



A student's academic index using multivariate techniques

Maria Elena López Zamora, Lelly Useche, Carlos Montes Rodríguez

► To cite this version:

Maria Elena López Zamora, Lelly Useche, Carlos Montes Rodríguez. A student's academic index using multivariate techniques. Bulletin of Computational Applied Mathematics, 2021, 9 (1), pp.33-62. <hal-03551705>

HAL Id: hal-03551705

<https://hal.science/hal-03551705v1>

Submitted on 1 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A student's academic index using multivariate techniques

Maria Elena López Zamora^{1 2}, *Lelly Useche*³,
*Carlos Montes Rodríguez*⁴

CompAMa Vol.9, No.1, pp.xx-xx, 2021 - Received April 12, 2021 - Accepted September 26, 2021

Special Issue III CICB 2020, Eds. Carmen Judith Vanegas and Saúl Buitrago Boret

Abstract

This work faced the construction of an index composed by the factors that influence the student's academic performance. To fulfill this objective, the information from students was collected by means of a questionnaire and later, the exploratory analysis of data was made, where the relevant characteristics of the sample were detected. To study the relations between the factors involved, a multiple correspondence analysis was made and the index was constructed by means of a factorial analysis. The index showed that the student's high performance is mainly associated to data related to academic achievement and personal relationships, owed to their factorial charges of 0.787 y 0.708 respectively. The compound index can be applied in other institutions, serving as basis for decision making and reforms in the educative sector, in order to obtain a lower desertion and a higher academic performance, among other aspects of the educational problems.

Keywords: Index; factor analysis; correspondence analysis; linear combinations;

¹Estudiante de Maestría en Estadística Aplicada, Instituto de Posgrado, Universidad Técnica de Manabí, Ecuador. (malena9461@gmail.com). (ORCID iD: <https://orcid.org/0000-0003-1183-9082>).

²Corresponding author.

³Instituto de Ciencias Básicas, Departamento de Matemáticas y Estadística, Universidad Técnica de Manabí, Ecuador. (lelly.useche@utm.edu.ec). (ORCID iD: <https://orcid.org/0000-0002-4294-9009>).

⁴Centro de Promoción y Apoyo al Ingreso, Área de Nivelación, Universidad Técnica de Manabí, Ecuador. (carlos.montes@utm.edu.ec). (ORCID iD: <https://orcid.org/0000-0001-8302-4056>).

academic performance.

Resumen

El presente trabajo contempló la construcción de un índice compuesto de los factores que influyen en el rendimiento académico. Para el cumplimiento del objetivo se recolectó la información de los estudiantes a través de un cuestionario, posteriormente, se realizó el análisis exploratorio de los datos donde se observaron las características relevantes de la muestra. Para estudiar las relaciones entre los factores se aplicó un análisis de correspondencias múltiple, y posteriormente, análisis factorial para la construcción del índice. El índice mostró que el rendimiento alto de los estudiantes se encuentra asociado a datos favorables principalmente para los factores de desempeño académico y relaciones personales debido a sus cargas factoriales 0.787 y 0.708 respectivamente. El índice compuesto puede ser replicado en otras instituciones, sirviendo de base para la toma de decisiones y reformas en el sector educativo a fin de obtener una tasa menor de deserción y un mejor rendimiento académico, entre otras problemáticas.

Palabras claves: Índice; análisis factorial; análisis de correspondencia, combinaciones lineales, rendimiento académico.

1 Introduction

The study of factors and patterns that allow explaining the influences on student's achievement, is of great interest for all academic organizations, because it enables them to structure plans and strategies focused in improving the academic performance and in consequence, the education quality that they offer [1]. The conceptualization of academic performance is complex and some authors indicate that there are limitations to explain this concept in a precise way [2, 3]. It can be said that the performance in the classroom is the acquisition of knowledge and its correct application in the resolution of problematics in a specific area or science [4, 5].

Studying the academic performance involves investigating all the possible factors that affect the individual's achievement. These factors are varied and not always act in the same way on a student, that is to say, they have different effects in different persons [6, 7]. The academic performance analysis by factors is made all over the world because of its relevance in the educational system. This analysis is made by various techniques or methodologies that aim to explain the positive and negative effects, on the academic achievement

[8, 9].

Data in the majority of these cases, are multidimensional and have effects on the matrix analysis and in the resultant distributions. The objective of multivariate techniques is, with no doubt, to particularize significant aspects of the input data [10, 11]. The statistical techniques as the correspondence analysis (CRA) and the principal components analysis (PCA), are consolidated as efficient methodologies for the information reduction. The frequent use of the CRA technique has its advantages in the factorial charges planes, that explain the relations between interdependent variables besides of the construction of indicators conformed by linear combinations of the dimensions. Some of the relevant examples related to the study of performance are: the construction of synthetic indicators [12], relation between residence regions and academic performance [13], familiar support and performance [14], among others.

Factorial analysis allows identifying homogeneity groups for the variables addressed in the study; that is to say, the correlations are analyzed and at the same time, the independences between groups are observed. The technique's relevance is centered in that each variable has the same importance, besides determining weights that explain some percentage of the original variance. Several authors have used the technique to explain the influence of factors on performance [15, 16].

The importance of the research on academic performance done through the years and in different educational centers, was observed from the bibliographic review made. Considering the background exposed, the objective of this research is to construct a student's academic index using multivariate techniques as correspondence and factorial analysis, to demonstrate that there exists a significative influence of a set of factors or variables on the student's performance.

2 Materials and Methods

The necessary methodology for the creation of a compound academic index by multivariate statistical techniques, is summarized in figure 1. In this research, the index is defined as the relation between the variables or independent factors influencing the academic performance (dependent variable) of students or individuals of research. The work takes as basis, relevant information from students of the Technical University of Manabi (UTM),

Ecuador, that is collected in an anonymous way to generate confidence in the studied individuals and to preserve their privacy.

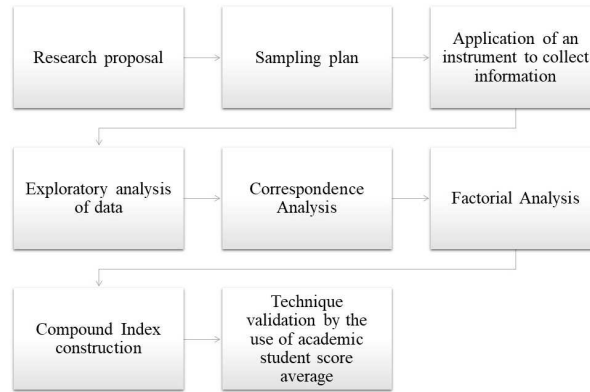


Fig. 1: General scheme of this research

2.1 Sampling plan

The sampling plan is centered in the identification of the variables studied that result from the bibliographic analysis of works that explain the influence of one or several factors on the academic performance. The analyzed information was obtained from databases as Scielo, Latindex, Science Direct, Dialnet, among others. This first analysis together with the instrument developed by [17], allowed to define five groups of variables that will be described below.

1. Economical data includes the variables of coexistence, parents' studies, economic income, basic services and transport.
2. Employment situation describes the work environment, the schedules flexibility and the work history of the group of students that has decided to work and study at the same time.
3. Personal relations address the factors associated to intra and interpersonal relations with subjects as the person's partner, family, teachers, schoolmates, among others.

4. The academic achievement was considered from the student's perspective, defining himself if subjects as family, person's partner, school-mates and others, affect positively or negatively his or her academic performance.
5. General data allow to identify the essential sample characteristics as sex, age, study career or major, semester, and associate them to the research on academic performance.

Some of these variables have been studied previously in relation with their influence in the students' academic achievement by several authors [16,18,19].

2.2 Instrument

This study collected personal information of a group of students that are enrolled in the UTM, including the score average or academic performance, whose range is from 0 to 100 points. The instrument is a survey that includes several scenes of factors that can influence in a significant way the student's academic performance. The survey was adapted to this research, taking the variant presented by [17], for the adaptation it was necessary to carry out small changes in the selection variants such as the currency exchange, the range of values for the admission to the institution and small modifications in the formulation of some questions (Annexes). It was asked to each student to complete the questions with the greatest possible veracity and additionally they were informed about the research objective. The questionnaire was validated through experts' judgment in the areas of education and statistics.

2.3 Sample calculation

The research's empirical scene took as object of study, students enrolled in the UTM distributed in 15 careers, different, but related to the subjects of mathematics and statistics, during the second ordinary academic period of 2019 or (2019-S2). Both subjects are of obligatory character, the semester level and the number of credits (3 or 4) vary by career.

The careers included in this study are: Mechanic, Industrial, Electric, Chemical and Civil Engineering, Administration, Social Work, Medicine, Economy, Agronomy, Veterinary, Informatics, Agriculture and Pedagogy. The population of students for the period 2019-S2 that take the mentioned

Tab. 1: Sampling stratus

<i>Faculty</i>	<i>Population (N)</i>	<i>Weight (w_i)</i>	<i>Sample (n)</i>
<i>F.C.A.E.</i>	890	0.11	56
<i>F.C.S.</i>	243	0.03	16
<i>F.C.H.S.</i>	728	0.09	42
<i>F.C.I.</i>	405	0.05	24
<i>F.C.M.F.Q.</i>	3481	0.43	206
<i>F.C.V.</i>	405	0.05	24
<i>F.F.L.C.E.</i>	567	0.07	32
<i>F.C.AGRI.</i>	486	0.06	29
<i>F.C.AGRO.</i>	891	0.11	53

subjects was of 8096 students. The calculation of the pilot sample is made using the equation:

$$n = \frac{z^2 \cdot N \cdot \delta^2}{\delta^2 \cdot z^2 + (N - 1)e^2} \quad (1)$$

where: N is the population, z is the critical value in function of the confidence level, δ^2 is the variance for the studied population, e represents the precision level and n is the sample.

For the moment in which the sample calculation was done, there was no evidence of variance difference between each stratum and, based on this fact, it was decided to take the stratified sample by proportional weight, considering also that the number of enrolled students per academic career is not the same in all cases.

The variance of the student's score average was estimated from a pilot sampling, as a first step to calculate the sample's size. There were considered values for $z = 2.58$ with respect to a high level of confidence (99%), $\delta = 8.509$ y $e = 1$, and it was obtained as result, that the minimum number of students necessary to be surveyed for this conditions was of 455. During the selection of students in each Faculty, it was established a slightly higher value of the calculated sample, of 482 students. The sample distribution is formed by several strata which correspond to one Faculty of the UTM (Table 1).

The analyzed individuals form a group of 482 students, where the 65% are men and the 35% are women. All of them participated voluntarily and their data were treated under standards of ethical principles.

2.4 Exploratory analysis of data

The data analysis by statistic descriptives, is the previous and necessary step to understand the nature of the collected data [20, 21]. Observing the frequencies of factors, as part of the exploratory data analysis, allows identifying the principal variables that could affect the student's achievement in the classroom, besides confirming the variables normal distribution assumptions.

This methodology has the objective of identifying the relevant characteristics of a particular set of data by the use of graphics and numerical tables and also allows to define theoretical models that make easier the representation of the selected population. This preliminary exploration of data considers atypical values, missing values, discontinuities, distributions, concentration or data frequency, among others. This analysis can be made in a simplified way to each group or in a joint way.

2.5 Correspondence analysis

The use of this technique has the objective of summarizing a massive data quantity, by the use of dimensions with a minimal as possible loss of information. In contrast to the factorial analysis methodology, the correspondence analysis has its fundamental application in categorical type variables or ordinals, as it is shown in different research works [22–24].

The analog measure for the total variation in the correspondence technique will be the total inertia, which gives value to the dependence grade for the categorical variables (x, y). This indicator is calculated from the relation of the typified residue:

$$I_t = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^n \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \sum_{h=1}^H \mu_h^2 \quad (2)$$

$$\sum_{h=1}^H \mu_h^2 = \sum_{h=1}^H \sum_{i=1}^n N_i \cdot a_{ih}^2 = \sum_{h=1}^H \sum_{j=1}^n N_j \cdot b_{jh}^2 \quad (3)$$

where: N is the population, n_{ij} is the observed frequency, e_{ij} is the expected frequency, u_h represents the singular values product of symmetric normalization, N_{i*} is the sum of cells in row of the contingency table, N_{*j} is the sum of cells in columns of the contingency table, a_{ih} represents the row points and b_{jh} represents the column points of the cartesian coordinate matrices.

For the determination of the dimensions' importance, the explained inertia that comes from the general or total inertia is calculated. The observed dependencies are calculated by means of:

$$\left(\frac{\mu_i^2}{I_t} \right)_{i=1,2,\dots,H} \quad (4)$$

With the accumulated inertia, the minimal quantity of dimensions that allows explaining the calculated or found dependencies is defined.

$$\left(\sum_{h=1}^i \frac{\mu_i^2}{I_t} \right)_{i=1,2,\dots,H} \quad (5)$$

Others of the important points in the correspondence analysis methodology, are the total and relative contributions. The used factorial axis interpretation is made through the total contributions, always observing the stronger or greater weight contributions. On the other hand, the partial contributions allow us to measure the point's proximity in the i-th file or column of the factors that have presented a high contribution.

2.6 Factorial analysis

It is of relevance to determine the minimal number of dimensions (factor) and that, at the same time, that these be the ones that explain great part of the initial information [15, 25]. The difference of this technique with others as the variance or regression analysis, is that each one of the variables in this plane has the same importance, it is to say, in a first instance, the variables are independent one from each other.

The factor results from a linear combination of variables. The answers of the individuals will not always be the same, therefore we will have a variety of linear combinations. The total variance will allow us having a better appreciation of the sample diversity. The factorial analysis technique is centered in finding weights that allow to explain the original variance of the variables that were presented in first instance. The technique's summarized process is showed below (Figure 2).

The analysis of the correlation matrix is indispensable to evaluate the factorial analysis effectivity, generally it must be observed that the variables show high correlations between each other. If this is not the case, the evident procedure is opting for another multivariate technique. For this technique

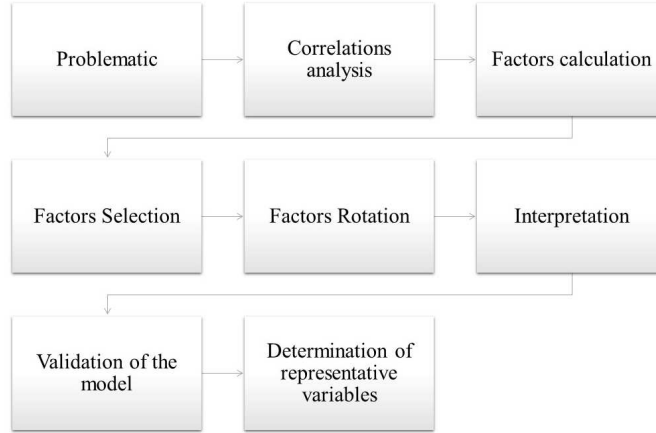


Fig. 2: Factorial analysis process

two relevant indicators are used, the Barlett Sphericity Test and the Kaiser Meyer Olkin (KMO) [26]. Also, the Kaiser rule will be used to determine the number of minimum factors and the rotation of factors by the Quartimax Method, which maximizes the variance for factorial charges.

2.7 Index Construction

The combination of variables from which the maximum correlations will be obtained, is defined by the instrument of collection of information, this fact takes us to generate the first five simple character indices that result from the multiple correspondence analysis linear combinations, which, after the factorial analysis, will result in the compound index.

The simple character indices correspond to: Economical data, Employment situation, Sentimental relations, Personal relations and Academic performance. To ease the compound index comprehension, the averages of the proposed index for each one of the students, are categorized by percentiles. The exploratory data analysis, the correspondence analysis, the factorial analysis and the index construction were made using the statistical software SPSS Statistics®.

3 Results

3.1 Exploratory Analysis

From the exploratory analysis, the fundamental characteristics of the sample of the students' population studied. In Table 2, the categories of each one of the variables groups with the maximum frequency, can be reviewed. This first analysis allowed us to determine the statistical techniques for the reduction of data and a deep analysis of the relations between variables and their incidence in the individual's achievement in classroom.

With respect to the economical data, it is observed that a 55% of the students live together with their parents and the most frequent scholar level for both parents is the Secondary level. A 49.6% of the sample declared that they economically depend on their parents and a 63.1% of the students receive in average an income from "0 to 100 USD". The most frequent services that the students have are the printer devices (35.1%) and internet connection (30.5%), also a 74.3% uses the public transport service for their attendance at the institution. For the students that work, it was observed that the average daily working hours is "from 1 to 7 hours" (20.41%), while the flexibility and study affectation is "regular". A 52.07% of the sample said that they had a sentimental partner and dedicate an average time for sharing coexistence "from 0 to 5 hours", they don't generally work or study together. Personal relations between the student and other individuals, are generally qualified as good or very good. On reference of other influences on the student's academic performance, the students say that family, friends in the institution and tutorials, positively affect their academic achievement

3.2 Correspondence Analysis

The categorization of each dimension for each one of the variables was obtained. Also, the Cronbach alpha coefficient (α_{Cr}) was analyzed which explains the consistency between data from the average of correlations, the data inertia and the variance percentage (Table 3). The dimension one was selected for the respective analysis.

The α_{Cr} value for the dimension one, except for the group of economical data, has superior values than the minimum acceptable (≥ 0.7), fact that indicates a high data internal consistency. With respect to data inertia, there exists a notorious dispersion that is explained by the different sets inte-

Tab. 2: Characteristics of the analyzed individuals

<i>Groups</i>	<i>Variables</i>	<i>Categories</i>	<i>n(%)</i>
Economical data	Coexistence	Parents	265(55)
	Mother's school level	Secondary	221(45.9)
	Father's school level	Secondary	237(49.2)
	Economic dependence	Totally	239(49.6)
	Incomes	(0-100)USD	304(63.1)
	Services	Printer device	169(35.1)
	Transport	Bus	358(74.3)
	Work	No	291(60.37)
Employment situation*	Time of working	(0-11)months	29(15.35)
	Hours of work	(1-7)hours	39(20.41)
	Flexibility	Regular	26(13.6)
	Affects study	Regular	28(14.7)
Sentimental relationships*	Partner	Yes	251(52.07)
	Time shared with partner	(0-5)hours	54(21.57)
	Partner studies	Yes	95(37.97)
	They study together	No	101(40.25)
	Works with partner	No	82(32.6)
Personal relationships	With family	Very good	318(65.97)
	With partner	Very good	171(35.44)
	With friends	Good	228(47.30)
	With classmates	Good	213(44.19)
	With workmates	Good	129(26.76)
	With teachers	Good	259(53.73)
Academic achievement influence of	Family	Very positively	225(46.68)
	Partner	Does not affect	114(23.65)
	Work	Does not affect	119(24.69)
	Friends out from the university	Does not affect	223(46.26)
	Friends in the university	Positively	201(41.7)
	Workmates	Does not affect	100(20.74)
	Classmates	Does not affect	204(42.3)
	Psycho-pedagogical Department	Does not affect	240(49.8)
	Tutorials	Positively	191(39.6)

*For the variable it is used a number of sample that fulfills the group criteria.

Tab. 3: Summary of the model for the variables of the different groups

Group	Dimension	Cronbach alpha	Accounted variance for		
			Total (self-value)	Inertia	Variance %
Economic data	1	.544	1.909	.239	23.860
	2	.437	1.618	.202	20.227
Employment situation	1	1.000	4.000	1.000	100.000
	2	.625	1.881	.470	47.029
Sentimental relationships	1	1.000	5.986	.998	99.771
	2	.445	1.588	.265	26.474
Personal relations	1	.745	2.636	.439	43.934
	2	.590	1.967	.328	32.790
Academic performance	1	.898	4.951	.550	55.008
	2	.831	3.822	.425	42.465

grated by persons with the characteristics that resulted from the exploratory analysis of data. The inertia's high value for all the the cases in dimension 1, shows us that the profiles row with respect to the column's vertices, are nearer, thus, there exists a good relation between inertia and the correlation between files and columns coefficient.

The most significant discriminant measures for the dimension 1, are those which form an angle near to zero with respect to the horizontal axis. In this way, for the group of economical data we have the example of the mother's scholar level (Figure 3).

In the graphic of the set of category points (Figure 4), it is observed that the first factor describes the levels of education of both parents and relates them with the other variables in the group or dimension of economical data. In the first quadrant there is a group constituted by individuals whose parents have university or superior studies and their economic income is high, and also have a familiar car for transportation to the institution. In the second quadrant we observe a medium income and partial economic dependence, in other words, the student works. For the third quadrant, the secondary level of study of both parents predominates and the student, with a lower income, uses public transport. Finally, the fourth quadrant shows total economic dependence, because the student does not work. From the linear combinations of the group, it was observed that some variables as the

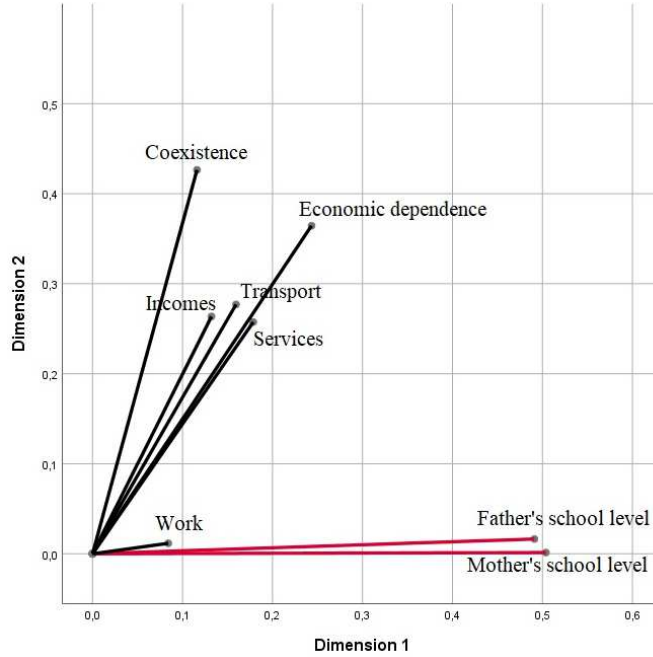


Fig. 3: Discriminant measures of economical data

coexistence with parents, mother's other studies, father's university studies, total economic dependence, among others, have the higher coefficients or weights, which become in a positive influence for the academic performance.

The analysis of discriminant measures and set of category points was made for each one of the variables groups, with the purpose of observing the most significant factors that influence the academic performance. For the dimension Employment situation, the variables daily working hours and work affectation stand out. With respect to the dimension of Sentimental relations, the variables they study and work together, are the most representative; while for the dimension of Personal relations, the relations with workmates are the most significant. The dimension Academic achievement showed that all the individuals influence in the student's academic performance, but in different extent.

The dimension 1 is composed of the sum of all the linear combinations of the variables generating a subscript (D_k); from the combinations it is possible to observe which one of the answers has a higher weight for the explanation of the information. For example, in Table 4, the group of variables

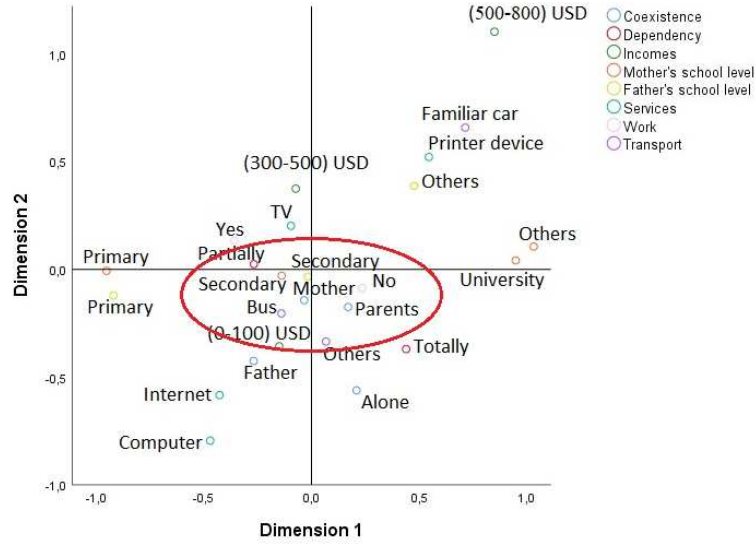


Fig. 4: Set of category points for economical data

of achievement factors, the influence of the individuals analyzed shows how they positively influence on performance, at the same time, these subjects have the major weight or coefficient in the combination. These observations vary from one variable to another.

Tab. 4: Subscripts (D_k) of the first dimension, centroid coordinates

Variables	Subscript
Coexistence	$D_1 = -0.269(Father) - 0.34(Mother) + 0.170(Parents) - 0.918(Partner) + 0.209(Alone)$
Mother's scholar level	$D_2 = 0.952(Primary) - 0.139(Secondary) + 0.948(University) + 1.031(Others)$
Father's scholar level	$D_3 = -0.919(Primary) - 0.017(Secondary) + 1.156(University) + 0.477(Others)$
Dependency	$D_4 = -1.081(Nothing) - 0.268(Partially) + 0.439(Totally)$
Income	$D_5 = -0.150[(0 - 100)USD] - 0.073[(300 - 500)USD] + 0.849[(500 - 800)USD] + 1.262[(800 - 1000)USD]$
Services	$D_6 = -0.470(Computer) - 0.427(Internet) - 0.095(TV) + 0.545(Printerdevice)$

Transport	$D_7 = 0.714(\text{Familiar car}) - 0.238(\text{Own car}) + 1.752(\text{Taxi}) - 0.140(\text{Bus}) + 0.067(\text{Others})$
Work	$D_8 = -0.358(\text{Yes}) + 0.235(\text{No})$
Working time	$D_9 = 0.810(\text{Does not apply}) - 1.234[(0 - \text{More than 5 years})]$
Working hours	$D_{10} = 0.810(\text{Does not apply}) - 1.234[(1 - \text{More than 48 hours})]$
Work flexibility	$D_{11} = 0.810(\text{Does not apply}) - 1.234[(\text{Very flexible}) + (\text{Flexible}) + (\text{Regular}) + (\text{Not very flexible}) + (\text{Nothing})]$
Affectation of work	$D_{12} = 0.810(\text{Does not apply}) - 1.234[(\text{Much}) + (\text{Regular}) + (\text{Little}) + (\text{Nothing})]$
Partner	$D_{13} = -0.954(\text{Yes}) + 1.036(\text{No})$
Time with partner	$D_{14} = 1.051(\text{Does not apply}) - 0.951[(0 - 5)h] - 0.948[(6 - 10)h] - 0.954[(11 - \text{More than 16h})]$
The partner studies	$D_{15} = 1.051(\text{Does not apply}) - 0.952(\text{Yes}) - 0.949(\text{No})$
They study together	$D_{16} = 1.051(\text{Does not apply}) - 0.954(\text{Yes}) - 0.950(\text{No})$
Partner Works	$D_{17} = 1.051(\text{Does not apply}) - 0.950(\text{Yes}) - 0.952(\text{No})$
Works with partner	$D_{18} = 1.051(\text{Does not apply}) - 0.928(\text{Yes}) - 0.953(\text{No})$
Family	$D_{19} = -1.641(\text{Very bad}) - 1.150(\text{Bad}) - 1.072(\text{Regular}) - 0.547(\text{Good}) + 0.339(\text{Very good})$
Partner	$D_{20} = 0.018(\text{Does not apply, Very bad}) - 1.372(\text{Bad, Regular}) + 0.146(\text{Good, Very good})$
Classmates	$D_{21} = -0.892(\text{Very bad}) - 0.567(\text{Bad}) - 0.801(\text{Regular}) - 0.189(\text{Good}) + 1.377(\text{Very good})$
Friends	$D_{22} = -0.040(\text{Very bad}) - 1.562(\text{Bad}) - 1.258(\text{Regular}) - 0.494(\text{Good}) + 0.821(\text{Very good})$
Workmates	$D_{23} = 0.024(\text{Does not apply, Very bad}) - 1.682(\text{Bad}) - 0.536(\text{Regular, Good}) + 1.457(\text{Very good})$
Teacher	$D_{24} = -0.495(\text{Very bad}) - 1.133(\text{Bad}) - 0.871(\text{Regular}) - 0.232(\text{Good}) + 1.253(\text{Very good})$

Family	$D_{25} = -1.173(\text{Very negatively}) - 0.708(\text{Negatively}) - 1.079(\text{Does not affect me}) + 0.023(\text{Positively}) + 0.638(\text{Very positively})$
Partner	$D_{26} = 0.096(\text{Does not apply}) - 0.499(\text{Very negatively}) - 0.912(\text{Negatively, Does not affect me}) + 0.515(\text{Positively, Very positively})$
Work	$D_{27} = 0.233(\text{Does not apply, Very negatively}) - 0.643(\text{Negatively}) - 0.458(\text{Does not affect me, Positively}) + 1.479(\text{Very positively})$
Friend outside	$D_{28} = -0.065(\text{Very negatively}) - 0.773(\text{Negatively}) - 0.719(\text{Does not affect me}) + 0.432(\text{Positively}) + 1.478(\text{Very positively})$
Friend inside	$D_{29} = -0.261(\text{Very negatively}) - 1.108(\text{Negatively}) - 0.843(\text{Does not affect me}) + 0.311(\text{Positively}) + 1.446(\text{Very positively})$
Workmates	$D_{30} = 0.225(\text{Does not apply}) - 1.240(\text{Very negatively}) - 0.796(\text{Negatively, Does not affect me}) + 0.388(\text{Positively}) + 1.609(\text{Very positively})$
Classmates	$D_{31} = -0.578(\text{Very negatively}) - 1.077(\text{Negatively}) - 0.699(\text{Does not affect me}) + 0.340(\text{Positively}) + 1.499(\text{Very positively})$
Psycho-pedagogical Department	$D_{32} = -0.843(\text{Very negatively}) - 1.173(\text{Negatively}) - 0.599(\text{Does not affect me}) + 0.485(\text{Positively}) + 1.903(\text{Very positively})$
Tutorials	$D_{33} = -1.016(\text{Very negatively}) - 1.254(\text{Negatively}) - 0.761(\text{Does not affect me}) + 0.347(\text{Positively}) + 1.068(\text{Very positively})$

3.3 Factorial Analysis

The first step of the procedure was to obtain the original subscripts for the groups of variables. These subscripts are composed with the factorial charges of the multiple correspondence technique, that is to say that, the process consisted in entering the first axis charges of all dimensions, in the analysis.

Tab. 5: Matrix of rotated components

	Component	
	1	2
Economical data	.451	.605
Employment situation	.265	
Sentimental relations	-.157	.841
Personal relations	.708	
Achievement factors	.787	

Some data of the original indexes have numerical values higher than 1, that can be negative or positive. To standardize data or represent them in form of percentages, each term is divided by the maximum absolute value in the column. With this procedure, the values for the standard index oscillate between 0 and 1.

The subscripts obtained explain aspects of each group of variables; in this way the index 1, can be used to study in depth the economical data of a sample, the index 2 has relation with the employment situation, index 3 is related to sentimental relationships, index 4 and 5 are related with personal relations and factors of academic achievement, respectively.

The Bartlett's sphericity indicator is calculated on the simple character indices developed, giving as result a value of 77.23, indicating that the correlations matrix is the identity matrix of our variables. Owed to the fact that we have high values for X^2 , it can be confirmed that there exist significant correlations between variables for the established reliability margin. For the KMO indicator of partial correlation between a pair of variables, it was obtained a value of 0.528, which is slightly superior to the minimum recommended ($KMO \geq 0.5$), proving that the technique is appropriated.

The extraction of factors was made by the principal components method, where the factorial punctuations were estimated. With respect to the minimum number of factors, in accordance with the Keiser Rule, the self-values higher than unity, are the firsts two components that explain an accumulated variance percentage of 50.19%. The rotation of factors by the Quartimax method, allowed us to maximize the factorial charges variances and to achieve the required pertinence concentration (Table 5).

The values of the first component express the linear combinations of each one of the subscripts for the final index. The linear combination for the

compound index is:

$$D_{index} = 0.451(Economical\ data) + 0.265(Employment\ situation) \\ - 0.157(Sentimental\ relations) + 0.708(Personal\ relations) \\ + 0.787(Achievement\ factors)$$

The personal relations and achievement factors coefficients, indicate that both subscripts have a high importance for the general index. Economical data and employment situation are complementary, while sentimental relations indicate a negative value that can be justified because the set of these data is divided in two groups, considering the fact of the sentimental partner existence or its absence. As can be appreciated for both components in a rotated space, the subscripts are located mainly in the first and fourth quadrant, that is, the majority of subscripts are represented by the component 1, in the right side of the graphic (Figure 5).

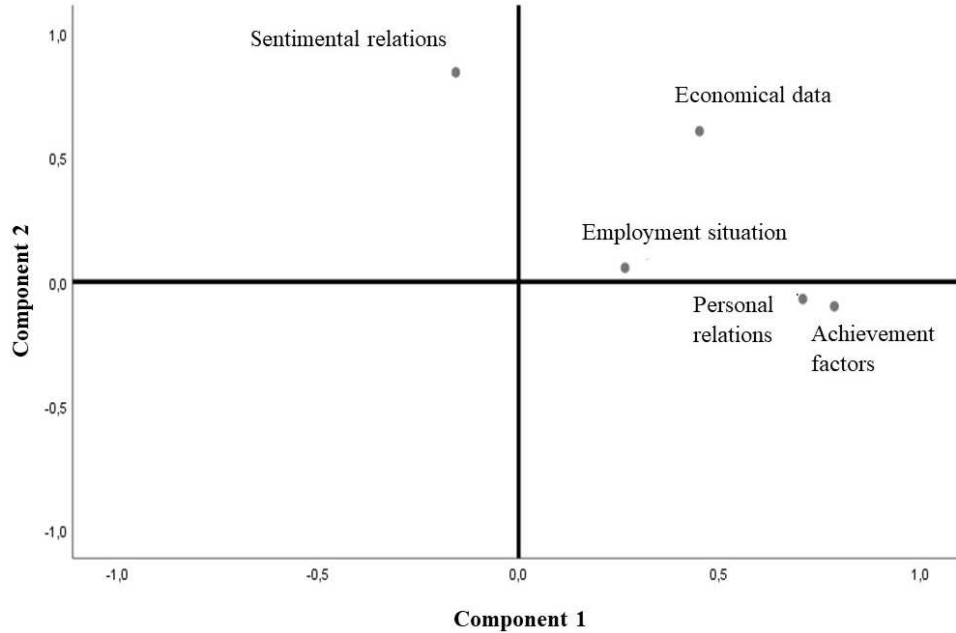


Fig. 5: Components in the rotated space

Later we proceeded to categorize the compound index by means of percentiles, were the respective categories will be: “low” for the individuals with

Tab. 6: Categorization of the compound index

N	Valid	482
	Lost	0
Percentiles	25	-.2091
	50	-.0158
	75	.1882

less or equal percentile of 25; “regular” for the individuals between the percentiles 25 to 50; “good” for individuals between the percentiles 50 to 75 and “excellent” for the individuals with percentiles higher than 75 (Table 6). The compound index was re-categorized in function of the label values and percentiles.

The validation of the technique was made by means of a factorial plane that allowed the comparison between the indexes of Table 5 and the scores’ averages. If we move to the right, in the last extreme is the higher average qualification, while that, by the left side with respect to the origin, we have the lowest values between 70 and 80 and finally, in the center we observe medium values of the scores’ average between 75 and 90 points (Figure 6).

This coincides with what the index describes, that is, that all subscripts that point to the right side are related with high scores’ averages: favorable economical data, favorable employment situation, positive personal relations and achievement factor. Personal relations have a significant influence in the index.

4 Discussion

The index developed in this work shows that the factors of academic achievement (subjects) and personal relations, have a higher factorial charge for the general index; that is, the subjects and good personal relations, positively influence in the academic performance. This result is in accordance with similar studies about the influence of the environment with which the student is interacting on the academic performance [27–29]. On the other hand, it was observed that the performance is moderately conditioned by economical data, that is, some students can have a high achievement, without their parents having made superior or university studies [14, 18, 30].

To have an academic index compounded by various dimensions or groups



The relations between variables observed by means of the correspondence analysis were verified by the alpha Cronbach, which for the majority of groups exceeded the minimum acceptable as mentioned by various authors [31,32]. The construction of the compound index from the factorial analysis technique was adequate because the minimum value of the KMO indicator was exceeded [33,34], also it was possible to define significant relations between the student's score average and the developed index, by the discriminant measures and the rotated space method.

5 Conclusions

The use of multivariate techniques as correspondence analysis and factorial analysis, simplified the work of the student's academic performance index construction; each technique played a fundamental role in the analysis of the collected information by means of the measuring instrument.

The analysis of multiple correspondences is an adequate technique as a previous step for the application of the factorial analysis, with the objective to reduce variables for achieving a better comprehension of factors that were used to construct the compound index. The use of the factorial analysis technique, allowed to construct efficiently the compound index and also the re-categorization of the academic performance in 4 levels.

The relation between the re-categorized index and general data (sex, age, semester, career, score's average) indicate that the index shows a significant relation with the score's average. Therefore, the index validation is demonstrated by the students' scores and this fact is proved in a descriptive way through the factorial plane and discriminant measures.

By means of the compound index developed, it can be evidenced that the student's performance is conditioned not only by the score, but by several factors that are related with the students economic situation and their personal relations, that is to say that, there are sceneries that must be treated with strategies to potentiate the academic performance results.

Acknowledgment

The National Secretariat of Superior Education, Science, Technology and Innovation (SENESCYT) is recognized as this research work's sponsoring entity, by its scholarship programs.

References

- [1] M. Zaffar, K.S. Savita, M.A. Hashmani, and S.S.H. Rizvi. A study of feature selection algorithms for predicting students academic performance. *Int. J. Adv. Comput. Sci. Appl.*, 9(5):541–549, 2018. doi: 10.14569/IJACSA.2018.090569.
- [2] S. Cardona, J. Vélez, and S. Tobón. Contribución de la evaluación socio-

- formativa al rendimiento académico en pregrado. *Educar*, 52(2):423–447, 2016. <https://racocat/index.php/Educar/article/view/311816>.
- [3] L.A. Morales Sánchez, V. Morales Sánchez, and S. Holguín Quiñones. Rendimiento escolar. *Rev. Elec. Humanidades, tecnología y Ciencia, del Instituto Politécnico Nacional*, 15:1–5, Jul-Dec 2016. http://revistaelectronica-ipn.org/Resources-Files/Contenido/16/HUMANIDADES_16_000382.pdf.
- [4] J. Gabalán Coello and F.E. Vásquez Rizo. Rendimiento académico universitario y asistencia a clases: Una visión. *Revista Educación*, 41(2):16–32, 2017. doi: 10.15517/revedu.v41i2.18477.
- [5] M. Vergel-Ortega, J.J. Martínez-Lozano, and S.L. Zafra-Tristancho. Factores asociados al rendimiento académico en adultos (factors associated with academic achievement in adults). *Revista científica*, 25(2):206–215, May-Aug 2016. doi: 10.14483/udistrital.jour.RC.2016.25.a4.
- [6] M.I. Barrios Gaxiola and M. Frías Armenta. Factores que influyen en el desarrollo y rendimiento escolar de los jóvenes de bachillerato. *Avaliação: Revista da Avaliação da Educação Superior*, 25(1):63–82, 2016. doi: 10.15446/rvp.v25n1.46921.
- [7] Alexander López Orozco, Elizabeth Genis Pérez, and Manuela Badillo Gaona. Incidencia del rendimiento escolar y sus factores asociados en el desarrollo económico. estudio de caso. In *Memoria del VI Congreso Red Internacional de Investigadores en Competitividad, la arquitectura financiera en las organizaciones para la competitividad*, ISBN 978-607-96203-0-1, volume 6, pages 1174–1189, Dic 2012. <https://riico.net/index.php/riico/article/view/478>.
- [8] L.M. Useche-Castro, J.C. Pérez-Parra, and V.E. Márquez-Pérez. Relación entre el índice de desarrollo de las TICs y el rendimiento académico medido mediante la prueba pisa. *Polo del Conocimiento: Revista científico-profesional*, 5(8):1030–1055, 2020. <https://dialnet.unirioja.es/servlet/articulo?codigo=7554385>.
- [9] R. Bajaña Nuñez, M. Quimis Zambrano, M. Sevilla Alarcón, L. Vicuña Monar, and J. Calderón Cisneros. Reducing feature selection bias using a model independent performance measure. *International Jour-*

- nal of Data Science*, 1(1):34–39, 2020. doi: 10.29076/issn.2602-8360vol1iss1.2017pp34-39p.
- [10] S. Díaz Cárdenas, K.A. Vergara, and M. Simancas-Pallares. Adición a internet y rendimiento académico de estudiantes de odontología. *Revista Colombiana de Psiquiatría*, 48(4):198–207, 2019. doi: 10.1016/j.rcp.2018.03.002.
- [11] C.R. Mejía, F.N. Ruiz-Urbina, D. Benites-Gamboa, and W. Pereda-Castro. Factores académicos asociados a la procrastinación. *Revista Cubana de Medicina General Integral*, 34(3):61–70, jul-sep 2018. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=89061>, <http://www.revmgi.sld.cu/index.php/mgi/article/view/954>.
- [12] A. Martínez Garza. *Diseños experimentales: métodos y elementos de teoría*. Trillas, 1988. ISBN 978-9682421556.
- [13] A. Beltrán Hernández, S. Rojas Sevilla, and M. Vertel Morinson. Análisis multivariado de la situación regional de la educación media en colombia (2015-2016): género, escuela y logro escolar. In *4to Encuentro Internacional de Investigación en Educación Matemática (EIEM)*, pages 75–81. Universidad del Atlántico, 2018. Repositorio Digital de Documentos en Educación Matemática, Universidad de los Andes, <http://funes.uniandes.edu.co/14298/>.
- [14] K. Lastre Meza, L.D. López Salazar, and C. Alcázar Berrío. Relación entre apoyo familiar y el rendimiento académico en estudiantes colombianos de educación primaria. *Psicogente*, 21(39):102–115, ene 2018. doi: 10.17081/psico.21.39.2825.
- [15] A.F. Aráuz. Aplicación del análisis factorial confirmatorio a un modelo de medición del rendimiento académico en lectura. *Revista de Ciencias Económicas*, 33(2):39–65, 2015. doi: 10.15517/rce.v33i2.22216.
- [16] L.J. Hayden, S.Y. Jeong, and C.A. Norton. An analysis of factors affecting mature age students’ academic success in undergraduate nursing programs: A critical literature review. *International Journal of Nursing Education Scholarship*, 13(1):127–138, Jan 2016. doi: 10.1515/ijnes-2015-0086.

-
- [17] N.G. Armenta, C.C. Pacheco, and Erika D.F. Pineda. Factores socioeconómicos que afectan el desempeño académico en los estudiantes universitarios de la facultad de ciencias humanas de la universidad autónoma de baja california. *Revista de investigación en psicología*, 11(1):153–165, 2008. doi: 10.15381/rinvp.v11i1.3888.
- [18] A.M. Cabrera Cruz and A. Alvarado Velarde. Diseño de un taller para la relación entre padre e hijos y su impacto en el rendimiento escolar en el nivel medio superior (cbtis no. 91 de cd. ixtepec, oaxaca). *EDUCATECONCIENCIA*, 11(12):205–220, 2016. <http://tecnocientifica.com.mx/educateconciencia/index.php/revistaeducate/article/view/213>.
- [19] A.R. Imlach, D.D. Ward, K.E. Stuart, M.J. Summers, M.J. Valenzuela, A.E. King, N.L. Saunders, J. Summers, V.K. Srikanth, and A. Robinson. Age is no barrier: predictors of academic success in older learners. *npj Science of Learning*, 2(13):1–7, nov 2017. doi: 10.1038/s41539-017-0014-5, <https://www.nature.com/articles/s41539-017-0014-5>.
- [20] R. Castro and M. Luna. Análisis exploratorio de datos geográficos voluntarios: estudio de caso en la administración zonal Manuela Sáenz de Quito. *Latin American Journal of Computing-LAJC*, 5(2):21–26, 2018. <https://lajc.epn.edu.ec/index.php/LAJC/article/view/151>.
- [21] A.T. Jebb, S. Parrigon, and S.E. Woo. Exploratory data analysis as a foundation of inductive research. *Human Resource Management Review*, 27(2):265–276, 2017. doi: 10.1016/j.hrmr.2016.08.003.
- [22] S. Das, R. Avelar, K. Dixon, and X. Sun. Investigation on the wrong way driving crash patterns using multiple correspondence analysis. *Accident Analysis Prevention Science*, 111:43–55, nov 2018. doi: 10.1016/j.aap.2017.11.016.
- [23] R.M. Lana, T.I.S. Riback, T.F.M. Lima, M. da Silva-Nunes, O.G. Cruz, F.G.S. Oliveira, G.G. Moresco, N.A. Honório, and C.T. Codeço. Socioeconomic and demographic characterization of an endemic malaria region in brazil by multiple correspondence analysis. *Journal of Applied Developmental Psychology*, 16(397):1–16, 2017. doi: 10.1186/s12936-017-2045-z,

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5625626/pdf/12936_2017_Article_2045.pdf.
- [24] A. Parchomenko, D. Nelen, J. Gillabel, and H. Rechberger. Measuring the circular economy-a multiple correspondence analysis of 63 metrics. *Journal of Cleaner Production*, 210:200–216, Feb 2019. doi: 10.1016/j.jclepro.2018.10.357.
 - [25] A.S. Agudelo Álzate, L.A. Figueroa Zapata, and L. Vásquez Benítez. Relaciones causales de los factores que afectan el desempeño de los estudiantes en pruebas estandarizadas en colombia). *Revista ESPACIOS*, 40(23):1–11, 2019. <http://www.revistaespacios.com/a19v40n23/a19v40n23p23.pdf>.
 - [26] E.P. Yildiz, M. Tezer, and H. Uzunboylu. Student opinion scale related to moodle lms in an online learning environment: Validity and reliability study. *International Journal of Interactive Mobile Technologies*, 12(4), 2018. <https://online-journals.org/index.php/i-jim/article/view/9205>.
 - [27] D. DeLay, L. Zhang, L.D. Hanish, C.F. Miller, R.A. Fabes, C.L. Martin, K.P. Kochel, and K.A. Updegraff. Peer influence on academic performance: A social network analysis of social-emotional intervention effects. *Prevention Science*, 17(8):903–913, 2016. doi: 10.1007/s11121-016-0678-8.
 - [28] H.Y. Kim, M.D. LaRusso, L.B. Hsin, A.G. Harbaugh, R.L. Selman, and C.E. Snow. Social perspective-taking performance: Construct, measurement, and relations with academic performance and engagement. *Journal of Applied Developmental Psychology*, 57:24–41, Jul-Aug 2018. doi: 10.1016/j.appdev.2018.05.005.
 - [29] J.V. Tomás-Miquel, D. Nicolau-Juliá, and M. Expósito-Langa. The social relations of university students: intensity, interaction and association with academic performance (las relaciones sociales de los estudiantes en la universidad: intensidad, interrelación y vinculación con el rendimiento académico). *Cultura y Educación*, 28(4):667–701, Oct 2016. doi: 10.1080/11356405.2016.1237340.
 - [30] B.E. Patiño Palma and L.Y. Apolinar-Joven. Nivel de escolaridad en padres y condición física relacionada con la salud en niños y adoles-

- centes (schooling level in parents and health-related fitness in children and adolescents). *Fisioterapia*, 43(2):68–75, Mar-Apr 2021. doi: 10.1016/j.ft.2020.09.002.
- [31] T. Caycho-Rodríguez, J. Ventura-León, C.H. García-Cadena, J.M. Tomás, J. Domínguez-Vergara, L. Daniel, and W.L. Arias-Gallegos. Evidencias psicométricas de una medida breve de resiliencia en adultos mayores peruanos no institucionalizados. *Psychosocial Intervention*, 27(2):73–79, ago 2018. doi: 10.5093/pi2018a6.
- [32] S.S. Sosa Palacios and E. Salas-Blas. Resiliencia y habilidades sociales en estudiantes secundarios de San Luis de Shuaro, La Merced. *Comuni@cción*, 11(1):40–50, ene-jun 2020. doi: 10.33595/2226-1478.11.1.394.
- [33] M.N. Anish, P.B. Dhanish, and R. Sridharan. Instrument for measuring service quality and dimensions of service quality affecting image, value and trust in medical tourism area. *International Journal of Services Operations Management*, 30(4):419–446, 2018. doi: 10.1504/IJ-SOM.2018.093511.
- [34] R.J.C. Yap, F. Komalasari, and I. Hadiansah. The effect of financial literacy and attitude on financial management behavior and satisfaction. *BISNIS BIROKRASI: Jurnal Ilmu Administrasi dan Organisasi*, 23(3), 2018. doi: 10.20476/jbb.v23i3.9175.

Appendix - Questionnaire for collection data

The following survey has the purely investigative objective of the construction of an academic index, by multivariant statistics in the Universidad Técnica de Manabí. We thank you in advance, your cooperation in answering the following questions related to the academic performance and the factors that influence it.

Instructions: You should mark your answer with an “X”.

GENERAL DATA

1. Sex: M ____ F ____ Age: ____
2. Career: _____

3. Semester: _____
4. Current average during the career: _____
5. Turn: Morning ____ Evening ____

ECONOMICAL DATA

1. You live with:
 - (a) Father ____
 - (b) Mother ____
 - (c) Both parents ____
 - (d) With your partner ____
 - (e) Alone ____
2. Which is the mother's scholar level ?
 - (a) Primary ____
 - (b) Secondary ____
 - (c) University ____
 - (d) Other ____
3. Which is the father's scholar level ?
 - (a) Primary ____
 - (b) Secondary ____
 - (c) University ____
 - (d) Other ____
4. Which is your monthly income?
 - (a) From \$0 to \$100 ____
 - (b) From \$300 to \$500 ____
 - (c) From \$500 to \$800 ____
 - (d) From \$800 to \$1000 ____
5. Which of the following services and apparatus have you at home?
 - (a) Computer ____
 - (b) Internet ____
 - (c) Television (Cable, Platforms) ____
 - (d) Printing machine ____
6. What type of transportation do you use to attend to the University?

- (a) Family's car _ _ _ _
- (b) Own car _ _ _ _
- (c) Taxi _ _ _ _
- (d) Bus _ _ _ _
- (e) Other (Specify) _ _ _ _

7. Do you work?

- (a) Yes _ _ _ _
- (b) No _ _ _ _

EMPLOYMENT SITUATION

8. How much time have you been working? (complete years)

- (a) From zero to 11 month _ _ _ _
- (b) From 1 to 2 years _ _ _ _
- (c) From 3 to 4 years _ _ _ _
- (d) More than 5 years _ _ _ _

9. How many hours per week do you work?

- (a) From 1 to 7 hours _ _ _ _
- (b) From 8 to 16 hours _ _ _ _
- (c) From 17 to 23 hours _ _ _ _
- (d) From 24 to 47 hours _ _ _ _
- (e) More than 48 hours _ _ _ _

10. How flexible is your work considering your student condition?

- (a) Very flexible _ _ _ _
- (b) Flexible _ _ _ _
- (c) Regular _ _ _ _
- (d) Not very flexible _ _ _ _
- (e) Not flexible at all _ _ _ _

11. How much do you think that your work affects your academic performance?

- (a) Very much _ _ _ _
- (b) Regular _ _ _ _
- (c) Little _ _ _ _
- (d) Nothing _ _ _ _

SENTIMENTAL RELATIONSHIPS

12. Have you a sentimental partner?
- (a) Yes ----
(b) No ----
13. How much time do you coexist with your partner weekly?
- (a) From 0 to 5 hours ----
(b) From 6 to 10 hours ----
(c) From 11 to 15 hours ----
(d) More than 15 hours ----
14. Does your partner study?
- (a) Yes ----
(b) No ----
15. Do you and your partner study together?
- (a) Yes ----
(b) No ----
16. Do you and partner work together?
- (a) Yes ----
(b) No ----

PERSONAL RELATIONSHIPS

17. How do you consider the relation you have with ...?

<i>Subjects</i>	<i>Very good</i>	<i>Good</i>	<i>Regular</i>	<i>Bad</i>	<i>Very bad</i>
Family					
Partner					
Friends					
Classmates					
Workmates					
Teachers					

ACADEMIC ACHIEVEMENT

18. How do you consider the following factors affect your academic performance?

<i>Subjects</i>	<i>Very positively</i>	<i>Positevely</i>	<i>It does not affect me</i>	<i>Negatively</i>	<i>Very negatively</i>
Family					
Partner					
Work					
Friends (outside university)					
Friends (inside university)					
Workmates					
Classmates					
Psico- pedagogical Department					
Tutorials					