Journal of Applied Statistics
Vol. 00, No. 00, January 2009, 1–12

Determinants of students' evaluation teaching according to their performance: an approach based on the relative importance metric

Vincenza Capursi[a] and Clara Romano[a*]
[a]*Dipartimento di Scienze Statistiche e Matematiche 'Vianelli', Università di Palermo, Italy*

(*Received 00 Month 200x; in final form 00 Month 200x*)

Student Evaluation of Teaching (SET) is an important tool to monitor teaching quality. In Italy, the SET is performed through the analysis of students' opinion who fill out a questionnaire including a set of items related the teacher's characteristics, the logistics, the organization of courses and the overall satisfaction on teaching quality.

In this paper, we want to give simple statistical tools to construct an indicator of teaching quality according to student's performance and to more important items which can influenced the student's satisfaction. To build an indicator of student's performance, we considered two variables, age and UEC (University Educational Credits) that influence significantly the student performance. Combining these two variables we obtained the ISP (Indicator of Student Performance). To investigate which items are more important, we use a relative importance metric based on Proportional Marginal Variance Decomposition (PMVD). PMVD metric allow us to overcome the problem of decomposition of variance analysis in an empirical study where the covariates are correlated.

Keywords: Student Evaluation Teaching; performance's indicator; relative importance measure; PMVD

1. Introduction

During the last few decades Student Evaluation of Teaching (SET) has been considered as an important tool in the improvement of teaching quality even if Marsh [27] and Wachtel [32] report that student evaluation programs were introduced at Harvard in 1915, and the first studies on SET effectiveness were written in the 1920s by Remmers [29–31]. Student evaluation research had a wide development in the decade 1970-1980, when much research was devoted to the utility and validity of student ratings of instruction [9]. Kulik [21] states that the initial aim of SET served two goals: mapping the quality of teaching in universities, and providing information and help to instructors in order to improve their teaching. For Marsh [27] students ratings are also very useful to make administrative decisions and to satisfy a fundamental principle of the evaluation: the accountability. Although the implementation of SET was spread in many faculties, a lot of univerties put up resistence to the use and the utility of these ratings. Supporters argue that evaluative judgements have a strong positive influence on the improvement of instructional skills. Marsh [26] states opinions about the role of SET vary from "reliable, valid and useful" to "unreliable, invalid and useless".

In Italy, the MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca) introduced, in 1999, as obligatory norm, the teaching evaluation in the Italian

[*] Corresponding author. Email: romano.clara@unipa.it

ISSN: 0233-1934 print/ISSN 1360-0532 online

DOI: 10.1080/0266476YYxxxxxxxx
http://www.informaworld.com

universities. The SET is performed through the analysis students' opinion who fill out a questionnaire including a set of items related to different dimensions: the teacher's characteristics, the logistics, the organization of the degree course and the overall satisfaction on teaching quality.

The literature on SET provides several studies that highlight the role of confounding variables [9, 27]. It seems that students background characteristics can explain a relevant portion of variability of evaluation items. On the other hand, teaching evaluation can be influenced by course or teacher characteristics that are not indicators of teaching quality, such as course difficulty, class size, grading leniency, teacher popularity, and so forth.

An usual measure of the teaching quality is a composite indicator with no attempt made to measure which variables, latent or manifest, determine the value of it. Statistical models more complex and useful, as Multilevel and Rasch models [3, 4, 15, 16], are used in literature for measuring the teaching quality, but in our case the primary objective is to give simple statistical tools to construct an indicator of teaching quality according to student's performance and to more important items which determine the student's satisfaction. In fact, we want to verify if teaching evaluation is conditioned by student's career. An important issue to construct a composite indicator is the weighting of simple indicators because all the dimensions (or items) are not equally important [2]. In our case, we want to investigate which items are more important to explain the overall satisfaction of teaching. To reach this aim we use a linear regression where the overall satisfaction is the response variable and the other items are the covariates. But, our data come, like all this kind of data, from observational study, where the covariates are usually correlated. Therefore, we cannot find the weight of each item by the usual decomposition of variance analysis. So, to overcome this drawback we use a relative importance metric based on Proportional Marginal Variance Decomposition (PMVD), introduced by Feldman [12].

The paper is organized as following. Section 2 deals with the statistical methodology used: relative important metric PMVD. In Section 3 we present the Indicator of Student Performance (*ISP*) and we construct a statistical test to test if the relative importance measures of the two set of students show a significative statistical difference. In Section 4 the data for application are described and the results are presented. Finally, Section 5 includes a discussion on substantive implications of findings.

2. Methodology

2.1 *Relative importance metric PMVD*

Weighting techniques based on a multiple regression model are widely used because of the numerous advantages that such techniques involve, like the possibility to determine the weight of the single simple indicators [28]. When regressors are uncorrelated each regressor's contribution is just the R^2 from univariate regression, and all univariate R^2-values add up to the full model R^2. But, when data come from observational studies, the covariates are usually correlated and such techniques are not appropriate because it is not simple to break down model R^2 into shares from the individual regressors. Let consider the linear regression model

$$Y = \beta_0 + X_1\beta_1 + ... + X_n\beta_n + \epsilon \qquad (1)$$

where random variables $X_i, i = 1, ..., n,$ denote n regressor variables and ϵ denotes an error term with expectation 0 and variance σ^2. This model implies the conditional moments $E(Y|X_1, ..., X_n) = \beta_0 + X_1\beta_1 + ... + X_n\beta_n$ and $var(Y|X_1, ..., X_n) = var(\epsilon|X_1, ..., X_n) = \sigma^2$. The marginal variance model is

$$var(Y) = \sum_{i=1}^{n} \beta_j^2 v_j + \sum_{i=1}^{n-1} \sum_{i+1}^{n} \beta_i \beta_k \sqrt{v_i v_k} \rho_{ik} + \sigma^2. \qquad (2)$$

Is X's are uncorrelated the explained variance decomposes into the contribution $\beta_i^2 v_i$ ($v_i = var(X_i)$), which can be consistlently estimated using the unique sum of squares for each regressor. If X's are correlated, it is no obvious how $var(Y)$ should be decomposed.

Some scholars have proposed analytical procedures able to underline the relative importance of each variable within a regressive model [14]. The various suggestions formulated, nevertheless, have not found unanimous agreement because of the different results reached in presence of correlation between the regressors. Solutions to this issue are proposed in literature by means of relative importance metrics for R^2 decomposition [12, 13, 24].

The metrics more used in literature are LMG (Lindeman Merenda Gold) and PMVD (Proportional Marginal Variance Decomposition). Both these metrics decompose R^2 into non-negative contributions that automatically sum to the total R^2. This is an advantage they have over all simple metrics.

The difficulty in decomposing R^2 for regression model with correlated regressors lies in the fact that each order of regressors yields a different decomposition of the model sum of square [1]. Generally the regressors enter into the model in the order they are listed. Sometimes, some researchers apply stepwise regression and decompose R^2 based on the order obtained by this automatic approach.

The approach taken by the metrics LMG [24] and PMVD is based on sequential R^2s, but taken care of the dependence on orderings by averaging over orderings [19, 20], either using unweighted averages (LMG) or weighted averages with data-dependent weights (PMVD). The LMG method produces a more equitable distribution of weights. Thus, LMG reflects the uncertainty of causal structure but does not describe it. For this reason we use PMVD metric. Let's describe the metric PMVD, introducing the following notation. In linear regression the coefficients β_k, $k = 0, ..., p$ are estimated by minimizing the sum of squared unexplained parts. Denoting $\hat{y}_i$ the fitted values and considering a set $\mathcal{S}$ of p regressors, R^2 is given by the ratio between regression deviance and total deviance:

$$R^2(\mathcal{S}) = \frac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}. \qquad (3)$$

R^2 measures the proportion of variation in y that is explained by the p regressors in the model.

The R^2, that gives the sequentially explained variance when adding the regressors in a set $\mathcal{M}$ to a model with that already contains the regressors with indices in $\mathcal{S}$, is given as

$$seqR^2(\mathcal{M}|\mathcal{S}) = R^2(\mathcal{M} \vee \mathcal{S}) - R^2(\mathcal{S}). \qquad (4)$$

The order of the regressors in any model is a permutation of the regressors $x_1, ..., x_p$ and is denoted by $r = (r_1, ..., r_p)$. Let $S_k(r)$ the set of regressors entered into the model before regressor x_k, in the order r, then the portion of R^2 allocated to regressor x_k in the order r can be written as

$$seqR^2(\{x_k\}|S_k(r)) = R^2(\{x_k\} \vee S_k(r)) - R^2(S_k(r)). \tag{5}$$

As said, PMVD can be seen as an average over orderings as well, with data-dependent weights for each order:

$$PMVD_k = \frac{1}{p!}\sum_{r=1}^{p!} p(r) seqR^2(\{x_k\}|r), \tag{6}$$

where $p(r)$ denotes the data-dependent weights. In this case, if all regressors have coefficient not zero, the permutation r has a weight proportional to

$$L(r) = \prod_{i=1}^{p-1} seqR^2(\{x_{r_{i+1}}, ..., x_{r_p}\}|\{x_{r_{i+1}}, ..., x_{r_{i+1}}\})^{-1} \tag{7}$$

and

$$p(r) = L(r)/\sum_r L(r) \tag{8}$$

is the probability associated to the order r, where summation in the denominator is over all possible permutations r. In other words, PMVD weights are obtained through a weighted mean of increases R^2 over all possible entry orders. Feldman's [12] proposal gives each order a weight as high as the first regressors catch a great portion of explained variance. This implies that the distribution of relative importance measures is concentrated on few regressors with high predictive power.

3. Our proposal: Indicator of Student Performance

As said, with the relative important metric we obtained determinants of satisfaction level (drivers of teaching quality). Our aim is to verify if students performance can influence the teaching evaluation. In particular, we concentrate our attention in student performance because in the last years, with the introduction of the Ministerial Decree 509/99, several changes of the Italian university system arise, as the introduction of UEC (University Educational Credits). One of the aims of the reform was to reduce the difference between the legal duration and the real duration which was too big before Ministerial Decree 509. Students that stay for a long time at university enter into the labour market at high age. This is also penalizing for universities, since they have to bear greater costs. In order to have information on students performance during the evaluation on the base of the questionnaire responses, we considered two variables, age and UEC. In fact, we have observed with simple descriptive statistics and inferential analyses, that age and UEC are the unique variables that influenced the student performance. To combine the two variables in a unique measure, we propose an Indicator of Student Performance (*ISP*) [22]:

$$ISP = 1 - ((A - 19) * 0.8 - C/60), \tag{9}$$

where A indicates the variable age declared by the student in the day when the questionnaire was fill out and C indicates the variable UEC (the credits he says to have acquired). UEC are divided by 60 to express them in terms of 'fruitful years', given that students should acquire 60 UEC per year. We subtract 19 from A, since it is the standard age students enter into the Italian university system. Therefore, the result is the number of years spent in university studies (assuming that students enter into the university system at the age of 19 exactly). This number is multiplied by 0.8, to adjust it to the standard of students performance, since students with an excellent career are very rarely observed. This is equivalent to assume that a student reaches, on average, 48 UEC per year. Indicator (9) can take negative values and it has not a theoretical maximum, since there is no theoretical maximum for student age. To sort out these drawbacks, ISP is standardized in the following way:

$$ISP^{\star} = \frac{ISP + k}{max(ISP + k)}, \tag{10}$$

where $k = -min(min(ISP), 0)$ is added to obtain a translation of values in the positive half-line and the denominator permits to obtain values between 0 and 1. $ISP^{\star}$ is equal to 0 when a student is 28 and he has just acquired 30 UEC; $ISP^{\star}$ is equal to 1 when ISP is equal to its maximum that is obtained crossing age 20 with the higher number of observed credits. $ISP^{\star}$ allow us (Section 4.2), to classify students in *bad* and *good* relating to their performance. Then, we obtained the drivers of teaching quality for these two groups by PMVD metric.

3.1 *Are good and bad students significantly different?*

To answer to this question, it is necessary to construct a statistical test to compare, for every item $k = 1, ..., K$ the weights obtained with PMVD metric for two grous. Because we have not standard error of PMVD, we utilize bootstrap procedure to construct an empirical sampling distribution and to assess the reliability of relative importance measures [11]. To build the statistical test, for two groups, we resample 500 times the values PMVD for every item, obtaining two matrices M_1 and M_2 of dimension 500xK. Then, relating these matrices, we obtain the ratio matrix R with generic element r_{ik}, where $i = 1, ..., 500$ is the sample dimension and $k = 1, ..., K$ indicates the item. The joint distribution of the K distributions r_k is a multinormal distribution. From R matrix we determine the variance and covariance matrix bootstrap $V^{\star}(\hat{R})$ of dimension KxK. The statistical test is the following [10]:

$$\hat{r}^T V^{\star}(\hat{R})^{-1} \hat{r}, \tag{11}$$

with a χ^2_K distribution, where $\hat{r}$ is the ratio vector of observed weights PMVD between two groups.

4. Application to teaching evaluation data of the University of Palermo

4.1 *Data*

In this study we analyze teaching evaluation data of a faculty of the University of Palermo in an academic year. We consider only the undergraduate courses. Students opinions are collected by means of a questionnaire that is filled out in the final part of the term. The evaluation form is made up of different sections concerning students personal characteristics and several aspects of university courses. These items (Table 1) are measured on a Likert scale with four categories: *decidedly no*, *more no than yes*, *more yes than no*, *decidedly yes*.

Table 1. Items of quality of teaching questionnaire

We exclude from the analysis the items related to practices since they are applicable only to a part of the teaching courses, the items, that we do not describe here, that are not well oriented toward the quality of teaching. Moreover, we eliminate questionnaires in which the percentage of attended lesson is less than 50%, as they could be not much reliable. In any case they should be treated separately. The dataset we analyze comprises 8503 questionnaires. The number of interviewed students is smaller than 8503, since each student could fill out as many questionnaires as the number of teaching courses he attended in the term.

Table 2 shows the percentage frequencies of evaluation item responses. Almost item distributions are positively skewed. In fact more than 50% of students gives positive responses to each item, with the exception of the item D2. Furthermore, there are some items concerning the teacher punctuality (F2, F3, F4, F5) that has the median in the best category.

Table 2. Percentage frequencies of evaluation items responses.

4.2 *ISP** *results*

In this section we present some considerations on indicator (10), justifying the classfication of students in relation to their performance. The graphic representation of conditional distribution of $ISP^\star$ given age (Figure 1) highlights an increasing monotonous trend of median level of non regularity to the growth of age. So, the variability of $ISP^\star$ is explained by age.

Figure 1. Boxplot of conditional distribution of $ISP^\star$ given Age.

Other considerations on indicator (10) can be drawn from Figure 2

- in this graphic we can observe the level curves of $ISP^\star$;
- the lowest values of the indicator are obtained for the students that have a very bad career;
- the indicator increases for decreasing values of age and/or increasing values of UEC;
- we can observe that in the top right side of the graphic there are not any observed values, because it is not possible a student is ahead of schedule;
- dots size highlights a very high frequency of students 19 years old who acquired 30 UEC. So we can consider that values between 0.6 and 0.7 correspond to a standard career. For example, this interval comprehends students who achieve a first degree (180 UEC) at 22 or 23 years old.

Figure 2. Level curves of $ISP^\star$ as a function of age and UEC with frequency classes of students.

In Table 3 we can observe the frequency distribution for classes of values of $ISP^\star$. The two classes of values greater than 0.7 represent good situations. In particular more than half (63.1%) of students have a excellent or standard career.

Table 3. Distribution of students for classes of values of $ISP^\star$.

These empirical considerations lead us to define the following dichotomous variable:

$$P = \begin{cases} 0 \text{ if } ISP^{\star} \leq 0.7 \\ 1 \text{ if } ISP^{\star} > 0.7. \end{cases}$$

P takes value 0 if the student time lag is greater than the standard one, i.e. if he has a bad career performance. On the other hand, when the time lag indicator is greater than 0.7 ($P = 1$), we consider the student has a good career performance. Differences in items responses between a bad and a good career performance can be observed in Table 4. Column I shows the indicator given in formula (12), following illustrated, that is a location indicator for ordinal distributions adjusted for the variability. According to that indicator, students with a bad career performance give higher ratings in almost all items. Results shown in Table 4 concern differences in marginal distribution and they do not give any suggestion on relations among items that will be investigated in the next section. It is interesting note that for item B8, that refers to adequacy of teaching material, the percentages of the two groups are equal and the two distributions perfectly overlap. Moreover, the item D2, that refers to the whole load of study, has lower percentages on best category both for $P = 0$ and $P = 1$.

Table 4. Percentage frequencies of evaluation items responses and indicator of formula (12).

4.3 *Relative importance metric results*

Students satisfaction depends on several aspects of the teaching activities, but not all with the same importance. We are interested in identifying which items are the drivers of quality of teaching in the students opinion [8], as in Capursi et al. [7], and above all, in highlighting possible differences between *good* and *bad* students. The complexity of the concept that we want to measure makes to necessary to pay attention to the exploratory analysis of data. In fact, evaluation items of the questionnaire are highly correlated, so it is difficult to identify those that greatly influence the global satisfaction. We use exploratory factor analysis in order to totally avoid subjective choice in the selection of variables. Particularly, we want to obtain weights (factor loadings) that, showing the correlation between items and the different dimensions. Before factor analysis, original ordinal data were aggregated by teaching course by means of the following transformation [5, 6]:

$$IS_{0.5} = 1 - \left(\frac{1}{3} \sum_{m=1}^{3} F_m^{0.5} \right)^2, \qquad (12)$$

where F_m is the cumulative distribution function of items responses in correspondence to the $m - th$ modelity of the ordinal variable. The transformation (12) is obtained as a prticular case of the complement to the unity of a relative index of dissimilarity between the ordinal empirical distribution of the judgments and the ordinal distribution 'excellent', namely the utmost agreement on the best judgment [5, 23]. So (12) gives a quantitative variable for each item and the statistical unit is the single teaching course. The results given by factor analysis allow us to identify which covariates enter into the regression model, whose parameters estimates are given by the PMVD metric.

4.3.1 Factorial analysis results

Factorial analysis [25] is obtained from a matrix whose rows are teaching courses of the faculty and whose columns are the items considered. The generic element of the matrix is the value that indicator (12) takes for a particular item in correspondence of a specific teaching course. To extract factors we use principal component method with varimax rotation. We consider factors with eigenvalues greater than 1. Table 5 shows the results of factorial analysis carried out separately for the two groups of students. For first group the extracted factors, with eigenvalues greater than 1, are four, for the second group are three. The first eigenvalues altogether explain 70% of variance and those of second group 68%. The two factorial analyses bring to similar results. As regards the first group, the first factor identifies aspect of teaching quality, associated to the single teaching, referring to preliminary information that a student receives at the beginning of each course: formative objectives (B3), examination procedures (B4), teaching material (B8), and lecturer's ability (F5, F6, F7). The second factor concerns items related to lecturer's punctuality (F2, F3, F4), the third one identifies aspects of organization of the degree course (D1, D2) and an organizational aspect of a single course (B10). The fourth factor highlights high correlation coefficients with items B11 that concerns the coordination of evaluated course with other courses and other two aspects of organization (D3, E1). In the second group the first factor is always correlated with items B3, B4, B8, F5, F6, F7 and also with item B11. The second component highlights the same items as for the first group; the third one identifies the dimension organization related item B10, D1, D2, D3, E1. Summarizing the results obtained, it can be highlighted that teaching quality in a strict sense, associated to the single course include three aspects: preliminary information (B3, B4, B8, B11) and teacher's punctuality (F2, F3, F4), teacher's ability (F5, F6, F7). These results brought our attention on such aspects.

Table 5. Rotated component matrix.

4.3.2 PMVD results

To carry out relative importance analysis, we consider the dimension for which the factorial weight of item C2 (overall satisfaction on teaching) is very high. This because it is thought that item C2 can express the general perception of teaching quality from students. To find the relative importance of such items, we use PMVD method on the basis of a linear model in which the indicator (12), that synthesizes item C2, is regressed on indicator of items identified by the first components of Table 5. Initially, we have eliminated 23 questionnaires in which students declared to attend the first year of their degree course at 24, 25, 26, 27 years; in fact it is our interest to consider only students that enrol at university at 18/19 years, i.e. "classical" students. Secondly, we consider a model in which ISP^* variable is present. Because of high correlation of item covariates, the effect of this variable is non relevant. For this reason we consider two separated models for the two groups of students (*bad* and *good*):

$$\begin{aligned} I_{C2i} = \beta_{0i} + \beta_{1i} IS_{B3i} + \beta_{2i} IS_{B4i} + \beta_{3i} IS_{B8i} + \\ + \beta_{4i} IS_{B11i} + \beta_{5i} IS_{F5i} + \beta_{6i} IS_{F6i} + \beta_{7i} IS_{F7i} + \epsilon_i, \end{aligned} \tag{13}$$

the first one ($i = 0$) for *bad* students and the second one ($i = 1$) for *good* students. The statistical unit, as said above, is teaching course, in particular we have 278 courses for *bad* students and 283 for *good* students. Results are shown in Table 6, where PMVD weights are scaled so that they sum to 1 to make interpretation easier. First of all, we can observe that item B4 has weight zero. R^2 is equal to 0.718 for the first model, and to 0.866 for second one. Observing the weights, the items that explain more the students satisfaction, in terms of relative importance, are F7 ("Is the lecturer clear in his/her exposition?"), F6 ("Does the lecturer stimulate/motivate interest toward his/her course?") and B3 ("Have been course educational objectives clearly exposed?"). However, there is a difference among the two models: in the model for the *good* students, the weight of F6 is three times greater than the weight that the item has in the first model (0.292 vs 0.104); *bad* students give a great importance than *good* students to the clarity of teaching the topics (F7) (0.727 vs 0.547). For both groups the explanation of formative objectives of teaching (B3) is important; for *bad* the adequacy of teaching material (B8) is more important than *good*. It seems that, somehow, the career performance, can be an element of discrimination to evaluate the teaching quality.

Table 6. PMVD weights of teaching quality items.

4.3.3 Boostrap results

The null hypothesis of statistical test (11) is:

$$H_0 : \beta_{k0} = \beta_{k1},$$

where β_{ki} are the coefficients of model (13) with $k = 1, ..., 7$ and $i = 0, 1$.

Summary statistics from bootstrap procedure are presented in Table 7. Mean bootstrap betas are consistent with the OLS (Ordinary Least Square) results (Table 8). For item F7 in *bad* group, PMVD component is slightly larger than share displayed in Table 8. For *good* students this consideration can be related to items B3, F6 and F7. This could be a consequence of the skew in component shares (these items have skew lower than other).

Considering that the excess kurtosis of the normal distribution is zero, the Bera Jarque p.values are based on the Bera Jarque test statistic and represents the confidence lvel in rejecting the hypothesis of asset return distribution normality based the sample values for the skew and kutorsis of the distribution. This test statistic is distributed χ_2. According this test, the hypothesis that residuals are normal can be accepted only for item B3 in *good* students. Figure 3 shows the univariate distribution of PMVD component shares for all items. It is evident that there are three types of distributions: approximately normal distribution observed for item B3 in *good* group; highly skewed distributions almost exponential in nature such as observed for items F7 for both groups; symmetric kurtotic distribution such as observed for F6 in *good* students. In particular items with a low weight PMVD are approximately exponential, items with a high explanatory power in terms of relative importance have skew and kurtosis values lower that others. The skew and kurtosis indicates that these items must be assumed to have more reliability than items such as B4, B8, B11 and F5. Moreover, for our aim, we can observe the non-overlapping between the two curve for all items.

Table 7. Boostrap statistics.

Table 8. OLS analysis.

In fact, considering the statistical test (11) with a χ^2_7 distribution, for $\alpha = 0.05$ we can reject the null hypothesis of equality of weights between two groups, that can be considered, in terms of relative importance, statistically different.

Figure 3. PMVD component boostrap distribution for bad and good students.

5. Discussion and conclusions

Recall the primary objective of this paper: to give simple statistical tools to build a composite indicator of SET in according to the student's career performance ancd on the basis of a questionnaire including a set of items. To reach this result we have chosen the item C2, i.d. the overall satisfaction of teaching, like synthesis of the quality of teaching in the opinion of the students and like the response variable in the linear regression model where the other items of the questionnaire are the covariates.

The most important regressors (or items) to predict the overall satisfaction are estimated by the PMVD metric that allows to single out a few elements [12, 13]. This aspect is meaningful from a twofold point of view: the first one concerns a methodological aspect related to the construction of a composite indicator, i.e. the weighting precedure and the second one concerns the nature of data typical of the social science, i.d. data are the result of an empirical study without any control on the variables and without a specific sampling strategy. The variables, therefore, often result correlated as in our case.

Indeed, the choice of weighting precedure is often characterized by a wide margin of subjectivity, the use of PMVD metric eliminates to a large of extent this problem. Besides, the R^2 decomposition used in PMVD metrci, by construction, holds under control the multicollinearity present among covariates and allows to get the weights to attribute to the select regressors in terms of their relative importance [13, 17, 18].

In this paper we introduce an indicator of student performance because we are convinced that the career performance can determine a different judgment on the quality of the teaching. This indicator can be a useful way to take into account the student performance because it is very simple to calculate and his graphical representation as a function of age and UEC (Figure 2) gives an immediate evidence of the order of magnitude of the regular students.

Acknowledgements

We would like to thank Vito Muggeo for his supervision and his helpful suggestions and Leonardo Librizzi for his availability.

References

[1] Achen C., *Interpreting and Using Regression*. Thousand Oaks, CA: Sage, 1982.

[2] Aiello F., Attanasio, M., *How to transform a batch of simple indicators to make a unique one?*, Atti del Convegno SIS June 2004, 327338.

[3] Aiello F., Capursi V., *Using the Rasch model to assess a university service on the basis of student opinions*, *Applied Srochastic Models in Business and Industry*, Vol. 24, 2008, pp. 459-470.

[4] Bacci S., *I modelli di Rasch nella valutazione della didattica universitaria*, *Statistica Applicata*, Vol. 18, 2006, n.1, University of Florence.

[5] Bernardi L., Capursi V., Librizzi L., *Measurement awareness: the use of indicators between expectations and opportunities*, SIS: Sezione Specializzata, Atti della XLIII Riunione Scientifica, Bari, 2004.

[6] Capursi V., Librizzi L., *La qualità della didattica: indicatori semplici o composti?*, *In: Capursi V., Ghellini G. (eds) Dottor Divago: Discernere valutare e governare la nuova Università*. Collana Valutazione AIV - Teoria, metodologia e ricerca. Franco Angeli, Milano, 2008, pp. 139-156.

[7] Capursi V., D'Agata R., Librizzi L., *A composite indicator of student evaluation teaching: from ordinal scales to metrics of relative importance*, Submitted.

[8] Campostrini S., Bernardi L., Slanzi D., *Le determinanti della valutazione della didattica attraverso il parere degli studenti*, VIII International Meeting on Quantitative Methods for Applied Sciences, 2006.

[9] Centra, J.A., *Reflective faculty evaluation*, San Francisco, CA, Jossey-Bass, 1993.

[10] Dobson A.J., *Introduction to Statistical Modelling*, London, Chapmann and Hall, 1983.

[11] Efron B., Tibshrani R. J., *An introduction to the bootstrap*, New York, Chapmann and Hall, 1993.

[12] Feldman B., *A Theory of Attribution. MPRA Paper 3349*, 2007, University Library of Munich, Germany.

[13] Feldman B., *Using PMVD to understand hedge fund performance drivers*, *Portfolio Analysis: Advanced Topics in Performance Measurement, Risk and Attribution*, Timothy Ryan (ed.), London: Risk Books, 2006.

[14] Firth D., *Relative importance of explanatory variables*, prepared for *Statistical 28 Issues in the Social Sciences, Stockholm*, Oxford: Nuffied College, 1998.

[15] Goldstein H., *Multilevel Statistical Models, 3rd edition*, London, Edward Arnold, 2003.

[16] Grilli C., Petrucci L., Rampichini C., *Analysis of university course evaluations: from descriptive measures to multilevel models*, *Statistical Methods & Applications*, 13 (2004), pp.357-373.

[17] Grömping U., *Estimators of Relative Importance in Linear Regression Based on Variance Decomposition*, *The American Statistician* 61 (2007), pp. 139-147.

[18] Grömping U., *Relative Importance for Linear Regression in R: The package relaimpo*, *Journal of Statistical Software*, 17, 2006.

[19] Kruskal W., *Relative Importance by Averaging over Orderings*, *The America Statistician*, 41, 1987a, pp. 6-10.

[20] Kruskal W., *Relative Importance by Averaging over Orderings*, *The America Statistician*, 41, 1987b, p. 341.

[21] Kulik, J.A., *Student ratings: validity, utility and controversy*, *New Directions for Institutional Research*, 27 (2001), pp. 925.

[22] Librizzi L., *Una proposta di indicatore composto della qualità della didattica al 'netto' delle caratteristiche degli studenti*, PhD. diss., Palermo University 2008.

[23] Leti G., *Statistica descrittiva*, Bologna, Il Mulino, 1983

[24] Lindeman R.H., Merenda P.F., Gold R.Z, *Introduction to Bivariate and Multivariate Analysis*. Scott, Foresman, Glenview, IL, 1980, p.199 ff.

[25] Mardia K.V., Kent J.T., Bibby J.M., *Multivariate Analysis*, London, Academic Press, Mcgraw-Hill, 1979.

[26] Marsh H.W., *Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research*, *International Journal of Educational Research*, 11 (1987), pp. 253-388.

[27] Marsh H.W., *Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and utility*, *Journal of Educational Psychology*, 76 (1984), pp. 707-754.

[28] Nardo M., Saisana, M., Saltelli, A. Tarantola, S., *Tools for composite indicators building*, EUR 21682 EN European Commission - JRC:Italy, 2005.

[29] Remmers H.H., *The relationship between students' markers and student attitudes towards instructor*, *School and Society*, 28 (1928) , pp.759-760.

[30] Remmers H.H., Brandenburg G.C., *Experimental data on the Pardue Rating Scale for instructors*, *Educational Administration and Supervision*, 13 (1927) , pp.519-527.

[31] Remmers H.H., Wykoff G.S., *Student rating of college teaching*, *School and Society*, 30 (1929) , pp.232-234.

[32] Wachtel, H.K., *Student evaluation of college teaching effectiveness: a brief review*, *Assessment and Evaluation in Higher Education*, 23 (1998), pp. 191210.

Table 1. Items of quality of teaching questionnaire

Items	Description
B3	Have the formative objectives of the teaching been explained in a clear way in the lecture hall?
B4	Have the modalities of the examination been explained in a clear way in the lecture hall?
B8	Is the teaching material (indicated or furnished) adequate for the studying of the subject?
B10	Is the load of study required by this teaching proportional to the credits assigned?
B11	Does the teaching have contents coordinated with other teachings?
C2	Are you satisfied of how this teaching has been carried out?
D1	Is the whole organisation (places, timetable, exams) of the teaching officially foreseen in this period acceptable?
D2	Is the whole load of study of the official teachings of the period acceptable?
D3	Does the teaching timetable take account of the movement time between two different lecture halls?
E1	Are the lecture halls adequate?
F2	Does the teacher inform the students with seasonableness when he is unable to hold the lesson?
F3	Does the teacher respect the scheduled teaching timetable?
F4	Does the teacher respect the sceduled consulting hours?
F5	Does the teacher express willingness to satisfy requests of clarification during the lessons?
F6	Does the teacher stimulate/motivate the interest in the subject?
F7	Does the teacher treat the topics in a clear way?

Table 2. Percentage frequencies of evaluation items responses.

Items	% frequency				# observations
	1	2	3	4	
B3	6	15	39	40	8373
B4	8	17	35	40	8395
B8	9	18	44	29	8435
B10	13	19	41	27	8435
B11	11	24	44	21	8398
C2	9	18	40	33	8446
D1	14	25	42	19	8432
D2	23	35	33	9	8404
D3	13	20	37	30	8335
E1	12	21	40	27	8448
F2	6	9	29	56	7489
F3	3	6	28	63	8387
F4	4	7	38	51	8122
F5	3	6	27	64	8380
F6	9	16	37	38	8401
F7	9	14	35	42	8393

1 = decidedly no 2 = more no than yes 3 = more yes than no 4 = decidedly yes

Table 3. Distribution of students for classes of values of $ISP^{\star}$.

Classes of values	Frequency	% frequency
0-0.1	34	0.4
0.1-0.2	755	0.4
0.2-0.3	3982	1.2
0.3-0.4	2400	1.2
0.4-0.5	742	2.6
0.5-0.6	266	7.5
0.6-0.7	140	15.8
0.7-0.8	100	34.0
0.8-0.9	36	35.9
0.9-1	33	0.9

Table 4. Percentage frequencies of evaluation items responses and indicator of formula (12).

	P = 0						P = 1					
Items	% freq.				# obs.	I	% freq.				# obs.	I
	1	2	3	4			1	2	3	4		
B3	5	13	40	42	2425	0.78	6	16	39	39	5909	0.75
B4	7	14	37	42	2428	0.76	9	19	34	38	5928	0.71
B8	9	19	43	29	2443	0.69	9	18	44	29	5956	0.69
B10	14	21	40	25	2439	0.63	13	19	40	28	5957	0.65
B11	7	20	46	27	2429	0.70	12	25	43	20	5932	0.62
C2	9	16	42	33	2446	0.71	10	19	39	32	5962	0.69
D1	14	25	41	20	2442	0.60	14	24	42	20	5952	0.61
D2	22	33	34	11	2427	0.48	23	36	33	8	5941	0.46
D3	12	21	39	28	2411	0.65	13	20	37	30	5888	0.65
E1	13	21	38	28	2443	0.64	12	20	41	27	5968	0.65
F2	5	9	31	55	2233	0.82	5	9	28	58	5225	0.83
F3	3	6	28	63	2421	0.87	2	6	28	64	5930	0.88
F4	4	7	38	51	2362	0.83	3	8	38	51	5728	0.84
F5	3	5	26	66	2421	0.88	3	6	28	63	5924	0.87
F6	8	14	38	40	2432	0.74	10	16	37	37	5935	0.71
F7	8	13	36	43	2426	0.75	10	15	35	40	5934	0.72

1 = decidedly no 2 = more no than yes 3 = more yes than no 4 = decidedly yes

Table 5. Rotated component matrix.

	P = 0				P = 1			
	Component				Component			
	1	2	3	4	1	2	3	4
B3	0.904	0.109	0.137	0.077	0.854	0.186	0.108	0.193
B4	0.755	0.150	-0.058	0.099	0.649	0.284	-0.032	0.354
B8	0.634	0.253	0.223	0.073	0.66	0.226	0.124	0.038
B10	0.182	0.226	0.748	0.015	0.340	0.127	0.664	-0.160
B11	0.456	0.022	0.013	0.165	0.392	-0.103	-0.115	0.642
C2	0.758	0.323	0.351	-0.076	0.889	0.218	0.207	0.102
D1	0.163	0.190	0.698	0.329	0.138	0.067	0.784	0.316
D2	0.126	-0.016	0.866	0.133	0.089	0.038	0.867	0.108
D3	0.117	0.130	0.350	0.753	0.008	0.132	0.412	0.646
E1	0.083	0.168	0.054	0.815	0.095	0.463	0.289	0.522
F2	0.196	0.732	-0.003	0.179	0.202	0.803	-0.050	0.102
F3	0.183	0.851	0.135	0.111	0.342	0.791	0.042	0.014
F4	0.193	0.812	0.199	0.026	0.437	0.701	0.156	0.016
F5	0.415	0.697	0.149	0.161	0.722	0.372	0.170	0.042
F6	0.792	0.313	0.219	-0.033	0.878	0.214	0.183	0.130
F7	0.793	0.363	0.243	0.029	0.905	0.149	0.131	0.048
Cum. % Var.	41.519	52.801	62.129	68.380	43.763	55.807	64.097	70.582

Table 6. PMVD weights of teaching quality items.

	PMVD	
Items	P = 0	P = 1
B3	0.115	0.135
B4	0.000	0.000
B8	0.042	0.024
B11	0.008	0.000
F5	0.004	0.002
F6	0.104	0.292
F7	0.727	0.547

Table 7. Boostrap statistics.

	bad						good					
Item	Mean Value	Std. Dev.	Skew	Excess Kurtosis	BJ- stat	BJ- p.value	Mean Value	Std. Dev.	Skew	Excess Kurtosis	BJ- stat	BJ- p.value
B3	0.134	0.118	1.061	0.852	108.86	0.000	0.140	0.069	0.584	0.035	28.464	0.328
B4	0.005	0.010	3.969	20.335	9927.784	0.000	0.001	0.002	2.035	4.818	828.709	0.000
B8	0.049	0.043	1.566	3.874	517.114	0.000	0.027	0.019	1.368	2.425	278.571	0.000
B11	0.017	0.027	3.127	15.449	5787.142	0.000	0.002	0.003	3.198	14.068	4975.390	0.000
F5	0.009	0.011	2.155	7.713	1626.31	0.000	0.007	0.010	2.862	10.905	3159.831	0.000
F6	0.127	0.111	1.284	1.694	197.132	0.000	0.292	0.122	0.390	-0.082	12.848	0.002
F7	0.659	0.132	-0.661	0.111	36.631	0.000	0.531	0.137	-0.327	-0.124	9.213	0.010

Table 8. OLS analysis.

	bad				good			
Parameter	Beta	Std. Err.	t-stat	p-val	Beta	Std. Err.	t-stat	p-val
Intercept	-0.006	0.060	-0.101	0.9208	-0.125	0.048	-2594	0.010
B3	0.173	0.078	2.217	0.027	0.261	0.054	4.792	0.000
B4	0.020	0.056	0.363	0.717	-0.020	0.040	-0.514	0.608
B8	0.161	0.048	2.067	0.001	0.136	0.034	3.989	0.000
B11	0.082	0.039	-1.425	0.040	0.023	0.027	-0.842	0.401
F5	-0.110	0.077	2.955	0.155	0.074	0.068	1.087	0.278
F6	0.193	0.065	2.955	0.003	0.259	0.053	4.932	0.000
F7	0.471	0.071	6.646	0.000	0.384	0.044	8.830	0.000

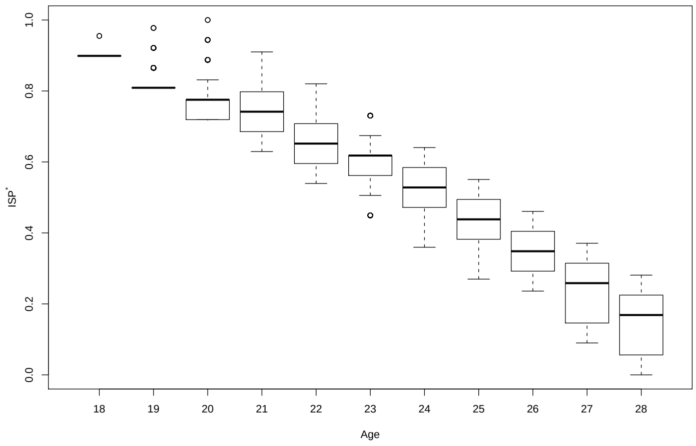

Figure 1. Boxplot of conditional distribution of $ISP^{\star}$ given Age.

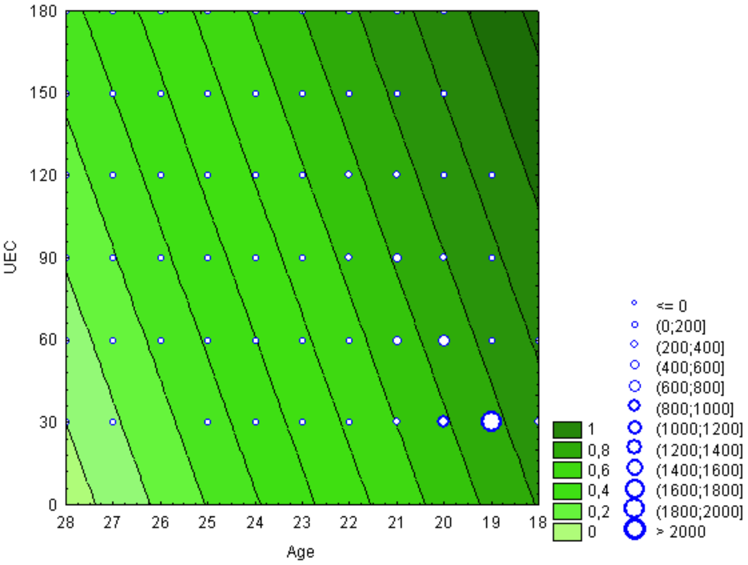

Figure 2. Level curves of $ISP^{\star}$ as a function of age and UEC with frequency classes of students.

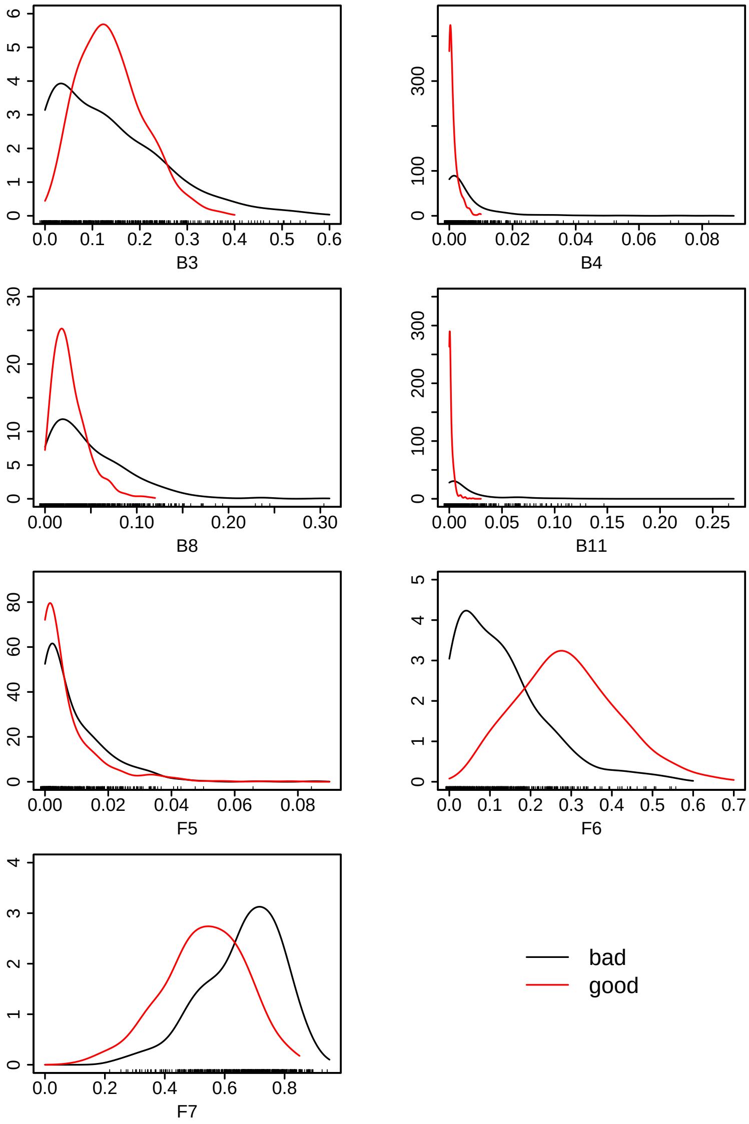

Figure 3. PMVD component boostrap distribution for bad and good students.