Available online at www.sciencedirect.com
SciVerse ScienceDirect
ELSEVIER
Procedia - Social and Behavioral Sciences 64 (2012) 402 – 411
Procedia Social and Behavioral Sciences

INTERNATIONAL EDUCATIONAL TECHNOLOGY CONFERENCE IETC2012

A New Approach to Evaluation of University Teaching Considering Heterogeneity of Students' Preferences

Marija Kuzmanovic*, Gordana Savic, Milena Popovic, Milan Martic

University of Belgrade, Faculty of Organizational Sciences, Jove Ilica 154, Belgrade 11000, Serbia

Abstract

Students' evaluations of teaching are increasingly used by universities to evaluate teaching performance. However, these evaluations are controversial mainly due to fact that students value various aspects of excellent teaching differently. Therefore, in this paper we propose a new approach to student evaluation of university teaching based on data from conjoint analysis. Conjoint analysis is a multivariate technique used to analyze the structure of individuals' preference. In particular, our approach accounts for different importance students attach to various aspects of teaching. Moreover, it accounts explicitly for heterogeneity arising from student preferences, and incorporates it to form comprehensive teaching evaluation score. We have conducted survey and confirmed applicability and efficiency of the proposed approach.

Keywords: Student evaluation of teaching; conjoint analysis; students' preferences; criteria importance; heterogeneity; overall rating.

1. Introduction

Students' evaluations of educational effectiveness are widely used to evaluate teaching effectiveness and educational quality in many countries across the world. These evaluations are important as feedback to students, teachers, departments, university administrators, governmental policymakers, and researchers. Hence, it is not surprising that substantive and methodological studies in this area have resulted in a huge research literature. The vast majority of these researches are based on the traditional approach to students' evaluations of teaching (SET), in which students in a specific course taught by a specific teacher evaluate the teaching effectiveness of their teacher, typically near the end of the term (Marsh et al. 2011). Ratings

E-mail address: marija.kuzmanovic@fon.bg.ac.rs

1877-0428 © 2012 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of The Association Science Education and Technology
Open access under CC BY-NC-ND license. doi:10.1016/j.sbspro.2012.11.047

by all students within the course are aggregated to form course-average ratings that are used as feedback to teachers to improve their effectiveness; the ratings are sometimes also used for personnel decisions by administrators, coursework selection by students, and research (Marsh 2007).

In higher education, there is a long history of research and much debate into the appropriate use of SETs (Marsh 2007; Marsh and Roche 2000; McKeachie 1997). One major shortcoming of current procedures is that it is uncertain if the survey items of SET questionnaires properly represent the underlying constructs for which they were developed. Marsh (2007) suggested that effective teaching is a hypothetical construct for which there is not a single indicator and the validity of the dimensions of student evaluations should be demonstrated through a construct validation. Both, researchers and practitioners (Abrami and d'Apollonia 1991; Cashin and Downey 1992; Marsh and Roche 1993) agree that teaching is a complex activity with multiple interrelated components (e.g., clarity, interaction, organization, enthusiasm, feedback). Hence, it should not be surprising that SETs, like the teaching they are intended to represent, are also multidimensional. Particularly formative/diagnostic feedback intended to be useful for improving teaching should reflect this multidimensionality. For example, a teacher can be organized but lack enthusiasm or he can be always available for students but lack clarity.

There are many ways of evaluating educational activity and therefore the teaching staff. Berk (2005), in a recent review, describes up to 12 varieties of evaluation. Most evaluations are represented by a Likert-format scale consisting of items that have been designed to assess some aspect of teaching. Responses to these items are then averaged to produce a mean teaching performance score (Witte and Rogge 2011). This average is then used as an index of teaching effectiveness and used for formative and/or summative evaluation. A somewhat similar approach consists of summing the ratings and expressing them as a percentage to the maximal attainable overall rating (Liaw and Goh 2003). A third way is asking students to rate the overall performance of the teacher on one single scale (Davies et al. 2007; Ellis et al. 2003).

However, some questions arise to be answered. First, whether all the aspects of teaching are equally important for students and whether students' attitudes are homogeneous, i.e. what is the structure of their preferences toward different aspects of teaching? Second, if there is heterogeneity of preferences, in which way and to what extent it influences the overall evaluation of teacher as well as his rank. Did averaged scores hide the real situation?

An appropriate approach to elicit students' preferences is conjoint analysis, a multivariate technique used to understand how an individual's preferences are developed. It originated in mathematical psychology (Luce and Tukey 1964), and was first introduced in marketing research to evaluate consumer preferences for hypothetical products and services (Green and Rao 1971). Nowadays, it is widely used for designing optimal products and product lines (Kuzmanovic and Martic 2012), to understand the preferences in various markets including retail (Kuzmanovic and Obradovic 2010), transportation (Hensher 2001), telecommunication (Kim 2004; Kim et al. 2008). It is also used in the health care field for eliciting patients' and community's preferences for health services (Kuzmanovic et al. 2012; Ryan and Ferrar 2000).

However, only a few studies have used the conjoint analysis within the education industry. Soutar and Turner (2002) used conjoint analysis and cluster analysis to suggest a better university education system for students. Hur and Pak (2007) attempted to identify the preferred subjects for an after-school computer education course in elementary schools by means of conjoint analysis. Sohn and Ju (2010) used conjoint analysis to assign the weights of those four components to effectively recruit the science high school students who have a high quality. Kim, Son, and Sohn (2009) used conjoint analysis to determine the most influential attributes of English Medium Instruction (EMI) classes, and to develop a customized EMI class system for university students whose mother tongue is not English.

In this paper we propose a new approach to student evaluation of teaching which takes into account the students' preferences for various aspects of teaching, as well as possible heterogeneity of their preferences. Using illustrative example we will confirm efficiency of proposed approach in practice. The approach is expected to serve as a practical guideline that, subject to possible modifications in the future, may become an essential tool used by universities for their continuous improvement programs.

The paper is organized as follows. Section 2 addresses some basic concepts of conjoint analysis, including the procedure for determining the criteria importance. The proposed methodological framework of student evaluation of teaching based on conjoint data is also given in this section. The proposed approach to evaluation of teaching is tested through the illustrative example given in Section 3. Section 4 provides the concluding remarks that summarize the most important contributions of the paper.

2. Theoretical framework

2.1. Conjoint analysis

Conjoint analysis, sometimes called ‘trade-off analysis’, reveals how people make complex judgments. The technique assumes that complex decisions involve not only one factor or criterion, but rather several factors ‘considered jointly’. It is based on the simple premise that consumers evaluate the value of a product or service by combining the separate utilities provided by each product attribute.

Conjoint experiments involve individuals being asked to express their preferences for various experimentally designed, real or hypothetical alternatives. These hypothetical alternatives are descriptions of potential real-world alternatives, in terms of their most relevant features or attributes (both quantitative and qualitative). Lists of attributes describing single alternatives are called profiles or concepts. Typically, the set of relevant attributes is generated by expert opinions, reviewing the research literature and performing pilot research with techniques such as focus groups, factor listings, or repertory grids. Two or more fixed values, or “levels”, are defined for each attribute, and these are then combined to create different profiles. The experimental procedure involves profiles being presented to respondents who are asked to express their preferences by rating or ranking these profiles.

Having collected the information on individual preference, the responses need to be analyzed. The simplest and most commonly used utility model is the linear additive model. This model assumes that the overall utility derived from any combination of attributes of a given good or service is obtained from the sum of the separate part-worths of the attributes. Thus, respondent *i*’s predicted conjoint utility for profile *j* can be specified as follows:

$$U_{ij} = \sum_{k=1}^{K} \sum_{l=1}^{L_k} \beta_{ikl} x_{jkl} + \varepsilon_{ij}, \; i = 1,...,I, \;\; j = 1,...,J, \qquad (1)$$

where I is the number of respondents; J is the number of profiles; K is the number of attributes; L_k is the number of levels of attribute *k*. β_{ikl} is respondent *i*’s utility with respect to level *l* of attribute *k*. x_{jkl} is such a (0,1) variable that it equals 1 if profile *j* has attribute *k* at level *l*, otherwise it equals 0. ε_{ij} is a stochastic error term.

The parameters β_{ikl} are estimated by a regression analysis. These beta coefficients, also known as part-worth utilities, can be used to establish a number of things. Firstly, the value of these coefficients indicates the amount of any effect that an attribute has on overall utility of the profiles– the larger the coefficient, the greater the impact. Secondly, part-worths can be used for preference-based segmentation. Namely, given that part worth utilities are calculated at the individual level, if preference heterogeneity is present, the researcher can find it. Therefore, part-worths can be used for preference-based segmentation. Respondents who place similar value to the various attribute levels will be grouped together into a segment. Segmentation of conjoint part-worths produces true "benefit segments". Widely used method for preference-based segmentation across industries is *k*-means cluster procedure. Thirdly, part-worths can be used to calculate the relative importance of each attribute, also known as an importance value. These values are calculated by taking the utility range for each attribute separately, and then dividing it by the sum of the utility ranges for all of the attributes. Thus the relative importance that *i*th respondent assigned to the attribute *k* is given by

$$W_{ik} = \frac{\max\{\beta_{ik1}, \beta_{ik2}, ..., \beta_{ikL_k}\} - \min\{\beta_{ik1}, \beta_{ik2}, ..., \beta_{ikL_k}\}}{\sum_{k=1}^{K}\left(\max\{\beta_{ik1}, \beta_{ik2}, ..., \beta_{ikL_k}\} - \min\{\beta_{ik1}, \beta_{ik2}, ..., \beta_{ikL_k}\}\right)}, \quad i = 1, ..., I, \; k = 1, ..., K \qquad (2)$$

These individual level impedances are then averaged to include the respondents with similar preferences. Thus, the importance of attribute *k* in segment *s* is given by:

$$W_{ks} = \frac{1}{I_s}\sum_{i=1}^{I_s} W_{ik}, \quad k = 1, ..., K, \; s = 1, ..., S, \qquad (3)$$

where I_s is the number of respondents from the segment *s*.

Part-worth utilities can be further used to obtain overall utility values for all possible combination of attribute levels. It will be done by inserting the appropriate part-worths into equation 1.

2.2. Methodological framework of proposed approach

The approach proposed here for the evaluation of teachers by students, includes two-part survey. The methodological framework of proposed approach is given in Figure 1.

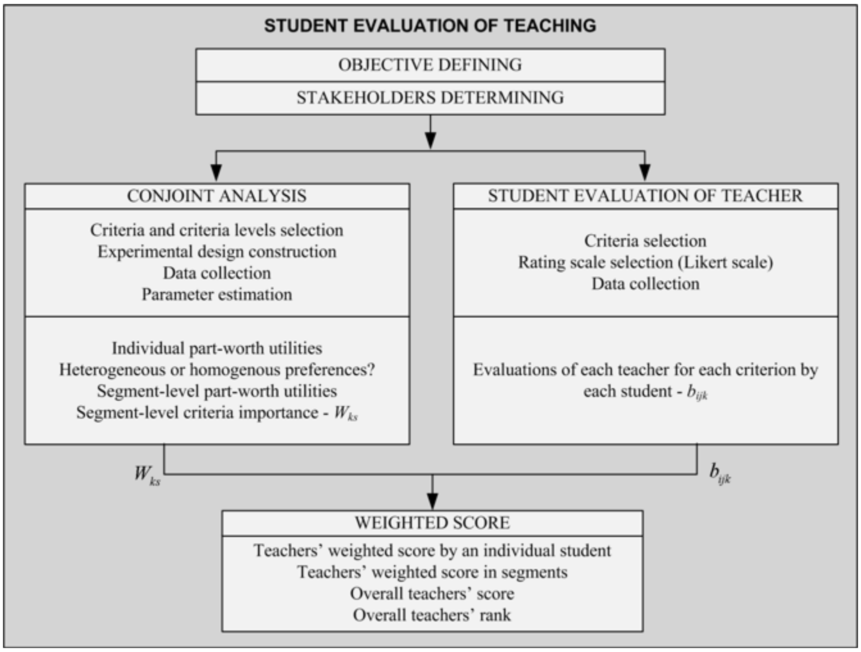

Fig. 1. Methodological framework of proposed approach

One aspect of procedure involved using conjoint analysis to determine the preferences of individual students according to specific attributes (here denotes as criteria), as well as criteria importance. Whether the population is heterogeneous is determined, as well as the extent of heterogeneity, thus segments of different preferences are insulated, and criteria importance in those segments are determined. Along with that, conventional evaluation of teachers by students is done, with the Likert scale, where students evaluate each teacher on every presented criterion. After collecting data on preferences and evaluations, they are summarized in order to formulate a comprehensive evaluation of the efficiency of teachers (here denoted as Weighted Score).

These two studies can be conducted simultaneously on the same sample of students, or if it is not possible from any reason, they can be done as two independent studies. In the second case, in order to create comprehensive evaluation of the teachers' efficiency, aggregated data should be used.

Because not all criteria are equally important for all students or segments, it is suggested that criteria importance are used as weighting factors (weights), i.e. to multiply scores assigned to individual teachers by the value of criterion importance, to produce overall score. Formally, if *i*th student belongs to *s*th segment, then the weighted score of the teacher *j*, by the student, equals:

$$WSI_{ij} = \sum_{k=1}^{K} W_{ks} b_{ijk}, \quad i = 1,...,I_s, s = 1,...,S, \; j = 1,...,J \;, \tag{4}$$

where W_{ks} is the importance of the *k*th criterion in the segment *s*, and b_{ijk} is a score by the *i*th student assigned to *j*th teacher, judging by the *k*th criterion. In the case of a sample of homogeneous preferences, the weights are determined by the aggregated criteria importance values.

In addition to weighted scores of an individual student for each teacher, the weighted scores in all isolated segments (*WSS*) can be determined, as well as in general at population level. Weighted score of the *j*th teacher in segment *s* is calculated as follows:

$$WSS_{js} = \frac{1}{I_s} \sum_{i=1}^{I_s} \sum_{k=1}^{K} W_{ks} b_{ijk}, \quad s = 1,...,S, \; j = 1,...,J \,, \tag{5}$$

while the overall weighted score (*WSO*) of the teacher *j* at the course level i.e. the whole population level, is equal to:

$$WSO_j = \frac{1}{I} \sum_{s=1}^{S} \sum_{i=1}^{I_s} \sum_{k=1}^{K} W_{ks} b_{ijk}, \quad j = 1,...,J \tag{6}$$

Using the obtained scores, the teachers can be further ranked by any of the criteria, and comprehensive. It can be shown that in some situations the same teacher is extremely differently evaluated in isolated segments. If these segments are significantly different in size, this should be borne in mind when considering rank and effectiveness of teachers. This will be shown at illustrative example.

3. Illustration

3.1. Conjoint data

In order to illustrate the proposed methodology, and confirm the possibilities of its application for student evaluation of teaching we used the conjoint data previously collected. The list of criteria and their levels used in the study as well as mean part-worths assigned to each of the criteria levels are given in Table 1, while the importance scores of attributes are shown on Figure 2.

It can be notice that the most important criterion on aggregate level is A1 (Clear and understandable presentation), followed by A2 (Methodical and systematic approach), A3 (Tempo of lectures) and A9 (Teacher availability). These four criteria cover almost 60% of the total importance. However, the question is whether these results reflect the student preferences truly. That is, if all the students in the sample most prefer clarity of the teacher's presentation, or whether they less prefer teacher's preparedness for lecture? And whether it is for all students the teacher-student relationship negligible significant?

Table 1. Criteria, corresponding levels and part-worth utilities (aggregated and segment level)

No	Criteria	Criteria levels	Part-worth utilities		
			Aggregate	Segment 1 n=53 (37%)	Segment 2 n=90 (63%)
A1	Clear and understandable presentation	Yes	0.836	0.46	1.05
		No	-0.836	-0.46	-1.05
A2	Methodical and systematic approach	Yes	0.692	0.46	0.83
		No	-0.692	-0.46	-0.83
A3	Tempo of lectures	Moderate	0.446	0.32	0.52
		Too fast	-0.123	-0.26	-0.04
		Too slow	-0.324	-0.06	-0.48
A4	Preparedness for a lecture	Good	0.279	0.14	0.36
		Poor	-0.279	-0.14	-0.36
A5	The accuracy of arrival to the lecture	On time	0.309	0.32	0.30
		Late	-0.309	-0.32	-0.30
A6	Encouraging students to participate in classes	Yes	0.316	0.54	0.18
		No	-0.316	-0.54	-0.18
A7	Informing students about their work	Yes	0.335	0.28	0.37
		No	-0.335	-0.28	-0.37
A8	Considering student comments and answering questions	Yes	0.316	0.46	0.23
		No	-0.316	-0.46	-0.23
A9	Availability (through individual teacher/student meetings or via e-mail)	Always available	0.220	0.31	0.17
		Mostly available	0.041	-0.19	0.18
		Mostly unavailable	-0.261	-0.11	-0.35

A more detailed analysis revealed heterogeneity in student preferences. Therefore, a cluster analysis was performed to classify respondents into more homogeneous preference groups.

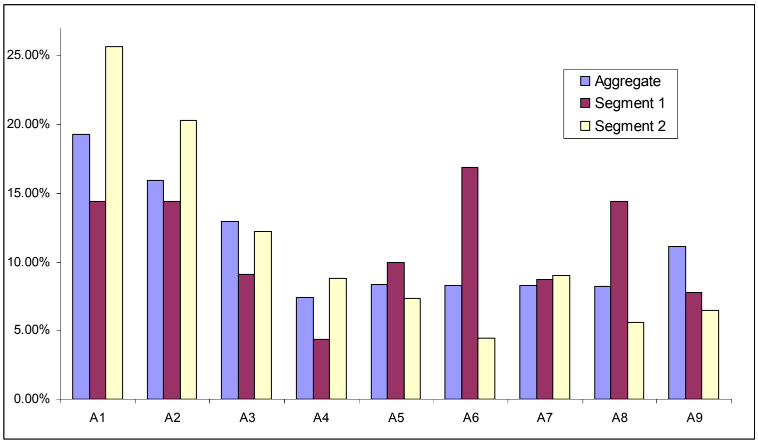

Figure 2. Aggregated and segment level criteria importance

The most important criterion to segment 1 is A6 (Encouraging students to participate in classes), followed by three equally important criteria, A1, A2 and A8 (see Fig. 2), while the least important criteria in this group of student is A4 (Preparedness for a lecture). The criterion with the greatest importance in segment 2 is A1 (Clear and understandable presentation), followed by the criterion A2 (Methodical and systematic approach). Other criteria are much less important. It is interesting that the criterion with the highest importance value in the first segment, A6, is by far in last place in this (second) segment.

3.2. The impact of heterogeneity of students' preference on the overall evaluation of teaching

Given that the survey revealed heterogeneity in students 'preferences for different aspects of teaching we analyzed its effects on the overall evaluation of certain teacher. For that purpose, Table 3 provides ratings of 20 teachers from a randomly selected *i*th student (columns A1 to A9). Ratings are on a scale from 1 to 5, with 1 being the lowest and 5 the highest score for given criterion. The analysis is performed first under the assumption that preferences are taken into account, but at the aggregate level, then the assumption is the student belongs to segment 1, and finally, that he belongs to segment 2. These results (Weighted Scores), as well as evaluations provided as average score by all criteria (Mean Score) are shown in Table 3. For both measures, corresponding ranks of teachers are given in the brackets.

Table 3. Teachers' evaluations by randomly selected student (using different evaluation measures)

	Scores per criteria									**Mean score**	**Weighted scores (WSI)**		
Teacher	A1	A2	A3	A4	A5	A6	A7	A8	A9	Value	Aggregate	Segment 1	Segment 2
Teacher 1	5	5	5	5	4	4	5	5	5	4.78 (1)	4.83 (1)	4.73 (2)	4.88 (1)
Teacher 2	4	5	5	5	4	5	5	5	5	4.78 (1)	4.72 (3)	4.76 (1)	4.67 (4)
Teacher 3	4	4	5	5	5	5	4	5	5	4.67 (3)	4.56 (4)	4.63 (3)	4.45 (7)
Teacher 4	5	5	5	5	5	4	4	4	5	4.67 (3)	4.75 (2)	4.60 (4)	4.81 (2)
Teacher 5	5	5	5	5	4	4	4	4	4	4.44 (5)	4.56 (5)	4.42 (6)	4.67 (3)
Teacher 6	4	4	4	4	4	5	5	5	5	4.44 (5)	4.36 (8)	4.48 (5)	4.26 (8)
Teacher 7	3	4	5	5	3	4	5	5	5	4.33 (7)	4.20 (10)	4.20 (9)	4.09 (11)
Teacher 8	5	5	5	4	4	4	4	3	5	4.33 (7)	4.51 (6)	4.31 (8)	4.59 (5)
Teacher 9	5	5	5	5	3	3	4	3	5	4.22 (9)	4.42 (7)	4.09 (10)	4.56 (6)
Teacher 10	4	5	3	3	4	5	4	5	5	4.22 (9)	4.23 (9)	4.40 (7)	4.16 (9)
Teacher 11	5	5	5	4	3	1	1	1	2	3.00 (16)	3.43 (14)	2.92 (17)	3.81 (12)
Teacher 12	5	5	4	4	5	1	1	1	1	3.00 (16)	3.36 (16)	2.95 (16)	3.77 (13)
Teacher 13	3	2	5	4	1	4	1	2	5	3.00 (16)	3.06 (18)	2.89 (18)	2.92 (19)
Teacher 14	3	3	3	3	4	4	4	4	5	3.67 (11)	3.55 (12)	3.66 (12)	3.39 (16)
Teacher 15	3	3	3	4	5	3	4	3	5	3.67 (11)	3.55 (13)	3.49 (14)	3.46 (15)
Teacher 16	2	2	2	4	5	4	5	4	5	3.67 (11)	3.31 (17)	3.51 (13)	3.06 (18)
Teacher 17	5	5	4	3	3	3	3	3	4	3.67 (11)	3.95 (11)	3.74 (11)	4.11 (10)
Teacher 18	1	1	1	5	5	5	5	5	1	3.22 (15)	2.62 (20)	3.18 (15)	2.41 (20)
Teacher 19	5	5	5	1	1	1	1	1	5	2.78 (19)	3.38 (15)	2.83 (19)	3.59 (14)
Teacher 20	5	5	5	1	1	1	1	1	1	2.33 (20)	2.93 (19)	2.51 (20)	3.33 (17)

Based on the results shown in Table 3, it can be seen that discriminatory power of the "Mean score" approach is very poor. Some teachers with qualitatively different characteristics achieve the same mean score value. For example, four teachers have the same mean score of 3.67, i.e. Teacher 14, 15, 16 and 17, while there is vast rating diversity according to criteria, what could be seen from table 3.

On the other hand, if the criteria importance derived from conjoint analysis is used as criteria weights (either at the aggregate level or at segment level) much better results are obtained (see Table 3). Discriminatory power of this approach, even viewed at the aggregate level, is much better. The ranks of teachers are clearly defined, and there are not teachers with the same rank. In addition, the ranks are different, some are lower and some higher than in the case of Mean Score approach. This makes sense because better rating of more importance criterion increases overall assessment score more than better rating of less criterion importance.

It can be further seen that the overall scores and ranks, for the same ratings, are significantly different from segment to segment. This is especially obvious in Teachers 9, 11, 14 and 18. This information may be of particular importance when the segments are of substantially different sizes, or when their preferences are significantly different. Simplification could hide substantially important information, and the results would be distorted image of reality.

We performed another analysis and compared the results. We analyzed overall ratings for 8 teachers given by 10 students. These are generated (hypothetical) scores for hypothetical teachers, because in conducting conjoint analysis, conventional evaluation of specific teachers is not conducted.

According to the identified segment sizes, we also assumed that 4 out of 10 students belong to the segment S1, while remaining 6 belong to the segment S2. Table 4 presents the standard summary measures derived from two teaching evaluation instruments: the conventional teaching evaluation instrument (Mean Score) and proposed approach based on conjoint data (Weighted Score). Corresponding teachers' ranks are given in the brackets.

Table 4. Summary measures for eight teachers (ranks are in the brackets)

	Conventional approach (Mean Score)	Proposed approach (Weighted Scores)		
Teacher	Overall	Segment 1	Segment 2	Overall
1	4.49 (1)	4.68 (1)	4.39 (3)	4.51 (2)
2	4.49 (1)	4.25 (4)	4.62 (1)	4.47 (3)
3	3.20 (8)	3.70 (6)	3.40 (8)	3.52 (7)
4	4.32 (3)	4.55 (2)	4.50 (2)	4.52 (1)
5	3.92 (4)	4.29 (3)	3.99 (4)	4.11 (4)
6	3.64 (6)	3.49 (8)	3.52 (7)	3.51 (8)
7	3.62 (7)	3.88 (5)	3.70 (6)	3.77 (5)
8	3.92 (4)	3.51 (7)	3.82 (5)	3.69 (6)

As in the previous analysis, data given in Table 4 indicate poor discriminatory power of the conventional approach based on the averaging ratings by criteria and at the level of the sample, without taking into account student preferences. Consider for example Teacher 5, whose Mean score value equal to 3.92. If we take into account the heterogeneity of preferences, and if these preferences are incorporated into an overall score, weighted score for Teacher 5 in segment 1 is 4.29, in segment 2, this value is 3.99, and the Overall score is 4.11. It may be noted that ratings, both at the segment and sample level, are significantly higher than the Mean score, so it can be concluded that this teacher is underestimated if preferences are not taken into account.

On the other hand, consider Teacher 8. His Mean Score value is also 3.92, but the teacher was, in both segments and the overall rated lower by Weighted Score approach (3.51, 3.82 and 3.69 respectively). In other words, the teacher may be overestimated if student preferences and their heterogeneity are not taken into account. A similar situation can be observed in the case of some other teachers.

Let us now consider the ranks of teachers obtained by the previously discussed two approaches, and compare them. Consider first Teacher 4. He is third-ranked according to conventional approach, but first-ranked according to approach we proposed. The same teacher is better ranked in both segments (rank is 2) than according to conventional approach. In the case of Teacher 2, situation is reversed. He is first-ranked according to conventional approach, but third-ranked in our approach. However, in this teacher case, particularly important are information on the rank in the segment. In fact, this teacher was at fourth place in segment 1, while first-placed in segment 2. A similar analysis could be done for other teachers.

Table 5. Correlation between evaluation measures

	Weighted Score (Seg. 1)	Weighted Score (Seg. 2)	Weighted Score (Overall)	Mean Score
Weighted Score (Seg. 1)	1			
Weighted Score (Seg. 2)	0.714	1		
Weighted Score (Overall)	0.929	0.881	1	
Mean Score	0.649	0.918	0.761	1

Table 5 provides correlation between measures previously discussed. Correlation of ranks in Segment 1 and Segment 2 is quite low (0,714), which makes sense, because these are groups of students with different preferences. It is also notable that correlation of ranks by the conventional approach and our approach (weighted score) is 0,761.

4. Conclusions

The paper proposes a new approach to student evaluation of university teaching based on data from conjoint analysis. Proposed approach includes two-part study. The first part involves the use of conjoint analysis in order to determine the students' preferences toward specific aspect of teaching, as well as importance of those aspects. It also determines whether and to what extent the population is heterogeneous. Accordingly, the groups of students with similar preferences are identifying. The second part of the study includes the conventional evaluation of teachers by students. Using the Likert scale students evaluate each teacher by each of the specified aspects. After the data on preferences and ratings are obtained, they are summarized in order to formulate the comprehensive evaluation of the teachers' efficiency.

Using real conjoint data, we tested proposed approach. First, we show that the proposed approach has better discriminator power than conventional approach to teaching and teacher evaluation. But it also offers the possibility of further deeper analysis of student preferences, which could be used to get a better idea of true picture, and assess more objectively the efficiency and effectiveness of teachers. It is possible to take into account the heterogeneity in the creation of a comprehensive evaluation.

Even in situations where the research is carried out separately, it is possible to identify the aggregate preferences of students based on conjoint analysis, and using those data to calculate weighted score for each teacher. This has proved the usage power of our approach in practice.

References

Abrami, P. C. & d'Apollonia, S. (1991). Multidimensional students' evaluations of teaching effectiveness; The generalizability of "N = 1" research: Comment on Marsh (1991). *Journal of Educational Psychology*. 83, 411-415.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching* and *Learning in Higher Education*, 17(1), 48-62.

Cashin, W.E., & Downey, R.G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology*, 84, 563–572.

Davies, M., Hirschberg, J. G., Lye, J. N., Johnston, C., & McDonald, I. M. (2007). Systematic influences on teaching evaluations: The case for caution. *Australian Economic Papers*, 46(1), 18-38.

Ellis, L., Burke, D. M., Lomire, P., & McCormack, D. R. (2003). Student grades and average ratings of instructional quality: The need for adjustment. *The Journal of Educational Research*, 97(1), 35-40.

Green, P.E., & Rao, V.R. (1971). Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research*, 8, 355-363.

Haddad, Y., Haddad, J., Olabi, A., Shuayto, N., Haddad, T., & Toufeili, I. (2007). Mapping determinants of purchase intent of concentrated yogurt (Labneh) by conjoint analysis. *Food Quality and Preference*, 18, 795-802.

Hensher, D. (2001). The valuation of commuter travel time savings for car drivers: evaluating alternative model specifications. *Transportation*, 28, 101–118.

Hur, J. S., & Pak, R. J. (2007). Conjoint analysis for the preferred subjects of elementary school computer education. *Journal of the Korean Data and Information Science Society*, 18(2), 357-364.

Kim, A., Son, Y.D., & Sohn, S.Y. (2009). Conjoint analysis of enhanced English Medium Instruction for college students. *Expert Systems with Applications*, 36, 10197–10203.

Kim, C., Choe, S., Choi, C., & Park, Y. (2008). A systematic approach to new mobile service creation. *Expert Systems with Applications*, 35, 762–771

Kim, Y. (2004). Estimation of consumer preferences on new telecommunications services: IMT-2000 service in Korea. *Information Economics and Policy*, 17(1), 73-84.

Kuzmanovic, M., & Martic, M. (2012). An approach to competitive product line design using conjoint data. *Expert Systems with Application*, 39(8), 7262-7269. doi:10.1016/j.eswa.2012.01.097.

Kuzmanović M., & Obradović, T. (2010). The role of conjoint analysis in the new product price sensibility research. Management – časopis za teoriju i praksu menadžmenta, 15(54), 51-58.

Kuzmanovic, M., Vujosevic, M., & Martic, M. (2012). Using Conjoint Analysis to Elicit Patients' Preferences for Public Primary Care Service in Serbia. *HealthMED*, 6(2), 496-504.

Liaw, S.-H., & Goh, K.-L. (2003). Evidence and control of biases in student evaluations of teaching. *The International Journal of Educational Management*, 17(1), 37-43.

Luce R. D., & Tukey J. W. (1964). Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. *Journal of Mathematical Psychology*, 1, 1-27.

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319–384). New York, NY: Springer.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist*, 52, 1187–1197. doi:10.1037/0003-066X.52.11.1187

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workloads on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, 92, 202–228.

Marsh, H. W., Ginns, P., Morin, A. J. S., Nagengast, B., & Martin, A. J. (2011). Use of Student Ratings to Benchmark Universities: Multilevel Modeling of Responses to the Australian Course Experience Questionnaire (CEQ). *Journal of Educational Psychology*, 103(3), 733–748.

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52, 1218–1225. doi:10.1037/0003-066X.52.11.1218

Ryan M., & Farrar S. (2000). Using conjoint analysis to elicit preferences for health care. *British Medical Journal*, 320(7248), 1530-1533.

Sohn, S.Y., & Ju, Y.H. (2010). Conjoint analysis for recruiting high quality students for college education. *Expert Systems with Applications*, 37, 3777–3783

Soutar, G. N., & Turner, J. P. (2002). Students' preferences for college: A conjoint analysis. *International Journal of Educational Management*, 16(1), 40–45.

Witte, K. D., & Rogge, N. (2011). Accounting for exogenous influences in performance evaluations of teachers. *Economics of Education Review*, 30(4), 641-653. doi:10.1016/j.econedurev.2011.02.002.