# Teaching evaluation questionnaire validation at Escuela Politécnica Nacional, applying the method of Factor Analysis with extraction of principal components

## Validación del cuestionario de evaluación docente de la Escuela Politécnica Nacional, aplicando el método de Análisis Factorial con extracción de componentes principales

Tarquino Sánchez-Almeida[1], Iván Sandoval-Palis[2], Raquel Gilar-Corbi[3], Juan Castejón-Costa[4], and Diego Salazar-Orellana[5]

**ABSTRACT**

This work validates a teaching evaluation instrument applied to professors in engineering, sciences and higher technological level programs of the Escuela Politécnica Nacional, using the method of Factor Analysis with extraction of principal components. The database used for the research was previously examined and refined due to inconsistency, eg. outliers, out of range values, etc. The result of the method described above was a reduced survey of 15 items, which was obtained from an original study of 33 items. This new questionnaire clearly identifies the four main dimensions or aspects required: teaching development and planning, teacher-student relationship, evaluation, and a global assessment question. The reduction of the evaluation scale will allow to improve the process of integral teaching performance evaluation of the faculty at Escuela Politécnica Nacional, and this method could serve as a benchmark for the teaching evaluation process of other universities that belong to the higher education system of Ecuador.

**Keywords:** factor analysis, teaching evaluation, questionnaire validation, principal component analysis.

**RESUMEN**

Este trabajo valida un instrumento de evaluación docente aplicado a profesores de las carreras de ingeniería, ciencias y de nivel tecnológico superior de la Escuela Politécnica Nacional, utilizando el método de Análisis Factorial con extracción de componentes principales. La base de datos utilizada en la investigación fue examinada previamente y refinada por inconsistencia - por ejemplo, valores atípicos, valores fuera de rango, etc. El resultado del método descrito anteriormente fue una encuesta reducida de 15 ítems, que se obtuvo de un estudio original de 33 ítems. Este nuevo cuestionario identifica claramente las cuatro dimensiones o aspectos: planificación y desarrollo de la docencia; relación profesor-alumno; evaluación; y una pregunta de valoración global. La reducción de la escala de evaluación permitirá mejorar el proceso de la evaluación integral del desempeño docente del personal de la Escuela Politécnica Nacional, y este método podría servir de referencia para el proceso de evaluación de la enseñanza de otras universidades que pertenecen al sistema de educación superior del Ecuador.

**Palabras clave:** análisis factorial, evaluación docente, validación de cuestionario, análisis de componentes principales.

[1] Electronics and Telecommunications Engineer, Escuela Politécnica Nacional, Ecuador. MBA. Project Management, Escuela Politécnica Nacional, Ecuador. Affiliation: Full-Professor, Escuela Politécnica Nacional, Ecuador.
E-mail: tarquino.sanchez@epn.edu.ec
[2] Electrical Engineer, Escuela Politécnica Nacional, Ecuador. MSc. University Pedagogy, Escuela Politécnica Nacional, Ecuador. Affiliation: Full-Professor, Escuela Politécnica Nacional, Ecuador. E-mail: ivan.sandoval@epn.edu.ec
[3] Degree in Psychopedagogy, Universidad de Alicante, España. Teaching Diploma, Universidad de Alicante, España. PhD. Design, Orientation and Psychopedagogical Intervention, Universidad de Alicante, España. Affiliation: Associate-Professor, Universidad de Alicante, España. E-mail: raquel.gilar@ua.es
[4] Psychology Bachelor, Universidad de Valencia, España. PhD. in Psychology Universidad de Valencia, España. Affiliation: Full-Professor, Universidad de Alicante, España. E-mail: jl.castejon@ua.es
[5] Mathematician, Escuela Politécnica Nacional, Ecuador. Affiliation: Escuela Politécnica Nacional, Ecuador. E-mail: diego.salazar@epn.edu.ec

## Introduction

Over the years, innovation and the appearance of new lines of research have found use incorporating new areas of knowledge as instruments of academic training. That is where the application of instruction psychology is proposed as a new teaching staff tool. This has served as a guideline for the appearance of new research that seeks to holistically understand the teaching process within the methodology

and the best alternative to transmit knowledge in the classroom.

Aparicio (2014) indicated that it is possible to interpret learning as the existing relationship between communication and interaction where interaction is seen as part of the teaching and academic development. Therefore, university professors require specific skills that allow them to enhance the quality of the teaching-learning process in the classroom. These competences enable them to achieve excellence in terms of results, which involves an evaluation culture and control of the learning process.

The normally used instruments to measure students' evaluation of their teachers, programs, and satisfaction with their instruction are known as standard rating scales. However, research on student evaluation of teaching ratings has not yet provided clear answers to some questions about their validity (Hornstein, 2017; Marsh, 2007 a,b; Spooren, Brockx, and Mortelmans, 2013; Uttl, White, and Gonzalez, 2017).

From a statistical perspective, there exist are records in Ecuador regarding the teaching performance evaluation in universities, and the existing limited evidence is of restricted nature. Nowadays, the "Ley Orgánica de Educación Superior", the law that governs the Ecuadorian educational system, establishes in the article 151 that teachers will submit to an integral periodic evaluation according to the program and teaching scale regulations of the professors and researchers of the Higher Education System and the statutory norms of each institution within it, in exercise of its responsible autonomy. The survey carried out by the students about their teachers will be considered as one of the evaluation parameters (Consejo de Educación Superior, 2018).

The current assessment instruments were designed considering the components established in the program and teaching scale regulations of the professors and researchers of the Higher Education System, such as self-assessment, co-evaluation, and hetero-evaluation. Some of the items are taken from other SET rating scales, like the SEEQ (Marsh, 2007a), STERS (Toland and De Ayala, 2005), and SET37 (Mortelmans and Spooren, 2009), and are adapted to the characteristics of the Escuela Politécnica Nacional. In general, the technical validation of the evaluation instrument is not considered as a criterion to guarantee the quality of the application of the instrument. The integral assessment of teacher performance is an essential component that allows a professor to enroll as Assistant Professor or Associate Professor. The requirements include a qualification of at least 75% of the score in the performance evaluation during his last two academic periods. Additionally, according to article 96 of the regulation (Consejo de Educación Superior, 2017), the academic staff will be dismissed if they have obtained:

1) an integral evaluation performance of less than 60% for two consecutive times, and

1) four integral performance evaluations of less than 60% throughout their career.

In addition, it establishes that the main titular teachers will be promoted to the next level if they comply with other requirements such as having obtained a score of at least 80% in the performance evaluation of their last two academic periods (Consejo de Educación Superior, 2017).

The proposed methodology arises as a necessity to validate the instruments to evaluate the teaching staff at the Escuela Politécnica Nacional of Ecuador. This validation is applied to teachers of engineering, sciences and higher technological level programs, using the method of factor analysis with extraction of major components. This research considers the reliability and validity requirements that questionnaires must have with Likert opinion rating scales (Alaminos and Castejón Costa, 2006).

The most used method to extract the initial factors of the matrix of correlation observed variables is the principal component method. It is characterized by an analysis of the total variance of the set of observed variables. The purpose is to discover the main components that define this set. Both factor analysis and principal component analysis are multivariable data reduction techniques.

The main metric characteristics to determine the accuracy of an evaluation instrument (questionnaire) are reliability and validity. Reliability is the property that designates the constancy and precision of the results obtained by an instrument when applied on different occasions. On the other hand, validity refers to whether the instrument can measure what it is intended to measure (Carvajal, Centeno, Watson, Martínez, and Sanz Rubiales, 2011). Reliability can be estimated by four means: internal consistency, stability, equivalence and inter-judge harmony. The method of choice is internal consistency, which uses the Cronbach Alpha ($\alpha$) statistical test. The objective of this approach is to compare the variability of each item against the total variability of the instrument.

Currently, a line of work has been developed to reduce the length of scales already used or to elaborate new scales with a reduced number of items. The lack of time for their application , fatigue, and possible stereotyped responses in scales that are too long or that are part of a set that is applied within the same study, among others, has led to proposals of short scales (Gogol et al., 2014; Lafontaine et al., 2016). These scales have to be small enough to allow for a rapid assessment of purposed constructs, but large enough to ensure appropriate reliability, validity, and accurate parameter estimation.

The objectives of the present work are two: on one hand, to analyze the construct validity of the teaching-learning questionnaire, and on the other hand, to propose a reduction of that scale, conserving its psychometric properties.

Finally, the development of this research leads to improvements and appliance of new strategies for the teacher evaluation instrument. Additionally, these methods allow to identify the most relevant items and constructs. The result of this validation is the design of a questionnaire whose

application brings accurate information that will improve the quality of the Higher Education System of Ecuador.

## Mathematical Model

In factor analysis, a linear model is assumed:

$$X = \mu + LF + \varepsilon$$

where

$X(p \times 1)$ is the observable random vector, with mean vector $\mu$ and covariance matrix $\Sigma$;

$L(p \times m)$ is the matrix of factor loadings;

$F(m \times 1)$ are common factors, unobserved values of factors which describe major features of members of the population;

$\varepsilon(p \times 1)$ are error specific factors, measurement error and variation not accounted for by the common factors;

$\mu_i$ is the mean of variable $I$;

$\varepsilon_i$ is the ith specific factor;

$F_j$ is the jth common factor; and

$L_{ij}$ is the loading of the ith variable on the jth factor.

We assume that the unobservable random vectors $F$ and $\varepsilon$ satisfy the following conditions:

$F$ and $\varepsilon$ are independent;

$E(\varepsilon) = 0$ and $Cov(\varepsilon) = \Psi$; where $\Psi$, is a diagonal matrix; and $E(F) = 0$ and $Cov(F) = 1$.

Thus, the factors are assumed to be uncorrelated. This is called the orthogonal factor mode.

Factoring, given the model: $X = \mu + LF + \varepsilon$

The implied covariance structure for $X$ is,

$Cov(X) = LL' + \Psi$

If, $V(X_i) = l_{i1}^2 + \cdots + l_{im}^2 + \Psi_i$ and;

$Cov(X_{(i)}, \ldots, X_{(k)}) = l_{i1}^2 l_{k1}^2 + \cdots + l_{(im)}^2 l_{km}^2$

Furthermore, $Cov(X, F) = L$ so that $Cov(X_{(i)}, \ldots, F_{(j)}) = l_{(ij)}$

The portion of variance of the ith variable that is explained by the m common factors is called the communuality of the ith variable: $\sigma_{ii} = h_i^2 + \Psi_i$ where $\sigma_{ii}$ is the variance of $X_i$, i.e., the ith diagonal of $\Sigma$; $h_i^2 = (LL')_{ii} = l_{i1}^2 + \cdots + l_{(i2)}^2 + l_{(im)}^2$ is the communality of $X_i$; and $\Psi_i$ is the specific variance or uniqueness of $X_i$.

Note that the communality $h_i^2$ is the sum of squared loadings for $X_i$ (Harman, 1968).

In this case, thirty-three items or quantitative variables are presented, so the factor analysis technique is applied with the extraction method of main components to obtain two dummy variables that allow to relate and summarize the teaching staff survey. This allows to evaluate the relevant aspects of the teacher, within the teaching-learning process.

## Methodology

### Analysis of the original information regarding its relevance and validity

An exploratory analysis is made of the data obtained from the application of the evaluation instruments of 33 items with 5 answer choices (see Table 1), which were carried out by 6 110 students of the engineering, science and higher level technological programs for the professors of the Escuela Politécnica Nacional. These students were enrolled in 8 faculties and schools, studying 24 different degrees. The higher percentage of male students is representative of the population of students of polytechnic studies, in which 68,60% were male and 31,40% were female. The average age was 22,30 years old. These 6 110 students attended 1 380 different subjects which were distributed into 1 812 class-groups. The teacher sample consisted of 670 teachers, who represented a varied sample in terms of age, category, and teaching experience. More than half of these teachers were male (62,80%). The application of the scale of 33 items was carried out at the end of semester 2017-A (October 2017-March 2018), before the students knew their final grades. All teachers were evaluated by the students in the same term. All students had to evaluate the teachers to be able to access their final grades. The student teaching evaluation was conducted through an electronic platform, obtaining 19 527 records (original data matrix) in which the data were recorded (the same student was able to evaluate several professors since he/she took several subjects).

From the **original data matrix**, a correlation matrix is elaborated between all the considered variables (items). Several tests are carried out to determine if it is pertinent, from a statistical point of view, to carry out factor analysis with the information available from the correlation matrix. The main tests are:

*The Bartlett sphericity test*: it is based on chi-square distribution, where high values lead to rejecting the null hypothesis (H$_0$) that states that the variables are not correlated within the population. Thus, Bartlett's test of sphericity determines whether the correlation matrix is an identity matrix, which would indicate that the factorial model is inadequate. If the significance value (p-value) is less than 0,050, we reject the null hypothesis (H$_0$) and continue with the factor analysis.

*The Kaiser-Meyer-Olkin Index* (*KMO*): it allows the comparison between the magnitude of the observed correlation coefficients and the magnitude of the partial correlation coefficients. The KMO statistic varies between 0 and 1. Those less than 0,500 indicate that factor analysis is not required for the data in question.

*The partial correlation coefficient:* it describes the linear relationship between two variables while controlling the effects of one or more additional variables. These coefficients should tend to zero, when they are lent for factor analysis (Montoya O. 2005).

**Table 1.** Evaluation instrument of 33 items in 5 constructs

| N° | QUESTION | N° | ANSWER |
|---|---|---|---|
| **I** | **DIDACTICS** | **IV** | **EVALUATION CRITERIA** |
| 1 | Did the teacher clearly explain the objectives and themes, indicating their interrelation and contribution to professional profile? | 17 | Has the teacher used objective methods to evaluate students? |
| 2 | Did the teacher select class activities appropriately, depending on the objectives? | 18 | Has the evaluation been used to reorient student learning? |
| 3 | Has the teacher been clear in his/her explanations and presentations? | 19 | Has the teacher considered aspects that are not merely cognitive? |
| 4 | Has the teacher related theoretical fundamental concepts and principles with practice? | 20 | Does the teacher evaluate fairly and impartially? |
| 5 | Does the teacher solve the difficulties that arise? | 21 | Has the minimum grade to approve the course been explained and why? |
| 6 | Does the teacher show the mastery of the subject? | 22 | Were the objectives defined in a clear and concise form? |
| 7 | Does the lecturer demonstrate planning his/her lectures before the class presentations? | 23 | Are the evaluation events related to the teaching taught? |
| 8 | Is the teacher creative and dynamic in the classroom? | **V** | **TEACHER-STUDENT RELATIONSHIP** |
| 9 | Does the professor show that he/she is up-to-date in the subject that is imparted? | 24 | Did the teacher ascertain that the students understand what they were being taught? |
| **II** | **RESOURCES** | 25 | Did the teacher encourage the initiatives coming from the students? |
| 10 | Does the teacher prepare didactic material additional to the textbook and makes it known? | 26 | Did the teacher create an environment of participation? |
| 11 | Does he/she organize didactic experiences such as visits, excursions, projects, discussions? | 27 | Did the teacher maintain a cordial relationship with the entire group of students? |
| 12 | Has the complementary, recommended or used material been interesting? | 28 | Did the teacher create an environment of trust and work during class? |
| 13 | Does he/she use means that benefit the learning process? | 29 | Has the teacher motivated students and increased their interest about the subject? |
| **III** | **METHODOLOGY** | 30 | Does the teacher have an attitude of availability outside the class? |
| 14 | Did the teacher use different teaching methods properly? | 31 | Has the teacher openly accepted the suggestions made by students? |
| 15 | Has the teacher used a varied methodology? | 32 | Was the teacher attentive with the evolution of the students? |
| 16 | Has the teacher explained the methodologies for evaluating the course? | 33 | Excluding limitations that are not due to the teacher, could he/she be considered as a good teacher? |

**Source:** Author

## Extraction of Main Components

Interpretation of the main components is often difficult, so the initial extraction is rotated to achieve a solution that facilitates it. Varimax with Kaiser Normalization (Kaiser, 1958) is the rotation method that uses the orthogonal rotation of factors previously normalized. In other words, it maintains the independence between the rotated factors. This method achieves that each rotated component presents correlations with only a few variables. Therefore, this method minimizes the number of variables with high loads by one factor and is adequate when the number of components is reduced.

## Results

*Statistical analysis of teacher evaluation instruments*

Bartlett's sphericity test was applied before using the multivariate factor analysis technique in order to verify if the correlation matrix is an identity matrix, which means that the correlations between the variables are zeros. The test consists of an estimation of the chi-square indicator, where high values lead to rejecting the null hypothesis. The test must have a significance value lower than the 0,050 limit, which would indicate that the variables are not correlated within the population. Table 3 shows the result of the Bartlett's Sphericity test that is 0,000. This demonstrates that the null hypothesis is rejected. Therefore, factor analysis is applicable in this case.

The analysis tool that was used was the Kaiser-Meyer-Olkin test (KMO). It is an index that compares the magnitude of the correlation coefficients observed with the magnitude of the partial correlation coefficients, eliminating the effect of the remaining variables included in the analysis. Since the partial correlation between two variables must be small when the factorial model is adequate, the denominator must increase a little compared to the magnitude of the correlation coefficients observed if the data corresponds to a factorial structure, in which case KMO will have a value close to 1.

Table 3 shows the result of the KMO test using the SPSS statistical analysis software, which has a value of 0,990, very close to the unit and therefore fulfills the requirement.

**Table 2.** KMO and Bartlett Tests

| KMO and Bartlett tests | | |
|---|---|---|
| Kaiser-Meyer-Olkin measure of sampling adequacy | | 0,990 |
| | Chi-square Aprox. | 338 959,305 |
| Bartlett's sphericity test | gl | 528 |
| | Sig. | 0,000 |

**Source:** Authors (data analysis performed by SPSS v.22)

The partial conclusion that can be reached about this first part is that the two types of analysis on the pertinence and validity of the data matrix are satisfactorily verified.

Now, we proceed with the second part, which consists of extracting the principal components by grouping the 33 items or original variables into new variables called "factors". It is based on an exploratory analysis and shows that there is a large number of stereotyped responses, defined as those

in which students respond with a single type of score along the whole scale, be it 1, 2, 3, 4 or 5. The data from these students is eliminated and, finally, the number of records on which the analysis is based is 15 771.
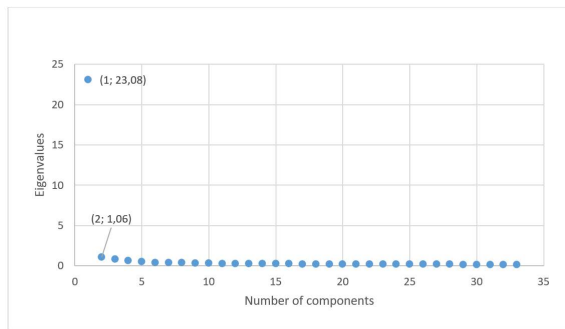


**Figure 1.** Sedimentation chart
**Source:** Authors

The results of the factor analysis of the sample reveal the existence of two factors, dimensions, or different constructs, as can be seen in the sedimentation chart in Figure I. These are chosen when the components have eigenvalues greater than 1.

The total variance explained in Table 3 analyzes in detail the selection of the two components, factors, or constructs: factor 1 explains 70% of the variation in the scores of the scale, and factor 2, 3,20%. Only the first two factors have eigenvalues greater than 1 and explain 73,20% of the original problem, resulting in a loss of 26,80% of the original information due to the fact that the survey has a very high number of items, among other aspects.

Factor 1 is composed of items 17 to 33. All the items have high saturations between them and the factor (0,780 to 0,680). This is related to the factor or set of items. The items with higher load or saturations are, in this order, items 32, 28, 31, 30, 25, 26, 24, 29, 27, 33, 22, and 20, to the less representative 23, 18, 19, 21 and 17.

Given that all these items refer to the teacher-student relationship, this factor can be called *Teacher-Student Relationship* and *establishment of a good learning environment*.

*Extraction Method: Principal Component Analysis*

**Rotation Method: Varimax with Kaiser Normalization**

Factor 2 is composed of items 1 to 16. All the items have saturations or high relationships between each one of them and the factor (0,761 to 0,670); the items with higher loads or saturations are, in that order, the items 3, 4, 5, 2, 6, 7, 1, 9, and 8. In addition, these items refer to what may be called *Planning, mastery, and clarity in the explanation of the subject*.

Given that Factor 1, displays a greater variance percentage than Factor 2, this indicates that the students of the Escuela Politécnica Nacional give the greatest importance to the teacher-student relationship, or, in other words, perform the

**Table 3.** Factor loadings and total variance explained

| Rotated Component Matrix | Loadings | |
| --- | --- | --- |
| | Factor 1 | Factor 2 |
| Item32 | **,782** | ,407 |
| Item28 | **,780** | ,398 |
| Item31 | **,789** | ,405 |
| Item30 | **,787** | ,407 |
| Item25 | **,784** | ,414 |
| Item26 | **,770** | ,405 |
| Item24 | **,767** | ,427 |
| Item29 | **,759** | ,442 |
| Item27 | **,759** | ,467 |
| Item33 | **,757** | ,427 |
| Item22 | **,727** | ,494 |
| Item20 | **,723** | ,470 |
| Item23 | **,723** | ,490 |
| Item18 | **,722** | ,494 |
| Item19 | **,712** | ,455 |
| Item21 | **,705** | ,483 |
| Item17 | **,689** | ,507 |
| Item15 | ,643 | **,697** |
| Item14 | ,643 | **,695** |
| Item16 | ,622 | **,690** |
| Item13 | ,618 | **,686** |
| Item12 | ,597 | **,680** |
| Item11 | ,525 | **,673** |
| Item2 | ,399 | **,759** |
| Item4 | ,430 | **,760** |
| Item3 | ,415 | **,761** |
| Item7 | ,442 | **,756** |
| Item5 | ,453 | **,760** |
| Item1 | ,369 | **,755** |
| Item6 | ,417 | **,758** |
| Item9 | ,454 | **,747** |
| Item8 | ,506 | **,699** |
| Item10 | ,547 | **,670** |
| Initial eigenvalues | 23,100 | 1,100 |
| % of Total Variance | 70 | 3,200 |
| Total Variance | | 73,20% |
| Sum of charges to the square of the extraction | 23,100 | 1,100 |
| % of Total Variance | 70 | 3,200 |
| Total Variance | | 73,20% |
| Sum of charges to the square of the rotation | 13,300 | 10,800 |
| % of Total Variance | 40,400 | 32,800 |
| Total Variance | | 73,20% |

**Source:** Authors

assessment of the teacher depending on the quality of this relationship, to a greater extent than the aspect of *Planning, mastery and clarity in the explanation of the subject*.

Another requirement that any questionnaire or rating scale must meet is reliability. If all the items amount to or contribute to measure the same, the reliability will be high. As indicated above, the most used statistical tool to calculate reliability is Cronbach's Alpha ($\alpha$) internal consistency coefficient. It evinces an adequate reliability when $\alpha$ values range from 0,650 to high values such as 0,800 and above.

To do this, the reliability of each of the factors obtained in the factorial analysis was calculated using Cronbach's $\alpha$ internal consistency coefficient; being $\alpha = 0,970$ the reliability of Factor 1, and Factor 2, $\alpha = 0,950$. I was very high in both cases.

Given that it is possible that both factors or aspects are related, the total reliability of the 33 item scale, that amounted to $\alpha = 0,980$, was obtained. This implies that a total score of the scale can be obtained, as well as scores for each of the previous factors or sub-scales.

Once it is confirmed that the reliability of each of the sub-scales is very high, it is possible to determine which item contributes more to the reliability of the scale and which items are redundant. Moreover, these can be eliminated without decreasing the reliability of the scale.

The sub-scale/*Factor 1*, composed of 17 items, has a reliability $\alpha = 0,970$. If redundant items are eliminated, this scale can be reduced to 6 items: 26, 28, 30, 31, 32, and 33, with reliability $\alpha = 0,950$, which is still very high.

The sub-scale/*Factor 2*, composed of 16 items has a reliability $\alpha = 0,950$. If redundant items are eliminated, this scale can be reduced to 5 items: 3, 4, 5, 6, and 7, with reliability $\alpha = 0,930$, which remains considerably high.

Thus, the 33-item questionnaire can be reduced to about 11 items without loss of validity or reliability $\alpha = 0,960$, and with practically the same informative value as the original evaluation instrument . These items would be, as indicated in Table 4:

Factor 1/scale 1: Items 26, 28, 30, 31, 32, and 33. 27 and 29 could also be included in this order, with reliability $\alpha = 0,960$.

Factor 2/scale 2: Items 3, 4, 5, 6, and 7. We could also include 9, with $\alpha = 0,940$ and 2, with $\alpha = 0,940$ in this order.

**Table 4.** Reduced questionnaire of 11 items

| Factor | Items | ($\alpha$) Cronbach | Additional Items | ($\alpha$) Cronbach |
|--------|-------|---------------------|------------------|---------------------|
| 1 | 26, 28, 30, 31, 32, 33 | 0,979 | 27,290 | 0,961 |
| 2 | 3, 4, 5, 6, 7 | 0,959 | 9,200 | 0,945, 0,944 |

**Source:** Authors

Then, the question that arises is: What happens with the other items and with the other theoretical aspects included in the scale as *Resources, Methodology and Evaluation*?

The answer explains that they contribute very little to the assessment of the teaching staff, given what the selected items of the reduced scale do.

However, the dimensions or aspects related to *Resources, Methodology and Evaluation* are important enough to be included in the scale, for which it is necessary to incorporate items that better represent these dimensions than the previous scale of 33 questions.

Therefore, a factor analysis was carried out to determine the extent to which the dimensions or aspects of *Resources, Methodology and Evaluation* influence the results, thus forcing the appearance of four factors, out of which two are new: *Evaluation, Methodology-Resources* and -two factors that had previously appeared- *Teacher-student relationship, and Planning, mastery and clarity in the subject's explanation*. Table 5 summarizes the variance parameters, factor loadings, as well as the Cronbach's Alpha internal consistency coefficients ($\alpha$) for each aspect.

**Table 5.** Factorial analysis forced to 4 factors

| Factor | Variance | Items | Factor loadings | ($\alpha$) Cronbach | Appearance |
|--------|----------|-------|-----------------|---------------------|------------|
| 1 | 70,98% | 24-33 | ,720-,620 | ,970 | Teacher-student re-lationship |
| 2 | 3,72% | 1-9 | ,730-,580 | ,960 | Planning, mastery and clarity in the subject's explanation |
| 3 | 2,85% | 16-23 | ,680-,460 | ,950 | Evaluation |
| 4 | 2,28% | 10-15 | ,770-,600 | ,950 | Methodology-resources |

**Source:** Authors

## Proposal of a reduced scale

Based on the information displayed in section A, it is considered convenient to better define the items on the *Evaluation* and *Methodology-Resources* aspects. To do this, 15 items are proposed, since they commonly appear in most universities (Alaminos and Castejón, 2006). This scale could include the most effective items of the original questionnaire, along with some new items introduced from other questionnaires, based on the theoretical dimensions of the aspects that are to be measured (Casero, 2008).

The analysis of the data obtained is structured in four aspects or dimensions: *Teaching Development and Planning, Teacher-Student Relationship, Evaluation* and a *Global Assessment question*, as indicated in Table 6, which shows each question with subscripts that express the following information:

1 = combined items of the aspects in the original questionnaire.

2 = relevant items of the original questionnaire.

3 = New items included

The analysis of the proposed reduced scale observed in Table 6 indicates that two of the included items, related

**Table 6.** Reduced scale proposal

| Aspect | N | Question |
|---|---|---|
| Planning and Development of Teaching | 1 | Does the teacher present and explain at the beginning of the period the contents (syllabus), methodologies and teaching activities, evaluation system, presentation of works, etc.?[1,2] |
| | 2 | Does the teacher demonstrate that his classes are based on the learning objectives and the syllabus of the subject?[1] |
| | 3 | Does the teacher demonstrate that he prepares and plans his classes (activities, methodologies, resources, evaluation, etc.)?[1] |
| | 4 | Does the teacher demonstrate mastery of the topics discussed in class?[1,2] |
| | 5 | Is the teacher clear in his expositions and explanations? Are the taught topics understood?[1,2] |
| | 6 | Does the teacher meet the established schedule?[1] |
| | 7 | Does the teacher use different didactic resources (for example, books, posters, maps, photos, slides, articles, videos, software, etc.) as support for the teaching of the subject? The resources and materials used or recommended are useful to take the course (bibliography, slides, virtual campus material, etc.)?[3] |
| | 8 | Does the methodology used by the teacher facilitate the learning of the subject and encourage interest in it?[1] |
| Teacher–Student Relationship | 9 | Has the teacher created an environment of trust and work in class?[2] |
| | 10 | Is the teacher accessible and willing to attend out of class consultations?[1,2] |
| | 11 | Has the teacher been made suggestions that he openly accepted? Has the teacher created an atmosphere of class participation?[2] |
| Evaluation | 12 | Are the evaluation events related to the topics presented in the course? Is the evaluation adjusted to the contents studied during the course?[3] |
| | 13 | Does the professor respect the weighing established by the institution that no evaluation must exceed 40% of the total score?[1] |
| | 14 | Does the teacher comply with the review of tests and/or previous exams the record of grades?[1] |
| | 15 | Does the teacher make the correction of the evaluations to the students? In the exams and work we the students have the possibility to know the mistakes made and comment on their valuation?[3] |
| Global Valuation | 15/16 | Excluding limitations that are not due to the teacher, can he/she be considered a good teacher?[2] |

**Source:** Author

to the *Teacher-student relationship,* were the same ones as in the original questionnaire because they provide relevant information about the evaluation to the professor. In addition to this table, there are items, such as 7, 12, and 15 that include two alternatives, which coincide in their meaning, but are expressed differently. Therefore, for those questions, it is necessary to choose between the options when applying the new survey.

From the results, two possible options are proposed for item 11, which are items 31 and 26 of the original questionnaires, offering the possibility of choosing between one or the other, since both alternatives have practically the same importance.

Finally, regarding item 16, which is related to the *Global Assessment* is included in the questionnaire to evaluate the general performance of the professor. However, as it not a relevant aspect, it can be considered as a replacement to item 15.

## Discussion and conclusions

The first objectives of the present work were to analyze the construct validity of the teaching-learning questionnaire. Factor analysis revealed that the scale was composed of two factors. However, when factor analysis was forced to 4 factors, the theoretical structure of the initial questionnaire was exactly reproduced.

The second objective of the present work was to propose a reduction of the teaching evaluation questionnaire. It is difficult to reduce a questionnaire while maintaining the fundamental aspects of teaching. However, if the objective is to reduce the questionnaire even further to condense it to 13 items, for example, it is recommended to eliminate item 2, which covers the *Planning* aspect, as well as items 14 and 15 that refer to the grading methodology. In addition, it would be optimal to eliminate item 16. These changes are proposed while taking into that the questionnaire would maintain the desired margin of reliability.

For the validation of the reduced questionnaire proposed in Table 6, the data obtained from a large sample would be subjected to the same analysis, along with other techniques such as Confirmatory Factor Analysis and Item Response Theory Analysis (TRI).

The items with the highest saturations are those that best define the factor, while the items with low saturations define the factor less accurately. Based on this, for the original questionnaire of 33 items, Factor 1 has a high level of saturation -within the range of 0,780 to 0,689- and determines a positive teacher-student relationship, as well as a good learning environment. Similarly, Factor 2 has a high saturation level and describes the planning, mastery and clarity in the explanation of the subject, leaving the rest with low saturation levels.

Based on reliability tests with Cronbach internal consistency coefficients ($\alpha$) and Bartlett's sphericity test, it is concluded that the two types of analysis about the relevance and validity of the matrix data are satisfactorily verified, which means that the original matrix data is reliable. In addition, all the questions have relevant information for the analysis of communalities.

The results obtained satisfy all the objectives established in this research paper and offer a proposal for a tool used for student evaluation of the university's teaching staff, based on the opinions of lecturers and students. The contribution that this work aims is to do is to present an available instrument to be used by universities and polytechnic schools, especially at the Escuela Politécnica Nacional, to validate and reduce the teaching evaluation questionnaires. The positive results of this study confirm it is possible to enter to a new phase for teaching evaluation using a new and well-defined survey.

A limitation of the study is that the assumption of randomness for factor analysis was not followed, because the questions are not arranged in a random order. On the other hand, another limitation is that the construct validity was examined but not the criterion of validity,

for example, correlating the questionnaire scores to some external criterion.

In addition to the validation analysis of the teacher evaluation instrument, it is recommended to carry out a multi-dimensional analysis including aspects of gender, academic record, admission examination score, subjects, degrees, among others, in order to relate the scores in the scales to other variables and their correlations.

## Acknowledgements

## References

Alaminos, A. and Castejón, J. L. (2006). *Elaboración, análisis e interpretación de encuestas, cuestionarios y escalas de opinión*. Alicante: Universidad de Alicante. http://hdl.handle.net/10045/20331

Aparicio, E. (2014). *Validación de un cuestionario de evaluación de la docencia universitaria*. (Doctoral thesis, Universidad de Alicante, Alicante, Spain). http://hdl.handle.net/10045/45168

Carvajal, A., Centeno, C., Watson, R., Martínez, M., and Sanz Rubiales, Á. (2011). ¿Cómo validar un instrumento de medida de la salud?. *Anales del Sistema Sanitario de Navarra 34*(1), 63-72. 10.4321/S1137-66272011000100007

Casero Martínez, A. (2008). Propuesta de un cuestionario de evaluación de la calidad docente universitaria consensuado entre alumnos y profesores. *Revista de Investigación educativa 26*(1), 25-44.

Consejo de Educación Superior (2017). *Reglamento de Carrera y Escalafón del Profesor e Investigador del Sistema de Educación Superior*. https://n9.cl/4scm

Consejo de Educación Superior (2018). *Ley Organica De Educacion Superior*, LOES. http://www.ces.gob.ec/documentos/Normativa/LOES.pdf

Gogol, K, Bunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischhach, A., and Preckel, F. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology 39*, 188-205. 10.1016/j.cedpsych.2014.04.002

Harman, H. (1968) *Modern Factor Analysis*, (2nd Ed.) Chicago: The University of Chicago Press, Ltd. https://archive.org/details/ModernFactorAnalysis/mode/2up

Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education 4*, 1. 10.1080/2331186X.2017.1304016

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika 23*(3), 187-200. 10.1007/BF02289233

Lafontaine, M.-F., Brassard, A., Lussier, Y., Valois, P., Shaver, P. R., and Johnson, S. M. (2016). Selecting the best items for a short-form of the experiencies in close relationships questionnaire. *European Journal of Psychological Assessment 32*(1), 140-154. 10.1027/1015-5759/a000243

Marsh, H. W. (2007a). *Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness*. In R. P. Perry and J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence based perspective* (pp. 319-383). New York: Springer. 10.1007/1-4020-5742-3_9

Marsh, H. W. (2007b). Do university teachers become more effective with experience? A multilevel growth model of students' evaluation of teaching over 13 years. *Journal of Educational Psychology 99*, 775- 790. 10.1037/0022-0663.99.4.775

Montoya, O. (2015). Aplicación del análisis factorial a la investigación de mercados. *Scientia Et Technica Scientia et Technica*, Año XIII 35, 281-286.

Mortelmans, D. and Spooren, P. (2009). A revalidation of the SET37 questionnaire for student evaluations of teaching. *Educational Studies - EDUC STUD.* 35. 547-552.

Mortelmans, D. and Spooren, P. (2009). A revalidation of the SET37 questionnaire for student evaluations of teaching. *Educational Studies - EDUC STUD*, 35, 547-552. 10.1080/03055690902880299

Spooren, P., Brockx, B., and Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. *Review of Educational Research 83*(4), 1-45. 10.3102/0034654313496870

Toland, M. D., and De Ayala, R. J. (2005). A Multilevel Factor Analysis of Students' Evaluations of Teaching. *Educational and Psychological Measurement 65*(1), 272–296. 10.1177/0013164404268667

Uttl, B., White, C. A., and Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation 54*, 22-42. 10.1016/j.stueduc.2016.08.007