



This journal provides immediate open access to its content under the [Creative Commons BY 4.0 license](https://creativecommons.org/licenses/by/4.0/). Authors who publish with this journal retain all copyrights and agree to the terms of the above-mentioned CC BY 4.0 license

DOI: 10.2478/seeur-2020-0013

QUALITATIVE ANALYSIS OF STUDENTS' EVALUATIONS – THE CASE OF THE SOUTH EAST EUROPEAN UNIVERSITY (SEEU) IN NORTH MACEDONIA

MSc. Irena Gjasimovska,
Max van der Stoel Institute,
South East European University,
Tetovo, North Macedonia
i.gjasimovska@seeu.edu.mk

Acad. Prof. Dr. Abdylmenaf Bexheti,
Faculty of Business and Economics,
South East European University,
Tetovo, North Macedonia
a.bexheti@seeu.edu.mk

Assoc. Prof. Dr. Veronika Kareva,
Faculty of Languages, Cultures and Communication,
South East European University,
Tetovo, North Macedonia
v.kareva@seeu.edu.mk

Assoc. Prof. Dr. Gadaf Rexhepi,
Faculty of Business and Economics,
South East European University,
Tetovo, North Macedonia
g.rexhepi@seeu.edu.mk

ABSTRACT

The use of student evaluations of teaching (SET) has become a widespread practice in higher education despite inconclusive evidence reported in the literature around its validity. Not surprisingly, the question of the validity of SET continues to be a current debate in higher education, pointing to the need for more research in this area. This paper is a part of a larger scale study, which aims to contribute to broadening the knowledge and understanding of SET validity by analysing the process within the South East European University (SEEU) in North Macedonia in order to determine whether student evaluations are objective and critical. A likert scale questionnaire, containing 9 questions, was designed for the purpose of the analysis. The questionnaire was sent to all students from the five (5) Faculties: Business and Economics, Law, Contemporary Sciences and Technologies, Contemporary Social Sciences and Faculty of Languages, Cultures and Communications, in both campuses, Tetovo and Skopje. Three hundred and thirty three (333) students participated in the survey. Statistical Package for Social Science (SPSS) was used for analysing the results. Findings revealed that the information students received about the reputation, experience and qualifications of the professors had the highest influence on their perceptions, which in turn influenced the evaluations. As an addition, the present paper also compares two methods on a data set of actual SET. For illustrative purposes, only data from one faculty have been analysed. It is shown that the traditional method of considering the average values can misrepresent a teacher's performance as it can be highly sensitive to any extreme grades, being either very positive or very negative.

Key words: Student evaluations, SEEU, higher education, data analysis.

INTRODUCTION

The effectiveness of student evaluation of teaching (SET) is one of the most researched areas in education and yet, there is still no agreement on the usefulness of this process. SET has been used in educational institutions around the world as a means of providing feedback about the quality of teaching. Nowadays, it is one of the most widespread instruments used to inform teachers and university management about the quality of provision in an institution (Wilson, 1986). Its necessity appears due to the fact that one of the goals of universities and other educational institutions is to provide the students with the relevant knowledge, skills and instruction to maximize their learning. This is why so much emphasis is placed on ensuring effective teaching year after year. Good teaching cannot happen without student learning and without good course administration. Therefore, effective teaching includes teachers, students,

the curriculum, teaching methods, assessment procedures, the climate created through interactions, and the institutional climate. The curriculum, the teaching methods, and assessment procedures are particularly important for the quality of the teaching process. Thus, student evaluations are considered as an effective measure that provide information about all above-mentioned components of good teaching and learning and contribute to quality assurance (QA).

QA in higher education

Quality is a broad notion and it becomes especially complex when referred to in higher education context. According to Pavel (2012), “the quality in higher education is a multi-dimensional, multilevel, and dynamic concept that relates to the contextual settings of an educational model, to the institutional mission and objectives, as well as to the specific standards within a given system, institution, programme or discipline” (Pavel, 2012, pg. 124). Teachers and their effectiveness represent just one segment contributing to quality. There can be no single definition of the effective or good quality teacher. The definition of effective teaching is different from the points of view of teachers, students and administration. The perception of effective teaching varies with the age of the student population, the background of the students etc. With this regard, the same author states that, “there has always been an individual and social need to improve the quality of peoples’ life, including the quality of what they learn over many years of schooling, how they learn it, and why they learn it” (Pavel, 2012, pg. 124). Nonetheless, for administrators, teachers and students, it is important to be sure that teaching is effective. Therefore, they try to find an appropriate measure of such a multidimensional process that combines different aspects of teaching. For both teachers and administrators it is very important to understand what kind of students’ opinion is presented by the teacher/course evaluation results. If the bias exists, it should be taken into account when analyzing SET results. The evaluations are intended to be a tool for quality assurance for the teachers and the university management.

QA system at SEEU

SEEU was founded in 2001, as a private public, not-for-profit higher education institution with strong international support and compared to the other universities in the country it had the advantage to organize its academic offer based on the main commitments resulting from the Bologna process: three year cycle degree structures, recognition of qualifications, quality assurance, learning and teaching, internationalization and employability. The University has a

well developed quality structure with mechanisms for quality enhancement and monitoring. There exist different instruments for getting student feedback and ensuring student-centered learning and teaching. The following instruments should be mentioned:

- Teaching Observation
- Student Evaluation
- Student Achievement
- Research Activities
- Academic Staff Digital Profile

The University is putting a lot of efforts in further development of its institutional QA system. As pointed out in a recent study by Kareva, Bexheti and Rexhepi (2019), facing major demographic changes and significant decrease of student population paired up with the existence of state funded universities require maintaining sustainability only through quality. This is why it is very important to test the existing QA system and be sure about the future activities with this regard.

THE PROBLEM

SET is one of the tools for measuring the quality of teaching at SEEU. The aim of this paper is to investigate whether individual student evaluations and thus the process in general are influenced by previously created perceptions about the academic staff or whether they are critical and objective. Through this qualitative analysis we attempt to understand if the behavior, attitudes, values and beliefs of the students, based on previously formed opinion, impact their evaluation or it is not a subject to previously created perceptions. The paper also considers how the results can be used for future improvement of the course/teacher evaluation system in general and how it can be used for various decision-making processes. Moreover, the paper contributes to the debate on the validity of student evaluations by exploring the relationship between students' perceptions and their impact on student evaluations of academic staff. The hypothesis is:

1. Students' evaluations are influenced by previously created perception of the academic and administrative staff.
- 1.1. Students' evaluations are objective and critical.

For this purpose, besides the SET from the South East European University for the academic year 2018/2019, an additional questionnaire was designed. We also believe it is important to

mention that this requires further studies in this area, especially given the prevalence of online evaluations. In addition, it should be noted that the evaluations were voluntary, and therefore, have some self-selection bias.

It is expected that the findings of this analysis will contribute to better understanding of student behavior and their way of thinking, which will further lead to more informed decision making important for successful functioning of the University in question (SEEU). Since SET is one of the most relevant instruments indicating the quality of teaching and as such it is widely used all over the world, other universities might also benefit from these findings and take them into considerations when interpreting individual teachers' evaluation results.

The rest of the paper is organized as it follows. The following section presents the findings from the relevant literature, the complex process of decision-making and the potential of bias in the current evaluations as well as the factors that influence those evaluations. A description of the methods used and the results of the analysis are presented in the fourth section. In the fifth section is presented a different way of analyzing student evaluations. The final section summarizes the key findings.

LITERATURE REVIEW

Having in mind that the aim of this paper is to better understand students' behavior, their thoughts and feelings, it seems wise to enter the field of social psychology as a science which main interest is human social behavior and the causes of such behavior. □ Social psychology is the scientific study of how the thoughts, feelings, and behaviors of individuals are influenced by the actual, imagined, and implied presence of others. Social psychologists typically explain human behavior as a result of the relationship between mental state and social situation, studying the conditions under which thoughts, feelings, and behaviors occur and how these variables influence social interactions.

The psychology literature identifies some difficulties related to student evaluations (Hackert, et. al. 2006; Keeley, et. al. 2013) which we can group in three categories. First, the behaviors that most influence these evaluations are rooted in physiology, personality, habit and culture. Those behaviors are difficult for any faculty member to change, and they often reflect characteristics like race, gender, nationality, or socioeconomic class. Second, the current evaluation process allows social stereotypes to filter students' perceptions of instructor behaviors; students see the behaviors of some faculty members differently than they view the

identical behaviors in other professors. These stereotypes further disadvantage faculty based on categories like race or gender. Finally, the ratings that students are making through the evaluation system that is commonly used bear little relationship to objective measures of learning (Cf. Clayson & Sheffet, 2006; Audesh K. Paswan & Joyce A., 2002). This implies that the way students are conducting the evaluations at contemporary universities, leads to rewarding faculty according to relatively trivial indicia of who they are, or seem to be, rather than what they do or accomplish in the classroom.¹

The cognitive foundation of judgement

As the Nobel Prize winning psychologist Daniel Kahneman (2003) explains, there are two types of human thought processes: System 1 and System 2. “System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control” (Kahneman, 2011, pg. 21). “System 2 allocates attention to the effortful mental activities that demand it, including complex computations. The operations of System 2 are often associated with the subjective experience of agency, choice, and concentration.” (Kahneman, 2011, pg. 21)

As rational adults, we assume that most of our judgments come from System 2; we perceive ourselves as rational creatures who think things through. However, almost all judgments begin as System 1 intuitions; we have gut reactions almost about everything. System 2 deliberations challenge and override our initial System 1 decisions in an amazingly limited number of those cases.² Instead, we are more likely to use System 2 cognitive processes to justify the intuitive judgments we make using the more emotional leaps of System 1.

The dominance of System 1 thinking offers a possible explanation for the high correlation between teaching evaluations and very quick judgments with minimum information of the same teachers. Students’ end-of-semester evaluations may not just correspond to those first impressions; they may be those impressions. Because the process used to collect evaluations does not encourage System 2, the students’ brains most likely retrieve their first impressions and report those impressions as their current assessment. Obtaining more meaningful evaluations from students, therefore, requires finding ways to involve System 2 thinking. There

¹ These flaws in the evaluation system do not mean that faculty members who receive high evaluations are bad teachers. A better evaluation process, more focused on student learning, might identify many of the same teachers as effective.

² Although it is impossible to know precisely how often System 2 overrides System 1 in daily life, Kahneman estimates that the “most common” paths of decision making are those that involve only System 1 or those in which System 2 only partially corrects System 1

are at least three reasons why the current evaluation system fails to encourage that system 2 thinking or deeper thought.

First, System 1 works with unusual efficiency in deciding whether things are good or bad. Kahneman (2003) quoting Joseph E. LeDoux (2000) suggests that the brain has specialized neural circuitry that renders these good/bad distinctions quickly and confidently. Therefore, when the brain is asked to make any good/bad judgment, it is less likely to invoke System 2's more reflective thinking because System 1 handles these decisions very efficiently. This does not mean that the brain invariably makes correct good/bad decisions using System 1. Evaluating a professor's teaching requires the brain to make good or bad distinctions especially when there are numerical judgments involved by most evaluation processes demanded from students. The very nature of these forms, which ask students to evaluate teacher/course effectiveness without requiring other types of reflection, discourages the brain from moving beyond System 1.

Second, System 2 thinking requires time. Our brains are good at intuitions, quick judgments and first impressions, but struggle to make more complex decisions. Humans have difficulty in processing multiple pieces of data because of the restricted working memory. To make thoughtful evaluations, the brain needs time to recall various bits of data, compare them, group pieces of information into larger chunks of partial judgments, and produce a reasoned response (Dijker & Koomen, 1996; Finucane et al., 2000; Sanbonmatsu & Fazio 1990). For students, evaluating a professors' teaching requires more complex thought than most faculty members realize. The professors shape their class preparation and class activities on a regular basis. Students, on the other hand, concentrate on the product of the professors' teaching rather than the process. They attend daily to what they are learning, but the underlying process is secondary. For students to make assessments, they must recall all the lectures, explanations, assignments, and other characteristics of a professor's work; examine those memories from a new perspective; and judge their efficiency. To do this well requires more time than five, ten, or even fifteen minutes. And finally, since the human brain has limited capacity, any cognitive distraction weakens System 2 thinking. Laboratory research shows that divided attention reduces reasoned thought, promoting reliance on intuitive, stereotyped, and other automatic System 1 processes. People have trouble thinking about two issues at once or quickly switching from one mental task to another. Students complete traditional evaluations under great cognitive load. They have just finished a challenging class; they may be filling gaps in their notes or maybe they are thinking about the upcoming exams or any other events with their

families. Therefore, the key to providing more meaningful evaluations of teaching lies in creating the circumstances that allow students to engage more reflective System 2 thought processes. Students need time and mental space to overcome intuitive System 1 judgments; they need mental prompts to help them review the semester's work; and also they need assistance in moving beyond reflexive good/bad judgments. This question is particularly significant given that the potential for bias has probably never been greater than it is today. This may be due to the circulation of a large amount of information among students, especially in a small urban study and living environment, prior to any interaction with the professor. Such information could distort subsequent perceptions and judgments.

Research on students' perceptions of SETs is limited (Gravestock and GregorGreenleaf, 2008). Some studies show that students think that the process of collecting students' feedback is useful and that students are valid evaluators of teaching. There is some evidence that students feel that evaluations have no effect on teacher performance and they also believe that faculty and administrators do not take their evaluations seriously (meta-study by Wachtel, 1998). Students do not know if anyone other than the professor sees the evaluations nor understand that ratings have an impact on personnel decisions. A study by Gaillard et al. (2006) shows that students are more likely to complete teacher/course evaluations if they understand how the evaluations are being used and believe that their opinions have an effect.

In response to the current concerns about the validity of the student evaluations, many researchers have investigated whether factors that are not related to teaching skills and course structure can explain the variability in ratings. There is much evidence that extraneous factors such as student characteristics, course characteristics or other environmental characteristics may influence how one student rates teachers/courses. Students are very different and some students' specific attributes can significantly affect how they rate their professors and courses. These characteristics can include gender, age, nationality, cultural/ethnic background, academic major, motivation, obtained grade, grade expectations, learning style, knowledge of essential material, students' interest in the subject matter prior to enrolling in the class etc. The mood of a student at the time of evaluation is another factor outside the control of the professor, and this may also have an impact on how the students rate a course and a professor. Based on a sample of 136 students, Munz and Munz (1997) found that student positive mood at the time of the evaluation accounted for only a modest 4% to 6% of the variance in the professor and course ratings. However, LaForge (2003) based on survey of 241 students, did not find the

correlation between professors' effectiveness ratings and students' mood measures to be significant.

The professors' characteristics like race, gender, experience, rank, weight may also influence the rate students give to their professors. In general, instructor age, rank and experience are not correlated with student ratings (Cashin, 1995). Marsh and Hocevar (1991) made a longitudinal study of student evaluations of the same professors across 13 years and found no systematic changes within teachers over time. Professors' entertainment level may positively influence the student ratings of the instruction. A study by Naftulin et al. (1973) concluded that professors' enthusiasm can influence student rating. Furthermore, different course characteristics like the size of the class, the difficulty of the course, course workload, whether the course is mandatory or elective, level of the class and time of day of the class may also influence student evaluations. Some environmental characteristics like the ambiance of the classroom should also be considered. To be fair to the professors, the effects of extraneous factors that are out of their control should be investigated and properly dealt with.

METHODOLOGY

Based on our hypothesis and in order to obtain a deeper insight in students' opinion a likert scale questionnaire was designed, containing 9 questions with 5 possible answers (Strongly disagree = 1, Disagree = 2, Neutral = 3, Agree = 4, Strongly agree = 5). The questionnaire was sent to all students in both campuses, Tetovo and Skopje, including all 5 faculties: Faculty of business and economics, Faculty of Law, Faculty of Languages, Cultures and Communication, Faculty of Contemporary Science and Communication, Faculty of Contemporary Sciences and Technologies. The number of students who fulfilled the questionnaire was 333. Statistical Package for Social Science (SPSS) was used for data analysis. Exploratory Factor Analysis using Factor analysis in SPSS was conducted. There is a strong relationship among variables found by Bartlett's test and samplings are adequate found by testing Kaiser-Meyer-Olkin (KMO).

Table 1 shows the Kaiser-Meyer-Olkin (KMO) test and Bartlett's Test of Sphericity. The standard measures of KMO returns are 0.0 to 0.49 is unacceptable, 0.50 to 0.59 is miserable, 0.60 to 0.69 is medium, 0.70 to 0.79 is middling, 0.80 to 0.89 is meritorious and 0.90 to 1.00 is marvelous. The KMO test result is 0.783 so sampling is acceptable. The Bartlett's test of Sphericity is 0.00 significance level indicating that factor analysis is useful for our data.

Table 1. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,783
Bartlett's Test of Sphericity	Approx. Chi-Square	700,283
	df	15
	Sig.	,000

Table 2 shows us the correlation between the variables. Therefore, the correlation between the information about the reputation, experience and qualifications of the professors ($P = 1$) is high with the opinion that is created about them based on the information obtained ($P = 2$) ($r = .675$), the opinion that is created based on the information obtained ($P = 2$) is highly correlated with the perceptions that students create about professors ($P = 3$) ($r = .710$). A strong correlation can be observed in other variables. Administrative staff (dean's office) ($P = 4$) and discussions with colleagues ($P = 5$) ($r = .552$), administrative staff (dean's office) ($P = 4$) and grades on midterms and other tasks ($P = 6$) ($r = .545$), discussions with colleagues ($P = 5$) and grades on midterms and other tasks ($P = 5$) ($r = .664$). The perceptions that students create about professors ($P = 3$) are strongly correlated with the administrative staff (dean's office) ($P = 4$) ($r = .582$), and with the discussions with colleagues ($P = 5$) ($r = .535$). There is also a moderate positive correlation between the administrative staff (dean's office) ($P = 4$) and the opinion that is created based on the information obtained ($P = 2$) ($r = .459$), the perceptions that students create about the professors. ($P = 3$) and grades on midterms and other assignments ($P = 6$) ($r = .473$) and information obtained about the reputation, experience and qualifications of professors ($P = 1$) ($r = .476$). There are no negative correlations between the variables.

The table 3 below shows us the Extraction Method in order to discover the principle components by reducing factors for further analysis together with the Scree Plot represent total variance within 2 components. Here among 6 dimensions, the model has been fitted with 2

Table 2. Correlation analysis

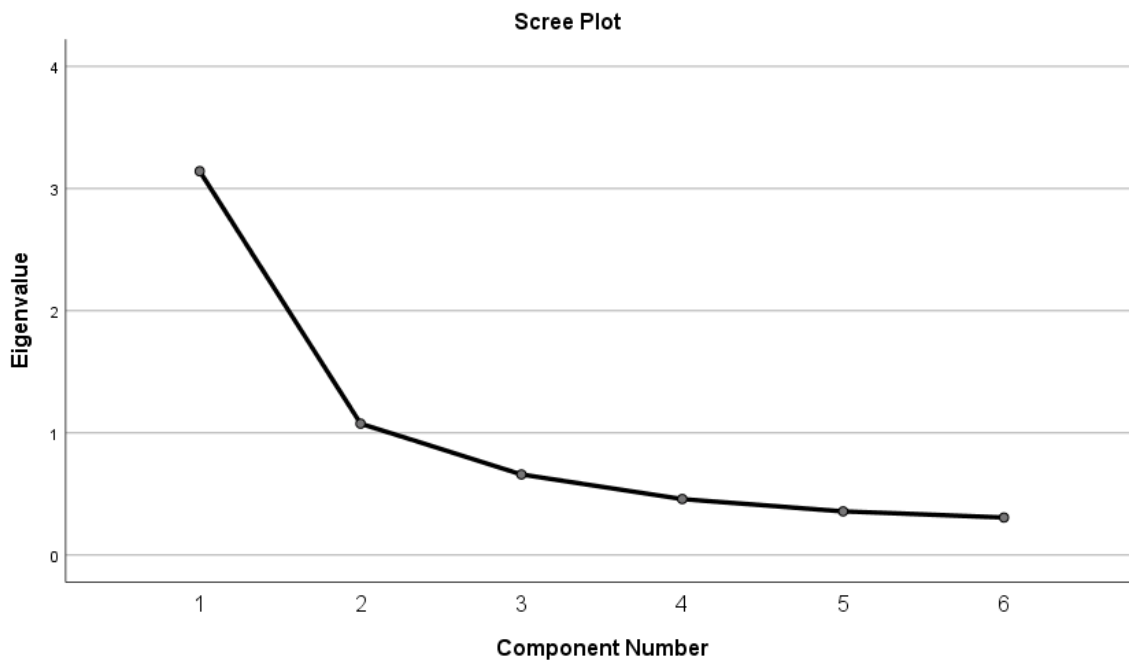
		Info. about prof. rep., exp. and quali	An op. based on the previous info.	The perception	Faculty admin. (Dean's office)	Discussion with the colleagues	Grades on the mid terms	The comm. with the students
Correlation	Info. about professors' rep., exp. and qual.,	1,000						
	An opinion based on the previous info.	,675	1,000					
	The percep. about the prof. infl. the eval.	,476	,710	1,000				
	Faculty admin.	,388	,459	,582	1,000			
	Discussion with the colleagues	,218	,389	,535	,552	1,000		
	Grades on the mid terms	,281	,374	,473	,545	,664	1,000	
	The comm. with the students outside classes	,307	,314	,362	,433	,416	,418	1,000
a. Determinant = ,040								

components by the analysis of factors. From table 3 it can be seen that 4 components have value of Eigen which is less than 1. From the column of “Extraction Sums of Squared Loadings” it can be seen that the loading for the components is 70,289%. In Scree Plot, component 1 is in the highest position for Eigenvalue (52.360%). Component 2 (17,929%).

Table 3. Extraction Method: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,142	52,360	52,360	3,142	52,360	52,360	2,456	40,939	40,939
2	1,076	17,929	70,289	1,076	17,929	70,289	1,761	29,351	70,289
3	,660	10,993	81,283						
4	,458	7,636	88,919						
5	,358	5,961	94,879						
6	,307	5,121	100,000						

Extraction Method: Principal Component Analysis.

Figure 1. Scree plot

The Scree plot and table 3 represent total variance within 6 components. Here among 13 dimensions, the model has been fitted with 2 components by the analysis of factors.

Table 4. Rotated Component Matrix

	Component	
	1	2
Discu. with the colleagues	,854	
Grades on the mid terms		
Faculty administration (Dean's office)		
The comm. with the students outside classes		
Information about prof. reput., exp. and qualif.		,916
An opinion based on the previous information		

From Table 4 it can be observed that there are 2 components that represent the highest possible rank positively out of 6 factors. Discussion with the colleagues is component 1 (.854), Information about professors' reputation, experience and qualifications is component 2 (.916). Here it was found that the information about professors' reputation, experience and qualifications has the highest influence on students' perceptions. Secondly, discussion with the colleagues influences students' perceptions.

Table 5. Correlations of Student perceptions and information about professors' reputation, experience and qualifications

			The perceptions about the professors influence the evaluations	Information about professors' reputation, experience and qualifications
Spearman's rho	The perceptions about the professors influence the evaluations	Correlation Coefficient	1,000	,523**
		Sig. (2-tailed)	.	,000
		N	330	329
	Information about professors' reputation, experience and qualifications,	Correlation Coefficient	,523**	1,000
		Sig. (2-tailed)	,000	.
		N	329	330

** Correlation is significant at the 0.01 level (2-tailed).

To test the hypothesis Spearman's Rho Correlation was made.

Table 5 presents the correlation between student perceptions and the information about professors' reputation, experience and qualifications. According to Spearman's Rho (Table 5), the information about teachers' reputation, experience, and qualifications has a strong positive correlation (.523) with students' perceptions, meaning that the information about teachers' reputation, experience, and qualifications mostly influences students' perceptions, which then affect their evaluations.

Student evaluations begin like most judgments (system 1 processes), meaning that they are influenced by first impressions and in this case the information about the professors' reputation, experience and qualifications. When the brain makes system 1 decision, he does not access all the information related to the subject collected over time. As a result, the students' brains most likely retrieve their first impressions and report those as their current evaluation. This dominance of system 1 thinking offers an explanation for the high correlation between student perceptions and the information about the experience, qualifications and reputation of professors. We believe it is also important to have in mind that one's reputation may invoke labeling and cause a halo error across not only ratings of distinct performance areas, but also extending to ancillary areas connected with the rate.

Finally, it should be borne in mind that individuals who are emotionally mature and who can stand firmly behind their views and opinions may be more rational in their evaluations rather than influenced by the information they have previously received.

DATA ANALYSIS: COMPARING MEANS AND PROPORTIONS

In addition, an interesting analysis of the performance of the professors is presented, a correct model for analysis of students' evaluations. It has been shown that the traditional method of analysis can lead to misinterpretation of professors' performance and can be extremely sensitive (McCullough & Radson, 2011). The use of proportions can put more focus on the individual performance of teachers.

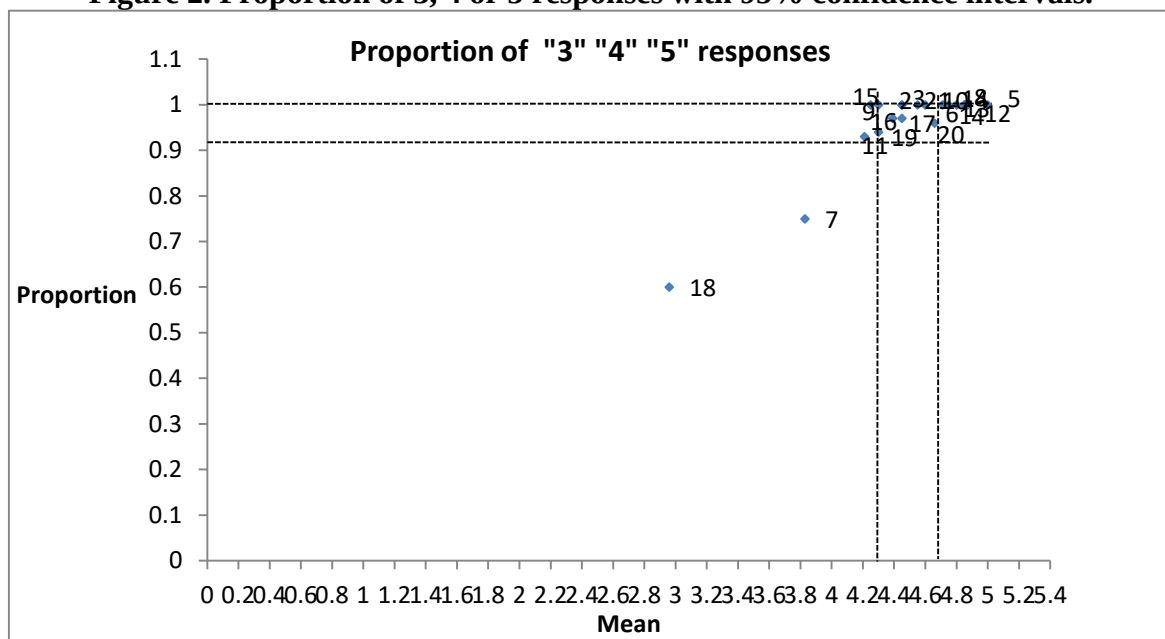
For this analysis, we received data from the South East European University for the academic year 2018/2019 for 5 faculties within the university. In this case, for illustrative purposes only, the Faculty of Contemporary Sciences and Technologies was analyzed, due to the large number of students who completed the student evaluations. We dropped all the classes that had less than 10 answers, we were left with 400 answers written in 21 classes taught by 20 teachers. We analyzed the question "Overall instructor rating". The response categories ('strongly agree', 'agree', 'neutral', 'disagree' and 'strongly disagree') are ordinal. The traditional method of analyzing SETs is to assign numbers to these ordinal categories, treat the data now as interval data and compute the mean (average). Using the data in Table 6, the professors' average score for the question is $109/31 = 3.52$. It is important to ask, '3.52 what?' What is the unit of measurement? There is none. What is being measured? If someone thinks that 'agreement with the question' is being measured, than the next time that person asks you, 'Do you agree?' you can answer '2.5' and expect the listener to know precisely how much you agree. Of course this cannot be true because 3.52 is the putative mean of an ordinal variable. Therefore, even if the calculation of the mean of ordinary data is legitimate, there will be no evidence that the professor is better than the average of the department. As Stark and Freishtat (2013) point out, the numbers placed on the answer categories are "labels", not "values". We cannot assume that the difference between "one" and "two" is the same as the difference between "four" and "five". There is no statistical sense in calculating the mean of ordinary data.

Table 6. Example data for SET

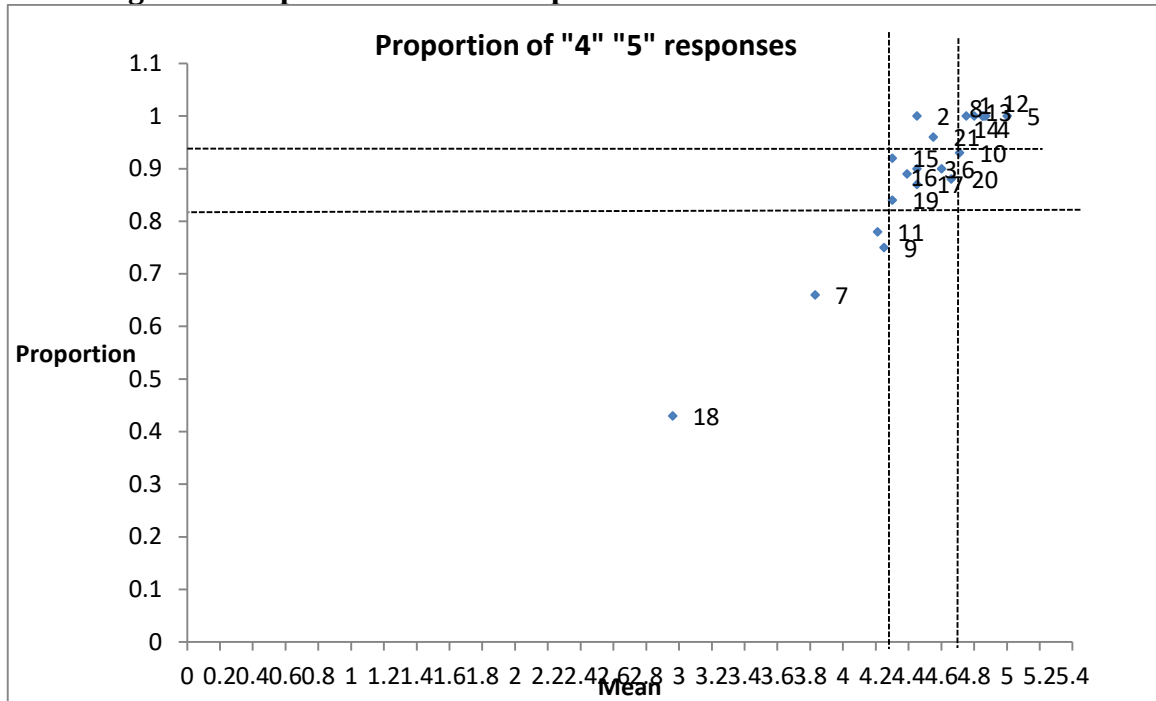
Response	Number of students	Assigned number	Total points
Strongly agree	7	7	49
Agree	15	2	30
Neutral	3	3	9
Disagree	5	4	20
Strongly disagree	1	1	1
Total	31		109

Figure 2 shows the proportions (p345) and the mean for the question "Overall instructor rating" for each class individually, where each data is represented by a class number. If we look at Figure 2 we will notice that most of the data is at the top where the proportion is equal to 1.0, and the ratio between p345 and the mean is nonlinear. There is little agreement between the two statistics. Professors who are considered "average" according to the mean (which fall within the lower and upper limit of the interval), are considered "above average" according to the proportion. It is not the case that the mean and the proportions are two ways of looking at the same thing. It is true that they will agree on extremes. For example, if all students rate a professor very well, both the proportion and the mean in this case will be high. Similarly, if all students rate a professor poorly, both the mean and the proportions will be low. When students do not register similar extreme responses, the average and proportion will vary and this is where most cases fall.

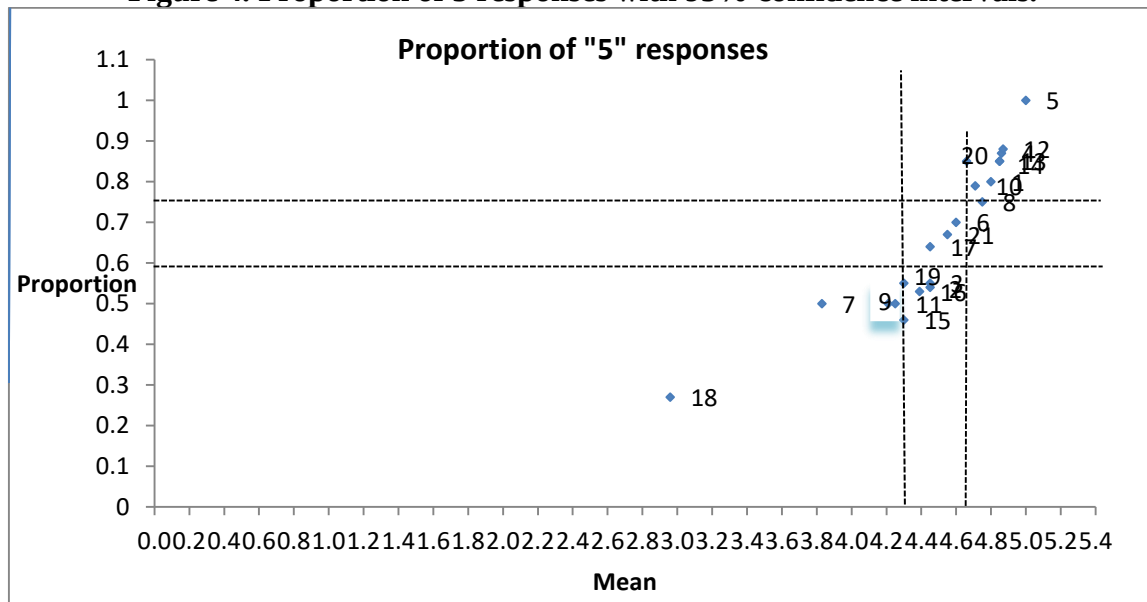
Figure 2. Proportion of 3, 4 or 5 responses with 95% confidence intervals.



Let us look at the statistics for p45 given in Figure 3. The ratio between the proportions and the average is more linear in contrast to the ratio between the proportions and the mean presented in Figure 2, which indicates a greater agreement between these two measures.

Figure 3. Proportion of 4 or 5 responses with 95% confidence intervals.

Finally, let us look at the ratio between p5 (the ratio of excellent or "5" answers) and the mean presented in Figure 4.

Figure 4. Proportion of 5 responses with 95% confidence intervals.

It can be noticed that most of the classes are within the average, but according to the proportions, they are below average. P45 statistics cannot distinguish between average and good student responses; however, p5 statistics make a better difference due to greater variability.

CONCLUSION

This paper was designed to shed light on the use of SET at SEEU as one of the instruments for quality improvement and attempted to measure student perceptions in the evaluation process. The University was a pioneer in the establishment and operation of such a system in the country and the region. For providing better quality education, positive students' perception about their teachers is required. Students' perceptions not only depend on teachers' internal characteristics but also influence other factors. These include students' decisions to participate in the student evaluation process, their severity; expectations for higher grades, gender of professors, scope of assignments and work, type of course, rank of professors, etc. A high correlation was found between the information that students had previously received about professors and their impact on evaluations. The evaluations offered by the students start as System 1 processes and are influenced by the information they have already received. We may suspect of the existence of a halo error caused by the information regarding the reputation, qualifications and experience of the professors. Evaluations rely heavily on automated brain processes, allowing subtle stylistic mannerisms and stereotypes to affect them. Designing evaluation systems that prompt more reflective, rational input would lead to obtaining reliable information from students, which further can contribute towards improved instruction and more fair treatment of the academic staff. Exploring such systems will also increase our own understanding of the intricate processes that drive decision-making - the knowledge we can apply to almost every field.

REFERENCES

- Cashin, W.E., 1995. Student Ratings of Teaching: *The Research Revisited*. IDEA Paper No. 32.
- Clayson, D.E. and Sheffet, M.J., 2006. *Personality and the student evaluation of teaching*. *Journal of Marketing Education*, 28(2), pp.149-160.
- Dijker, A.J. and Koomen, W., 1996. Stereotyping and attitudinal effects under time pressure. *European Journal of Social Psychology*, 26(1), pp.61-74.
- Finucane, M.L., Alhakami, A., Slovic, P. and Johnson, S.M., 2000. The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making*, 13(1), pp.1-17.
- Gaillard, F.D., Mitchell, S.P. and Kavota, V., 2006. Students, faculty, and administrators perception of students evaluations of faculty in higher education business schools. *Journal of College Teaching & Learning (TLC)*, 3(8).
- Gravestock, P. and Gregor-Greenleaf, E., 2008. Student course evaluations: Research, models and trends. Toronto: *Higher Education Quality Council of Ontario*.
- Harpe, S.E., 2015. How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, 7(6), pp.836-850.
- Heckert, T. M., Latier, A., Ringwald-Burton, A., & Drazen, C. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: is it possible to "buy" better evaluations through lenient grading?. *College Student Journal*, 40(3), 588-597.
- Kahneman, D., 2003. A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, 58(9), p.697.
- Kahneman, D., 2011. Thinking, fast and slow.
- Kareva V., Bexheti A. and Rexhepi G., 2019. Thinking out of the box as a precondition for sustainability and innovation development.
- Kareva V., Dika Z., Memedi Xh., 2018. Towards a Measurement of Academic Performance: The case of the South East European University (SEEU). *Journal of the European Higher Education Area*, 2018, No.2.
- Keeley, J. W., English, T., Irons, J., & Henslee, A. M. (2013). Investigating halo and ceiling effects in student evaluations of instruction. *Educational and Psychological Measurement*, 73(3), 440-457.
- LaForge, M.C., 2003. Student mood and teaching evaluation ratings. *Journal of The Academy of Business Education*, 4.
- Marsh, H.W. and Hocevar, D., 1991. Students' evaluations of teaching effectiveness: The stability of mean ratings of the same teachers over a 13-year period. *Teaching and teacher education*, 7(4), pp.303-314.
- McCullough, B.D. and Radson, D., 2011. Analysing student evaluations of teaching: Comparing means and proportions. *Evaluation & Research in Education*, 24(3), pp.183-202.
- McNatt, D.B., 2010. Negative reputation and biased student evaluations of teaching: Longitudinal results from a naturally occurring experiment. *Academy of Management Learning & Education*, 9(2), pp.225-242.
- Merritt, D.J., 2008. Bias, the brain, and student evaluations of teaching. . *John's L. Rev.*, 82, p.235.
- Munz, D.C. and Munz, H.E., 1997. Student mood and teaching evaluations. *Journal of Social Behavior and Personality*, 12(1), p.233.
- Naftulin, D.H., Ware, J.E. and Donnelly, F.A., 1973. The Doctor Fox lecture: A paradigm of educational seduction. *Journal of medical education*, 48(7), pp.630-635.

- Paswan, A.K. and Young, J.A., 2002. Student evaluation of instructor: A nomological investigation using structural equation modeling. *Journal of Marketing Education*, 24(3), pp.193-202.
- Pavel, A.P., 2012. The importance of quality in higher education in an increasingly knowledge-driven society. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 2(1), pp.120-127.
- Sanbonmatsu, D.M. and Fazio, R.H., 1990. The role of attitudes in memory-based decision making. *Journal of Personality and social Psychology*, 59(4), p.614.
- Sliusarenko, T., 2013. Quantitative assessment of course evaluations.
- Stark, P.B. and Freishtat, R., 2013. Evaluating evaluations: Part 1.
- Wachtel, H.K., 1998. Student evaluation of college teaching effectiveness: *A brief review. Assessment & Evaluation in Higher Education*, 23(2), pp.191-212.
- Wilson, R. C. (1986). Improving faculty teaching: Effective use of student evaluations and consultants. *The Journal of Higher Education*, 57(2), 196-211.