPV, assuming the four branches of engineering in a joint mode. Table 4 shows these parameters differentiated by study field. As can be seen in both tables, the goodness of fit indicators, show a good overall fit and also in the different fields. It should also be pointed out that in the construction engineering field the fit of the model is not so good.

INDICATOR	ESTIMATIÓN
Goodness of Fit Index (GFI)	.976
Adjusted Goodness of Fit Index (AGFI)	.958
Bentler Bonnet Normed Fit Index (BBNFI)	.961
Bentler Bonnet Non Normed Fit Index (BBNNFI)	.969
Comparative Fit Index (CFI)	.952
Coefficient of Determination (R2)	.491
Root Mean Standard Errors (RMSE)	.006
Standardized Root Mean Standard Errors (SRMSE)	.002
$\chi 2$ with df = 321	915.67 (p<.05)
Ratio $\chi 2$ / df	2.85
Values of reliability for latent factors	
OPINION OF THE UNIVERSITY	α = .799
PERCEIVED UTILITY OF DEGREE	α = .801
SOCIAL CONSIDERATIONS	α = .812
VOCATIONAL ASPECTS	α = .799
INFLUENCE OF SURROUNDINGS	α = .791
GEOGRAPHICAL LOCATION	α = .798

Table 3. Overall setting for Structural model only for UPV engineering students, with an indication of reliability (estimation of α of Satorra-Bentler).

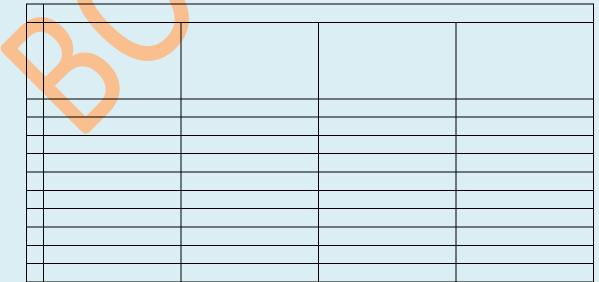|			

Table 4. Overall settings of each specialty

The structural parameters derived from each of the four models described in Table 4 were also estimated to make it possible to analyze any differences from the effects in the general model, dividing the degree courses into different groups. Table 5 summarizes the standardized estimates of each structural parameter value.
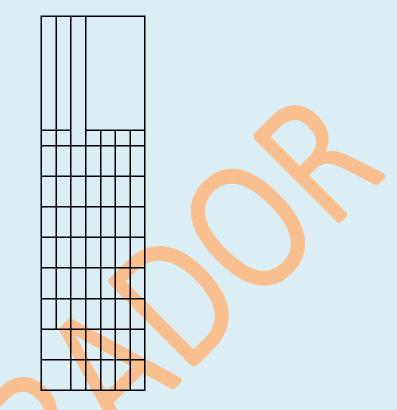

* p <.05 * p <.01

Table 5. Estimation of each standardized structural parameter for degrees at the UPV

Conclusions

The first issue to highlight has to do with the adjustment values of the models analyzed; while the fit of the χ^2 statistic was not particularly good, we consider that the general model proposed to explain the demand for the students' first choice in engineering studies at the UPV under the Spanish public system could be a suitable model. This conclusion is based on the fact both in the case of the overall adjustment of the UPV total and in the adjustment for the different engineering courses, the adjustment rates are good, since the values of GFI, AGFI, BBNFI, BBNNFI, or IFC indices are over .90 and in some cases more than .95. The SRMR values are lower than .011 and SRMSE values are below .005, (as proposed by Hu & Bentler [11]). These indices are acceptable when they are superior to .90 (GFI, AGFI, BBNFI, BBNNFI and CFI) and for SRMR ≤ .05; and RMSEA ≤ .06), also the χ^2/df ratio < 2 indicates an excellent fit, χ^2/df < 3 a good fit and χ^2/df < 5 an acceptable fit [11], and in our case for all the models fitted these ratio are lower than 3.These results coincide with the model proposed by Guardia et al [10] for the first version of the present model

It can also be pointed out that in general the hypothesized parameters in the model are statistically significant in all cases, which is another argument in favor of the proposed model's ability to explain the demand for the first choice of university degree courses. All the estimated values are statistically significant with a confidence level of 95%. However, there exist certain differences that should be pointed out. In Table 5 the value of the estimated parameter is very different for the specialties analyzed. The standardized estimates provide a simple descriptive analysis. Thus, in the case of ICT, the end endogenous variable to be explained (the demand for first choice (Y_2)) presents higher parameter values due to the impact of the exogenous variables (ξ_1, ξ_2 and ξ_3), in most cases γ_{ij} is of the order of .60 or

higher. This could be explained by the fact that ICT is composed of a range of courses identified with the present time and thus is given preference in the demands of the students. Despite new offers, new UPV students see ICT as an attractive choice with a strong impact of individual and social factors for selection as the first option.

The cut-off mark also has a great influence on the perception of the offer/demand in the access year, especially in ICT (γ_{23} = .676). This is consistent with the above, since the UPV is particularly strong in engineering, which therefore reaffirms the predominant role of the ICT field as compared to the other three fields of Construction, Agricultural and Industrial Engineering.

This work has certain limitations that should be considered and which mainly involve two issues. The first is the size of the asymmetrical samples, which means that some estimates are somewhat skewed by this effect. The use of standardized estimates facilitates presentation and study, but does not solve all the problems of sampling. The second issue is that, in order to keep the results as simple as possible, we did not carry out strict statistical comparisons between the various parameters using the usual mechanisms (LM or Wald Test). The additional parameters and models would have involved extra complexity without providing special information. For the same reason we also decided not to incorporate the global adjustment values based on the Akaike (AIC) or Bayesian criteria (BIC).

References:

[1]   Murphy, P.E. y McGarrity, R.A. (1978). Marketing Universities: A Survey of Student Recruitment Activities. *College and University*, 53, 3, 249-61.

[2]   Yurtseven, T. (2002). *How does the image of engineering affect student recruitment and retention? A perspective from the USA*. New York: Global Journal of Engineering Education.

[3]   Turner, C.S.V. y J.R. Thompson, J.R. (1993). Socializing Women Doctoral Students: Minority and Majority Experiences. *Review of Higher Education*, *6*, 232-241.

[4]   Ford, D.Y. (2008). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *The Journal of Special Education*, 31(1), 4-14.

[5]   Capilla, R. (2009). *Análisis estratégico de los estudios TIC en la Universidad Politécnica de Valencia*. Tesis Doctoral no Publicada. Valencia: Universitat Politècnica de València. (http://riunet.upv.es/handle/10251/5767).

[6]   Raymond, S. (2001). Predicting Academic Succes of First-Year Engineering Students from Standardized Test Scores and Psychological Variables. International Journal of Engineering Education, 17 (1), 75-80.

[7] Huang, Sh. and Fang, N. (2013). Predicting student academic performance in an enginnering dynamics course: A comparison for four types of predictive mathematical models. Computers & Education, 61, 133-145.

[8] Veenstra, C.P.; Dey, E.L. and Herrin, G.D. (2008). Is modeling of freshman engineering succes different from modeling of non-engineering succes? Journal of Engineering Education, 467- 479.

[9] Guàrdia, J.; Peró, M.; Hervás, A.; Capilla, R.; Soriano, P,P. & Porras, M. (2012). Factores asociados con la decisión de cursar estudios universitarios de Psicología. Una aproximación mediante modelos de ecuaciones estructurales. *Anuario de Psicologia*, *42* (1), 87-104.

[10] Guàrdia. J.; Peró, M.; Hervás, A.; Capilla, R.; Soriano, P,P. & Porras, M. (2013). Factores asociados con la elección de titulación en el sistema universitario público español. Estudio de un modelo de ecuaciones estructurales. En revisión editorial.

[11] Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

[12] Ory, D.T. & Mokhtarian, P.L. (2010). The impact of non-normality, sample size and estimation technique on goodness-of-fit measures in structural equation modeling: evidence from ten empirical models of travel behavior. Quality & Quantity, 44(3), 427-445.

[13] Palomo, J.; Dunson, D.B. & Bollen, K. (2007). Bayesian structural equation modeling. In S.Y.Lee (Ed.). Handbook of latent variable and related models, pp. 163-179. New York: Elsevier.

[14] Poon, W.Y. & Lee, S.Y. (1994). A distribution free approach for analysis of two-level structural equation model. New York:Elsevier.