Italian University student evaluation of teaching: does it changes over time?

G. Boscaino[a]

[a]Department of Scienze Statistiche e Matematiche 'Silvio Vianelli', University of Palermo, Palermo, Italy.

Email: *gioboscaino@unipa.it*

Abstract

This paper concerns a study on students' perceived teacher activities quality. The aim is investigate about the changes over time of the students' opinions. To evaluate the perceived teachers' activities quality a Partial Credit Model is considered. This model allows to transform the ordinal data – collected through a questionnaire filled in by the students of a Faculty of University of Pavia, Italy – into interval scale data. To evaluate changes over time, the outcomes of the Rasch analysis are compared for three academic years: results show invariance – during the three years – of the assessment of the students on the aspects of the teachers' activity and variability in the mean levels of satisfaction.

keywords: Student Evaluation of Teaching, Partial Credit Model, changes over time, ANOVA.

1 – Introduction

During the last decade the Italian Higher Education evaluation has became more and more a crucial matter. To improve the quality of the University System (US) (see Appendix A for the Italian University System organization) is basic knowing its features, its functioning, every aspects that is typical of it. Only through the knowledge it is possible to make decisions, with the aim of managing and improving the System itself. Among these aspects, there is the teacher's quality. Since 1999 the Italian University System has monitored teachers' activities with the aim of improving the service offered to the students, and this practice has been set also by the Ministerial laws. Several changes have been made during the last years as regard both US and monitoring system itself. The principles that inspired such changes are based on the improvement of the whole service "University", of the competitiveness and of the accountability, and with the aim of getting closer to the US of the other European Countries. The reform – introduced by several bills, first D.L. 270/04 and after D.M.s 544/07 e 362/07 – introduced new concepts and new actions as the propensity to accountability of all the US processes, the accuracy in the management of financial and human resources, the monitoring and the evaluation of the main processes of the US. Those concepts have become the philosophy of the governance of the US.

The Italian literature offers several works about the evaluation of teaching, e.g. considerations on the whole system itself (Vittadini 2002), studies on measurement of the teaching quality using multilevel models (Rampichini, Grilli and Petrucci 2000), or through the specification of suitable indicators (Rampichini and Petrucci 2000; Capursi and Librizzi 2008), or by Rasch Model (Pagani and Zanarotti 2003; Bacci 2006; Boscaino 2006), or by multiple correspondence analysis (Iezzi 2005). Most of those works aim at evaluating students' satisfaction, the quality of teaching or to tune the questionnaire, just for a single teaching, for a single Degree Course or for a single Faculty.

The international literature is wider, and it focuses mostly on tuning the questionnaire of teaching evaluation, and on comparison among students ability, teachings, colleges, and among universities. Rasch models are widely adopted as powerful instrument to manage the students' responses (typically measured in ordinal scale) as interval scale measures with respect to the measurement theory (Bond 2005; Wolfe, Ray and Harris 2004; Beltukova, Stone and Fox 2004; Waugh 2003; just to mention a few).

This paper is organized in the following way: the second section regards the Italian University Student Evaluation of Teaching (SET) system and the structure of the questionnaire; the third section regards the aims and the data; the forth one introduces the methodology; last two sections concern with the results and the concluding remarks. An appendix about the Italian University System is reported at the end of the paper.

2 – The Italian University SET

Since almost 15 years the Italian Universities have performed a survey on students' opinions on teaching, e.g. on teachers punctuality, clearness and willingness, and on suitability of the facility, the classrooms and the laboratories, and on course scheduling and on managing.

Each University performs a survey with own procedures and through a questionnaire consisting of a core of items, required by the Ministry and the CNVSU – i.e. the National Board for the Evaluation in University System –, and other customizable items. The goal is to guarantee a homogenous national survey and ensuring data comparability at least for the core items set.

The aim of the survey is to gather information on teachers' teaching skill, education goals, coordination among teachings, and adequacy of resources. This information is useful to the policy makers to improve the quality of teachers' activities and of the Degree Courses.

In the matter of the SET, the survey is carried out three weeks before the end of the lessons period: the questionnaires are administer to the students who are attending the class – at that moment

of the survey (they are not informed when the questionnaires are administrated) – and they fill in it anonymously. Then data are elaborated, and their results are sent to Dean's office, as percentage distributions of the answers for each course, cancelling out any individual information. That report will be sent to the Rector, published on website of the Nucleo (i.e. the Evaluation Committee) and included in the Annual Report on Evaluation. It is sent to the MIUR (i.e. Italian Ministry of University and Research) and to CNVSU. The impact and the subsequent effects of this evaluation procedure is not still defined. So, if bad level of satisfaction is reported there is no consequences against the teacher.

The data used for the analysis regard the survey conducted at the University of Pavia. The questionnaire consists of 30 items divided into five sections:

A) the student

B) the teacher

C) the additional activities

D) the tutoring

E) the interest and satisfaction

Items reported below according to the 5 sections measured in a Likert-type scale with 4 ordinal categories of response: 'definitely no", "more no than yes", "more yes than no", "definitively yes". I report any differences in the item categories when they occur. Those items marked with * belong to the core set required by CNVSU.

Section A collects information about the student who fills in the questionnaire:

A01: Enrolment status

(this item has 3 categories of response: 'regular'; 'repeating the year'; 'out of normal course duration')

A02: Working activity during this term

(this item has 3 categories: 'no activity'; 'part time jobs'; 'full time')

A03*: Are basic access competences enough to understand courses topics?

A04: If ‘definitely no’ or ‘more no than yes’, why?

(this item has 4 specific categories: ‘Knowledge out of my academic curriculum are needed to understand the topics’; ‘I have not the adequate knowledge because I did not study previous subjects’; ‘Previous topics did not give the adequate knowledge to understand these new topics’; ‘My last exam is so remote that I am not able anymore to follow well this course’)

A05: How many class hours did you attend? (in percentage)

(this item has 4 categories: <10%; about 50%, about 70%, about 100%)

Section B regards the evaluation the teacher quality in student’s opinion:

B01*: Is study material (suggested or supplied) convenient for the course?

B02: Is the teacher’s teaching quality effective?

B03: Does the teacher stimulate/motivate interest towards his/her course?

B04: Is the teacher clear in his/her exposition?

B05*: Is the workload for this course proportioned to its credit?

B06: Does the teacher follow the scheduled timetable for the office hours?

B07: Is the teacher available for explanations during classes?

B08: Does the teacher follow the scheduled timetable for this course?

B09*: Have been examinations procedures clearly exposed?

B10: Percentage of course hours have been taught by the course regular teacher

(this item has 3 category of response: <80%; >80%; 100%)

B11*: Are teaching halls suitable? (it is possible to see, to hear, to have a seat...)

B12*: Is the overall workload for the scheduled courses during this term sustainable?

B13*: Is the overall course organization (teaching halls, course timetable, exams, etc.) during this term acceptable?

Section C gathers information about practical activities, laboratories, seminars, etc.

C01: Are additional activities provided for this teaching?

(this item has 2 categories: ‘yes’, and ‘no’. If ‘no’, skip to section D)

C02: Percentage of additional activities attended.

(this item has 4 categories: <10%; about 50%, about 70%, about 100%)

C03*: Are additional activities useful for learning?

C04: Is the teacher’s quality effective?

C05: Does the teacher follow the scheduled timetable for this course?

C06: Are study material and teaching instruments convenient?

C07*: Are the facilities and the tools for additional activities suitable?

Section D regards the evaluation of the tutoring service:

D01: Did you make use of tutoring service for this teaching?

(this item has 5 categories: ‘yes’; ‘no, it was not expected’; ‘no, I did not know it was expected’; ‘no, I was not interested in’; ‘no, but I think I am going to make use of it before the examination’)

D02: If ‘yes’ on D01, how often did you make use of it?

(this item has 3 categories: ‘often’; ‘sometimes’; ‘just once’)

D03: Are you satisfied about this service?

Section E regards the overall level of interest and satisfaction:

E01*: Are you interested in the topics of this course irrespective of the teacher performance?

E02*: Are you satisfied how this course has been carried out?

3 – Aims and data

The main aim of this work is to investigate the changing over time of quality of teacher's activity as it is perceived by the attending students. Are the students' opinions constant over time? Do best and worst aspects of the teacher's activity change over time? Teacher quality is obviously a concept that can not be directly measureable. It is a latent concept and it is accounted as the good practice of the teacher in pursuance of his/her teaching activity. The above mentioned questionnaire includes items (among others) strictly linked to the teachers' activity. The quality of a teacher is a concept that the US tries to measure through a set of items: if the level of agreement of the students on each item – that measure a single aspect of the concept – is high, than the teacher quality is supposed to be high. To hit the mark, a comparison among students' opinions has been performed. A Rasch model has been adopted, and just a subset of items – among those listed above – has been considered. Rasch model gives item calibration and person measure on the same continuum and on an interval scale (and with the same scale unit, the logit). This is very useful to make comparisons among items, among students satisfaction levels, among items and satisfaction levels, and over time. Those items regarding student general information, tutoring and additional activities, the facility, workload and overall organization of the course are not taken into account in the analysis: they concern other dimensions, rather faraway from teacher's activity quality. Therefore, the items of Section A, C, D and from B10 to B13 have not been considered. The remaining items are particularly important because they easily combine to measure the tecaher activities quality. Regarding the E section items, item E01 could be considered as a "causal variable", in the Fayers and Hand way (2002): an item that "if it is present (score highly, say) then the concept in question is present". The a-priori interest in the course topic (item E01) cannot be considered as an aspect of the quality of the teacher, but it is possible to consider it as a predictor of the level of satisfaction of the student. Consequently, the final set of items that pertains to the analysis covers with those from B01 to B09, and E02.

The above mentioned questionnaire was adopted by University of Pavia since the a.y. 2004/05 to the a.y. 2006/07, afterwards it was modified. The dataset used for the analysis regards the survey – involving 1st cycle undergraduate students attending to the lessons – carried out during the three a.y.s at the Engineering Faculty. It numbers 340 teachings and 360 teachers during the first a.y., 386 teachings and 409 teachers in the second one, and 365 teachings and 384 teachers during the third a.y. (the difference between teachers and teachings number, for each a.y., is due to the fact that the same teaching may be carried out by two or three teachers). I have considered just 5 teachers for the analysis, because I have chosen those subjects with a lot of attending students (questionnaire filled in) and which have been taught by the same teacher during the 3 a.y.s. Table 1 summarizes the distributions of the students who have filled in the questionnaire – or the number of questionnaires collected –, by teacher and academic year.

TABLE 1

4 – Method

The analysis of teacher perceived quality has been performed using the Partial Credit Model (PCM) (Masters 1982). This model belongs to the Rasch Models (Rasch 1960) which are a subset of the Item Response Theory. In the Rasch Models, taking into account a one-dimensional latent trait to be measured, the probability of a specified response (e.g. right/wrong or satisfied/unsatisfied) is modelled as a function of person and item parameters. Specifically, in the dichotomous Rasch model, the probability of a "correct" response is modelled as a logistic function of the difference between the person ability (satisfaction) (β) and item difficulty (quality) (δ) parameter. The parameters of the model pertain to the level of a quantitative trait possessed by a person or item. The purpose of applying the model is to obtain measures (in physical sense) from categorical response data. Let us suppose that a person has to solve an item with ordinal response

categories, and each category represents an increasing "difficulty step" to solve the item. If so, the PCM conventional representation is:

$$\pi_{nix} = \frac{\exp \sum_{j=0}^{x} \left(\beta_n - \delta_{ij} \right)}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} \left(\beta_n - \delta_{ij} \right)} \quad (1)$$

where:

π_{nix} is the probability that a person n encountering item i is observed in category x;

β_n is the ability (satisfaction) measure of person n;

δ_{ij} is the difficulty (quality) measure of the aspect of the concept catch by item i and observed in the category j.

The numerator contains only the "difficulties" of the x completed steps (from 1 to x), while the denominator is the sum of all the m_i+1 (from 0 to m_i) possible numerators (Wright and Masters 1982).

To check the unidimensionality, following the statement of Linacre (2006) the "unidimensionality is never perfect", I used the Linacre suggested rule of thumb to evaluate dimensionality. This is performed through the Rasch-residual-based Principal Components Analysis (PCAR) – these components show contrasts between opposing factors, not loading on one factor (Linacre 2006). The rule of thumb consists in the variance explained by measures more than 60%, unexplained variance explained by the 1st contrast (size) less than 3.0, and unexplained variance explained by the 1st contrast less than 5%. Moreover, Linacre suggests to investigate on the correlation between two subsets of items, based on loadings on the first residual contrast. If R1 and R2 are the reliabilities of the two subsets, and r is the correlation among the person 'ability' measures, then their latent (error-disattenuated) correlation approximates to $C = r/\sqrt{R1 \cdot R2}$. If C

approaches 1.0, then the two subsets are statistically equivalent (Linacre 2006) – “the problem is not if there is multidimensionality but if it is big enough to merit dividing the items into separate tests”.

Rasch analysis has been performed using Winsteps software (Linacre 2006), and through the Joint Maximum Likelihood estimation method because it has several advantages, like the independence from specific person and item distribution form, unbiased measure estimates, independence from the length of the test and from the sample size. Further analyses has been performed using Statistica software.

5 – Results

The results arise from 15 PCM analyses: 5 teachers for 3 academic years. The analyses focused on the survey results not on the tuning of the questionnaire. The questionnaire already exists, I am not developing a new test. Then, no item selection or thresholds analysis have been performed. I have used all the considered items because “even an item with a very low mean-square tell us a little something that is new and useful. We do not want to waste any useful information” (Linacre 1994).

I have just reported the analysis of the adequacy of the measures and of the unidimensionality.

Measures adequacy

Even if I do not want to validate the questionnaire or improve it modifying, deleting or adding any item, I want to check if data accord to the Rasch model. Table 2 reports the range for the

15 reliability indexes – one for each PCM analysis – of the item calibrations and the person measures: all the reliabilities support the evidence of consistence. In this table are also reported the ranges for the item and person INFIT for the 15 analyses: as the expected INFIT value is 1, we can be satisfied about the overall person and item fit.

Moreover, the range of the means of the persons measures suggests that the student are quite satisfied of the teaching quality. These values are useful to check the targeting of the items calibrations and the persons measures. As example, the Figure 1 reports the item/person map relative to the teacher D and the year 2005: the vertical dotted line is the continuum of the teaching quality – on the continuum is reported the scale in logit unit. On the right side there are the item calibrations and on the left side the person measures. Each X is a person. M label marks the Mean, S label is placed 1 standard deviation away from the mean, and T two standard deviation away. M is equal to 0 for the items calibration because it is a constrain in the estimation process (and represents the conventional 0 of the interval logit scale). Through the map it is possible to appreciate the good targeting of the items calibrations and the persons measures. They share the same part of the continuum.

TABLE 2

FIGURE 1

Unidimensionality

The analysis shows evidence of unidimensionality of the set of items taken into account, but the third academic year for the teacher E. Whereas there is some outcome which is borderline – even if C is close to 1 – for the last dataset (E-3) the evidence of multidimensionality is clear, although in the previous years (and for the dataset concerning the other teachers) the set of items shows unidimensionality (Tab. 3).

TABLE 3

Satisfaction

The analysis of the satisfaction of the students on the teacher's activity points out high levels of the mean satisfaction if compared to the quality level expressed by the set of items (mean equals to zero, by constrain) (Tab. 4). An one-way ANOVA was performed to check differences among mean levels of satisfactions measured for each academic year. During the three years the mean satisfaction levels are significatively different, with the exception of the teacher A (Tab. 4). The analysis of the contrasts does not give indication of any "trend" of satisfaction: there is no evidence of increasing – or decreasing – satisfaction level over the three academic years (except for teacher D). A longer period is needed to study the satisfaction trend…

TABLE 4

Items locations, or best/worst aspects of the teachers

The specific objectivity of the Rasch Models allows comparisons of the items locations among the three years, among the teachers, and among years and teachers. In Figure 2 is reported the continuum (horizontal axis) and the location of the items calibrations for each year conditioned by teacher. Therefore, it is possible to compare simultaneously the locations among year for each teacher, and among teachers, and to check which are the best and worst teacher's activities during the three years. Items B07, and B08 are those that have lower locations on the continuum – that correspond to the aspects that are best appreciated – whereas items B04 and B05 seem to be the worst ones. Therefore, the teacher "punctuality" and the availability for the explanations are the

most appreciated aspects by the students. The workload and the clearness of the exposition seems to be not so appreciated. In the figure are also reported the mean student satisfaction measures (the 'X'). Thanks to this representation it is possible to compare and summarize the measurement process: it is possible to check how the satisfaction changes, over time and for each teacher, and it is possible to compare the satisfaction levels of the student with the single item/aspect calibration.

FIGURE 2

6 – Concluding Remarks

This paper reports the analyses conducted on 15 dataset consisting on students evaluation teaching collected for 5 teachers during 3 academic years. The Partial Credit Model was adopted to transform ordinal scale data into interval scale ones, to check unidimensionality of the set of items taken into account, to investigate if the evaluation of the teachers' activities and the satisfaction level change over time. Some constant elements rise: the adequacy of the measures and the unidimensionality hold; the worst and best aspects of the teachers' activity seem to be the same for all the teachers and over time; and the mean level of student satisfaction seems to change over time. The graphical representation of these results (Fig. 2) is a very useful tool to check best and worst teachers' activities aspects, to monitoring the satisfaction levels of the student, and to make comparisons. The results may give some hints for policy makers, if they want to improve the teaching quality and the students satisfaction. For example, they and the teachers should improve the aspects of the course suggested material and the workload of the courses, because the students seem do not appreciate them very well. Among further developments, it is possible to use the quantitative measures to build an Indicator of 'good teacher'. Regarding the satisfaction level, it is obvious that a longer period is needed to analyse if a satisfaction trend exists: three years are too little. Moreover, an ad hoc survey with a questionnaire built in the Rasch logic is preferred if we

want to investigate the teacher activity better: the used questionnaire measures different dimension of the teaching and of the degree course organization. Other methodologies exist to analyse the student evaluation of the teaching. Among these, the Relative Importance Metrics is another way to obtain quantitative measures from qualitative measure.

Appendix A – The Italian University System

The Decree no. 509 of November 3, 1999 and the Ministerial Decree no. 270 of October 22, 2004 have changed the Italian University System and the organization of Degree Courses. In Table A is reported a scheme o the Degree Courses currently granted by the Italian Universities (article 3 of Ministerial Decree no. 270/2004):

TABLE A

Degree Courses follow the European Credit Transfer and Accumulation System (ECTS) structure. A University credit corresponds to 25 hours of workload per student. The average annual workload of a full time student is conventionally fixed at 60 credits. Each University and Faculty is free to organize itself the a.y.. The Academic Year starts at the middle of September and stops at the end of the February of two years after: e.g. the a.y. start on 20th September 2010 and stops on 28th February 2012. Most of the Univeristies organizes the a.y. into two parts: one aimed to the lessons and one aimed to examinations. The lesson period is generally divided into two semesters, and each course is developed during a single semester. At the end of this period related examinations start. When a student enrols at the University, he/she enrols at a specific Degree Course that is a set of fixed teachings – with the exception of about two teachings that can be chosen by the student – and fixed teachers, one for each teaching – with the exception of those class with many students: in this case the teaching is divided among two or more teachers.

The first-cycle degree is attained at the end of a 3-years course and it aims at providing “adequate command of general scientific methods and knowledge", namely the basic skills allowing students to find a job. The Second-Cycle Degree is attained at the end of a further 2-years course and it aims at providing “advanced training for the activities requiring high qualification in specific fields". It provides specialized knowledge to gain access to specific professions. The Research

Doctorate Programme aims at training postgraduates for very advanced scientific research or for professional appointments of the highest consequence; they envisage the use of suitable teaching methodologies such as updated technologies, study periods abroad, stages in specialistic research centres. The 2nd (level) Specialization degree course is devised to provide postgraduates with knowledge and abilities as required in the practice of highly qualifying professions; they may be established exclusively in application of specific Italian laws or EU directives. Finally, the 2nd (level) University Master degree course consists in advanced scientific courses or higher continuing education studies. For further information visit the CRUI (The Conference of Italian University Rectors) website http://www.crui.it/crui/ECTS/english/italian_univ_system.htm.

Reference

Andrich, D. 1985. An elaboration of Guttman scaling with Rasch models for measurement. In *Sociological methodology*, ed. N.B. Tuma, 33-80. San Francisco CA: Jossey-Bass.

Bacci, S. 2006. I modelli di Rasch nella valutazione della didattica universitaria. *Statistica Applicata* 18, n. 1

Beltyukova, S.A., G.F. Stone, and C.M. Fox 2004. Equting Student Satisfaction Measures. *Journal of Applied Measurement* 5, n.1

Bond, T.G., and C.M. Fox 2001. *Applying the Rasch Model – Fundamental Measurement in the Human Sciences*. NJ: Erlbaum

Bond, T.G. 2005. Accountability in the Academy: Rasch Measurement of Student Feedback Survey. In *Frontiers in Educational Psychology*, ed. Russell F. Waugh, 119-129. Nova Science Publisher.

Boscaino, G. 2006. La qualità della didattica: la calibrazione dello strumento di misura con il modello di Rasch. PhD diss., University of Palermo, Italy.

Capursi, V., and L. Librizzi 2008. La qualità della didattica: indicatori semplici o composti?. In *Dottor Divago: Discernere Valutare e Governare la nuova Università*, ed. V. Capursi, and G. Ghellini, 139-155. MILANO: Franco Angeli (ITALY)

Curtis, D.D., and P. Boman 2007. X-ray your data with Rasch. *International educational Journal* 8, n. 2.

Fayers, P. M., and D.J. Hand 2002. Causal variables, indicator variables and measurement scales: an example from quality of life. *Journal of the Royal Statistical Society*, Series A 165, n. 2: 233-261

Fischer, G.H., and I. Molenaar 1995. *Rasch Models Foundations, Recent Developments, and Applications*. New York: Springer.

Holland, P.W., and H. Wainer 1993. *Differential Item Functioning*. Lawrence Erlbaum Associates

Iezzi, D.F. 2005. A method to measure the quality on teaching evaluation of the university system: the Italian case. *Social Indicators research*. Springer 73, 459 – 477

Linacre, J.M. 2006. *WINSTEPS Rasch measurement computer program*. Chicago: Winsteps.com

Linacre, J.M., and B.D. Wright 1994. (Dichotomous Mean-square) Infit and Outfit Chi-square fit statistics. *Rasch Measurement Transactions* 8, n. 2, http://www.rasch.org/rmt/rmt82a.htm

Masters, G.N. 1982. A Rasch model for partial credit scoring. *Psychometrika* 47, 149-174.

Masters, G.N., and B.D. Wright 1982. *Rating scale analysis: Rasch measurement*. Chicago: MESA Press

Pagani, L., and M.C. Zanarotti 2003. Analisi della qualità di un servizio: un confronto tra scale mediante il modello di Rasch. *Statistica e Applicazioni* 1, 35-54

Rampichini, C., L. Grilli, and A. Petrucci 2000. Analisi della didattica universitaria attraverso modelli multi-livello. In *Metodi e tecniche per le rilevazioni assistite da computer*, ed. M. Civardi, and L. Fabbris, vol. 3. Padova: CLEUP.

Rampichini, C., and A. Petrucci 2000. Indicatori statistici per la valutazione della didattica universitaria. In *Metodi e tecniche per le rilevazioni assistite da computer*, ed. M. Civardi, and L. Fabbris, vol. 3. Padova: CLEUP.

Rasch, G. 1960. *Probabilistic Models for some Intelligence and attainment tests*. Copenhagen Danish Institute for Educational Research [expanded edition (1980) with foreword and afterword by Wright, B.D.. Chicago: The University of Chicago Press]

Vittadini, G. 2002. La valutazione dell'Università: misurazione del prodotto. In *La valutazione dell'Università, accreditamento del processo, misurazione del prodotto*, ed L. Modica. FrancoAngeli.

Waugh, R.F. 2003. Evaluation of quality of student experiences at University using a Rasch measurement model. *Studies in Educational Evaluation* 29, 145-168. Elsevier Science Ltd.

Wright, B.D., and J.M. Linacre 1994. Reasonable mean-square fit values. *Rasch Measurement Transactions* 8, n. 3, http://www.rasch.org/rmt/rmt83b.htm

Wolfe, E.W., L.M. Ray, and D.C. Harris 2004. A Rasch Analysis of Three Measures of Teacher Perception Generated from the School and Staffing Survey. *Educational and Psychological Measurement* 64, n. 5, 842-860.

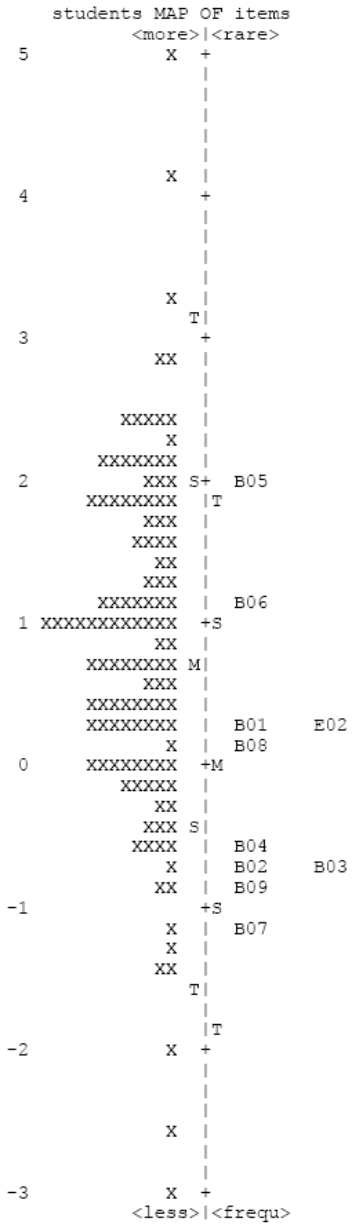
Table 1 – Distribution of questionnaire collected, by teacher and academic year

	Academic Year			Total
Teacher	2004/05	2005/06	2006/07	
A	93	78	69	240
B	108	117	110	335
C	100	91	96	287
D	92	123	106	321
E	71	94	84	249

Table 2 – Summary statistics for the 15 PCM analyses

Item reliabilities	[0.75; 0.97]
Person reliabilities	[0.73; 0.86]
Item INFIT	[0.98; 1.07]
Person INFIT	[0.95; 1.05]
Mean person measure	[0.57; 1.88]

Figure 1 – Item/Person Map for teacher D and the year 2005



Note: X = 1 student; M = Mean; S = 1 s.d. away from M; T = 2 s.d. away from M

Table 3 – Summary of unidimensionality analysis

		Variance Explained by Measure		Unexplained Variance in 1st Contrast		C
Teacher	A.Y.	Empirical (%)	Modeled (%)	Empirical	Empirical (%)	
A	1	83.3	87.6	2.1	3.5	0.75
	2	69.6	76.7	2.1	6.4	0.86
	3	79.1	79.1	2.1	4.4	0.84
B	1	59.9	64.3	2.3	9.2	0.78
	2	66.1	65.8	2.0	6.8	0.83
	3	69.8	69.3	2.3	7.1	0.79
C	1	61.6	69.4	1.9	7.3	0.89
	2	69.6	72.4	2.5	7.7	0.99
	3	69.2	69.9	2.0	6.2	0.94
D	1	69.0	70.0	2.7	8.5	0.90
	2	69.8	70.4	2.0	5.9	0.94
	3	69.5	72.6	2.0	6.2	0.91
E	1	70.4	72.1	2.7	7.9	0.83
	2	69.1	71.2	2.0	6.1	0.98
	3	52.0	52.6	2.7	12.9	0.58

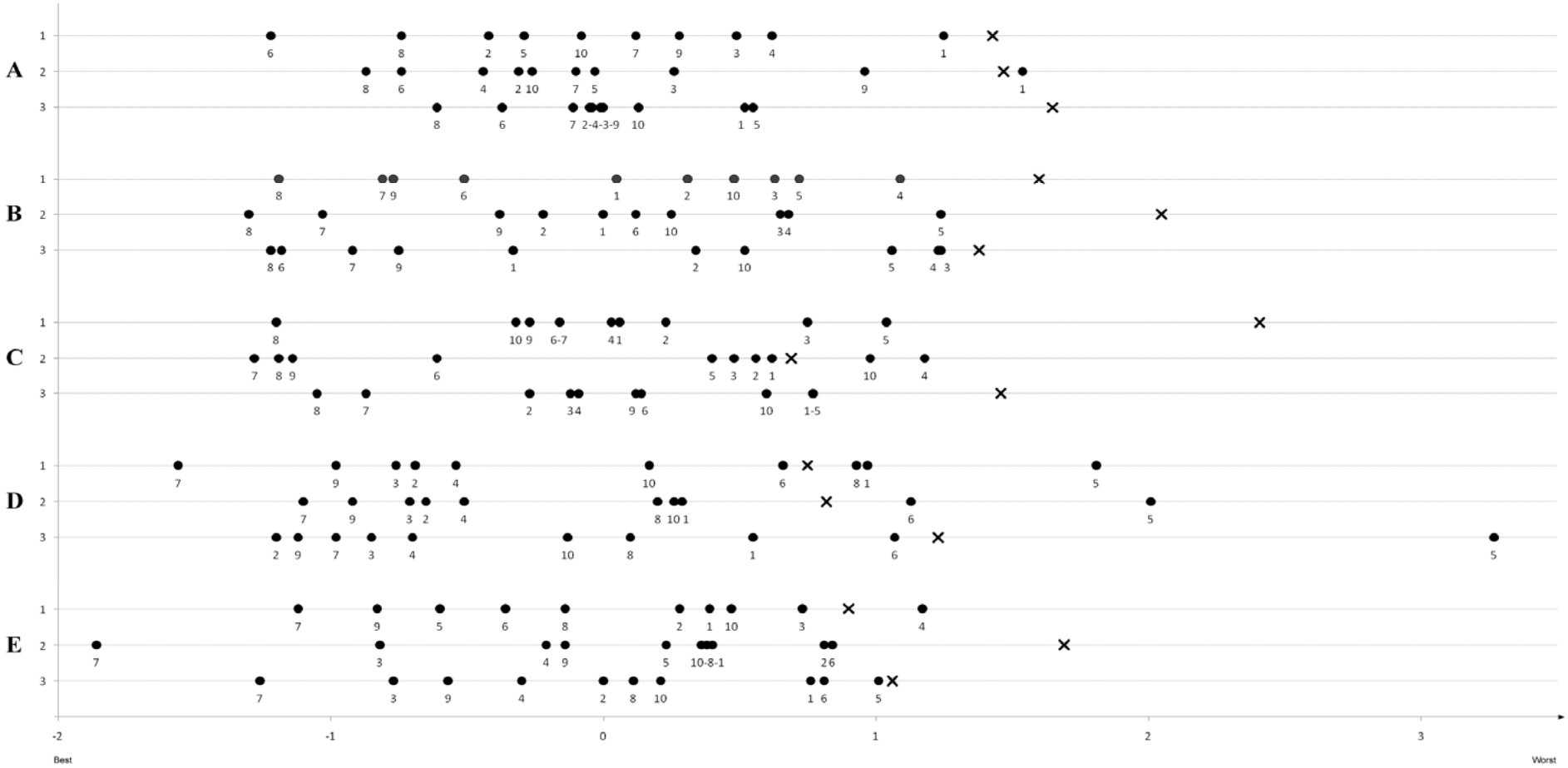
Note: academic year: 1=2004/05; 2=2005/06; 3=2006/07

Table 4 – Summary of one-way ANOVA of student satisfaction level, and analysis of contrasts

		Measure		p-value		
Teacher	A.Y.	Mean	S.D.		Contrast	p-value
A	1	1.43	1.17	0.525	1 vs 2	0.806
	2	1.47	1.28		1, 2 vs 3	0.273
	3	1.65	1.37			
B	1	1.60	1.39	0.012	1 vs 2	0.048
	2	2.05	1.91		2 vs 3	0.004
	3	1.38	1.81			
C	1	2.41	1.90	0.000	1 vs 2	0.000
	2	0.69	1.94		2 vs 3	0.006
	3	1.46	1.85			
D	1	0.75	1.26	0.015	1 vs 2	0.724
	2	0.82	1.22		1, 2 vs 3	0.004
	3	1.23	1.40			
E	1	0.90	1.58	0.004	1 vs 2	0.002
	2	1.69	1.75		2 vs 3	0.011
	3	1.06	1.54			

Note: academic year: 1=2004/05; 2=2005/06; 3=2006/07

Figure 2 – Items locations (●) and Person Mean Satisfaction Level (X) by teacher and academic year



Note: for simplicity, items B01, B02, …, B09 are labelled from '1' to '9', and item E02 is labelled with '10'
teacher: A, B,C, D, E
academic year: 1=2004/05; 2=2005/06; 3=2006/07

Table A – The Italian University System

Degree Courses	**Degrees**	**Credits earned**	**Years**
1st cycle - undergraduate studies/students:			
Laurea degree course	1st degree / Laurea degree	180	3
2nd cycle - graduate studies/students:			
2nd (Laurea) degree course	2nd degree (2nd Laurea)	120	2
1st (level) Specialization degree course	1st (level) Specialization degree	120-180	2-3
1st (level) University Master degree course	1st (level) University Master degree	60+	1+
3rd cycle - postgraduate studies/students:			
Research Doctorate programme (RDP)	Research Doctorate degree		3+
2nd (level) Specialization degree course (SC2)	2nd (level) Specialization degree	60-300	1-5
2nd (level) University Master degree course (UMC2)	2nd (level) University Master degree	60+	1+

TOTAL WORDS NUMBER: 5205