Exploring Correlation among Different Elements of Student Evaluation of Teaching

Wen Cheng
Department of Civil Engineering
California State Polytechnic University, Pomona
Pomona, USA
wcheng@cpp.edu

Rani Vyas
Department of Industrial and Manufacturing Engineering
California State Polytechnic University, Pomona
Pomona, USA
rrvyas@cpp.edu

Ranjithsudarshan Gopalakrishnan
Department of Industrial and Manufacturing Engineering
California State Polytechnic Univerity, \Pomona
Pomona, USA
ranjith@cpp.edu

Edwars R. Clay
Department of Civil Engineering
California State Polytehcnic University, Pomona
Pomona, USA
erclay@cpp.edu

Mankirat Singh
Department of Civil Engineering
California State Polytechnic University, Pomona
Pomona, USA
mankirats@cpp.edu

***Abstract*— The Research to Practice Full Paper intends to explore the possible correlations among the distinct elements of the student evaluations. The students have their own perception of the course and instructor. The individualism of students in a class reflects the need to change instructional elements of the course when the student expectation is not satisfied thoroughly. Students ought to have their mind flooded with the things which they want to learn from the instructor. They entail improving their skills through university education for the upcoming competition of the professional world. In addition, some of the previous research indicates student evaluations are a crucial source of data available to increase the student performance of the course.**

As a result, the authors are dedicated to the analysis of the instructor-based questions used in the evaluations. The questionnaire usually used in a feedback form is generalized and branded according to instructional requirements. The correlation of course structure of assignment in the mind of an instructor is different than in the mind of a student. The data used include courses offered by a selected professor at the Engineering Department at University. The student evaluations cover instructor and course-based questions. This include, course structure and organization, course objective-defining and course-meeting, knowledge demonstration, concept-explaining, communication effectiveness, class discussion-stimulating, student question-answering, consultation availability, student learning facilitation by assignments, concept-reflecting by exams, timely, and constructive feedback on student work, and overall rating of the instructor. The rating scale with five response categories are employed which range from very good (1) to very poor (5). The panel data used contains mean of each question from all students enrolled during a semester. Layer in the panel data differs for a lecture course and a lab course.

To capture the reliable correlation amongst instructor-based questions, some of the more popular rank correlation statistics are used including Spearman's ρ, Kendall's τ, Pearson's r, Wilcoxon T-test. Some of the teaching activities are found to have a stronger connection with the method of structuring assignments, which tend to have isolated impacts on student learning. It is anticipated that the research findings would assist the faculty in doing reflective work to comprehend student requisites.

***Keywords*—Student Evaluation, feedback assessments, Correlation analysis, Panel data.**

I. INTRODUCTION

University education is for students to enhance their critical thinking and develop skills for the professional world. Cambridge dictionary defines education as a service occupation, with the university students being considered as a consumer. Consumer perceived service quality is one of the most crucial features for every university. In general, student evaluations are the most popular platform for students to convey their satisfaction towards the courses being offered [1].

Evaluations can be gathered from either online or offline feedback assessments [2]. The feedback assessment forms consist of a list of multiple-choice, categorical, or descriptive questions related to teaching effectiveness. The effectiveness of the methods used to teach critical concepts in a classroom by universities can also be analyzed through these assessments [3].

There is extensive research dedicated to student evaluations. For example, Peterson et al. (2001) demonstrate a positive impact on student performance due to the use of student evaluations [4]. Another research article compares student evaluation and student learning through meta-analysis which showed that Student learnings/performance is not affected by a having highly ranked instructor. [5]. However, very few research articles have been centered on the correlation analysis of different elements of student evaluations with the intention to verify whether there is any association among distinct teaching performance areas.

To this end, this paper aims to address the correlation between different questions used in student evaluations. Multiple correlation analysis methods are used with the Panel

data gathered from student evaluation at University. The panel data consists of both different years of data and multiple students' data for the same year.

The contribution of this paper mainly resides in two areas: First, for the statistics field, it provides additional insights to the identification of the difference and similarity of various correlation statistics relying on the rarely used educational empirical data. Second, the common results found in the different correlation tests would lead to a better understanding of the relationship of various teaching elements, which can then be used by the university community for the betterment in student learnings.

II. DATA

The panel data has been gathered by one University using a student evaluation questionnaire acquired from the Department of engineering. The evaluations include the courses taught by one same instructor from Fall 2009 to Spring 2019 with the exclusion of Fall 2015 due to instructor sabbatical leave during that quarter. The university has changed from a quarter system to a semester system in Fall 2018, but the questionnaire remains the same. One different questionnaire was used for each lecture, and lab courses, respectively. The panel data spans 10 years and includes winter, fall and spring terms (quarters or semesters). The total number of lecture courses are 51, and of lab courses are 17 during this period.

TABLE 1 SUMMARY OF STUDENT EVALUATIONS FROM LECTURE COURSES

Question Number	Question	Min.	Max.	Mean	Standard Deviation
3	How effectively does the instructor organize and structure the course (e.g.: course objectives, grading policy, textbook and/or reference material, office hours, number of exams and schedule of exam times)?	1.00	1.33	1.07	0.07
4	How well did the instructor define and meet the objectives of the course?	1.00	1.32	1.06	0.07
5	How well does instructor demonstrate knowledge of the subject?	1.00	1.36	1.06	0.07
6	How effective is the instructor in explaining the concepts of the course?	1.00	1.64	1.13	0.12
7	How effectively does the instructor communicate with the class (e.g., speaks clearly, writes legibly, uses appropriate visual aids, listens effectively)?	1.00	1.96	1.17	0.16
8	How well does the instructor stimulate and/ or encourage appropriate class discussions?	1.00	1.36	1.09	0.08
9	How effectively does the instructor answer student question?	1.00	1.36	1.10	0.08
10	How available is the instructor for student consultation?	1.00	1.30	1.09	0.08
11	How well do the instructor's assignment facilitate student learning?	1.00	1.32	1.09	0.08
12	How well do the exams reflect the concepts of the course?	1.00	1.32	1.07	0.07
13	How effective is the instructor in giving timely and constructive feedback on student work?	1.00	1.25	1.06	0.06
14	How would you rate this instructor compared to other instructor in the university?	1.00	1.44	1.05	0.07

[a] Questions 1~2 & 15~17 are excluded from the research analysis since those are not related with instructor performance. [b] Min and Max refer to Minimum and Maximum values of the data.

TABLE 2 SUMMARY OF STUDENT EVALUATIONS FOR LAB COURSES

Question Number	Question	Min.	Max.	Mean	Standard Deviation
1	How well prepared for the lab does the instructor appear?	1.00	1.26	1.07	0.08
2	How well did the instructor define and meet objective of the lab?	1.00	1.30	1.09	0.08
3	How well does the instructor arouse interest and transmit enthusiasm in lab?	1.00	1.45	1.17	0.12
4	How helpful and available is the instructor for consultation during the lab and office hours?	1.00	1.30	1.07	0.08
5	How would you rate this lab instructor as compared to other lab instructors?	1.00	1.40	1.12	0.11
6	How well did the lab coordinate with and reinforce the lecture?	1.04	1.38	1.21	0.12
7	How clearly were the experiments written?	1.04	1.59	1.26	0.13

[a] Questions 8~10 are excluded from the research analysis since those are not related to the instructor's performance. [b] Min. and Max. refer to Minimum and Maximum values of the data

The students were asked to answer the questions using a score of 1 to 5, with 1 being "very good" and 5 being "poor". Only instructor-based questions are selected from both data banks for this paper. Tables 1~2 show that Q3 to Q14 for lecture courses and Q1 to Q7 for lab courses are selected, which are instructor-based questions. The tables also show other basic statistical measures of the panel data, including minimum, maximum, mean, and standard deviation.

III. Methodologies

The correlation analysis is one of the statistical methods to compare two variables and find a relationship between them. The method could be parametric or non-parametric. The parametric method assumes a specific distribution of the population and is usually used for continuous and unranked data. Whereas for ranked data, non-parametric tests are used which uses data with nominal or ordinal scale. The range of resulted correlation coefficient is usually between -1 to +1. Here the sign of the coefficient shows a negative or a positive trend between the two variables. This paper worked with the mean value of ranked data, for which ranked correlation methods can be used as well as parametric tests such as Pearson's T-test. The usual absolute threshold value of a correlation coefficient of 0.6 is used to determine whether two student evaluation elements are highly correlated or not.

Four popular correlation methods are selected for the correlation analysis, which includes Spearman's ρ, Kendall's τ, Pearson's r, and Wilcoxon T-test. The details of each correlation tests are shown in the following sections in order.

A. Spearman correlation

Spearman's rank correlation (ρ) is a nonparametric technique for evaluating the degree of linear association or correlation between two independent variables. It operates on the rank of the data rather than the raw data value. It assesses how well is the relationship between two variables using a monotonic function. The advantage is that results are unaffected by the distribution of the population because the technique operates on the rank values. It is not required for the data to be collected over a regular space interval. It can be used with a small sample size and it is easy to apply [6]. The correlation coefficient can be calculated using the following equation:

$$\rho(rho)\frac{\sum(x'-m_{x'})(y_i'-m_{y'})}{\sqrt{\sum(x'-m_{x'})^2\sum(y'-m_{y'})^2}} \quad (1)$$

Where

x′=rank of x

y′=rank of y

mx'= means of rank(x)

my'= means of rank(y)

B. Kendall correlation

The Kendall rank correlation (tau) is a nonparametric measure of a coefficient. This evaluates the degree of similarity between two sets of ranked objects. This coefficient depends upon the number of inversions of pair of objects which would be needed to transform one rank order into the other. Kendall coefficient of correlation is obtained by normalizing the symmetric difference such that it will take values between -1 and +1 with -1 corresponding to the largest possible distance, and +1 corresponding to the smallest possible distance [7]. The correlation coefficient can be calculated using the following formula:

$$tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (2)$$

Where

n_c = *total number of concordant pairs*

n_d = *total number of discordant pairs*

n = *size of x and y*

C. Pearson correlation

Pearson's is a parametric correlation (r) method used for continuous data. It measures the association between variables of interest as it is based on the method of covariance. Pearson is defined as the covariance of the two variables to the product of their respective standard deviation [8]. The data consists of the average score of some students enrolled in each course. Even though the score from an individual student is rank value, however, the average score is a continuous variable. This is why Pearson's correlation coefficient r can be calculated between -1 and 1 [3]. The correlation coefficient can be calculated using the following expression:

$$r = \frac{\sum(x-m_x)(y-m_y)}{\sqrt{\sum(x-m_x)^2\sum(y-m_y)^2}} \quad (3)$$

Where

m_x = *mean of x variable*

m_y = *mean of y variable*

D. Wilcoxon correlation

Wilcoxon T-test is also called a Mann-Whitney U test. It is another nonparametric test of the null hypothesis. Under this test, two independent variables are tested to check whether they are from populations having the same distribution [9]. For a population of one of the two variables, the number of lower ranks and higher ranks for the sample rank value are calculated and these are called wins and ties, respectively. The sum of wins and ties is called U or W. The correlation coefficient can be calculated using the following equation:

$$W = \sum_{i=1}^{N_r}[sgn(x_{2i} - x_{1,i}).R_i] \quad (4)$$

$$z = \frac{W}{\sigma_W} \quad (5)$$

Where

sgn = *sign function*

R_i = a rank of the data of two variable

W = test statistic

σ_W = standard deviation of data

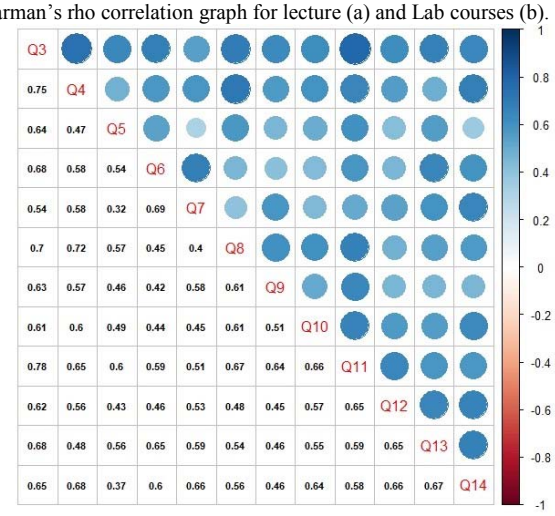
P value of the Wilcoxon T-test is the probability of the ranked data after the sample value of z critical is compared. [6]

IV. Results

The detailed correlation results containing correlation graphs, summary of correlation matrix, and some highest correlated questions are presented in the following subsections.

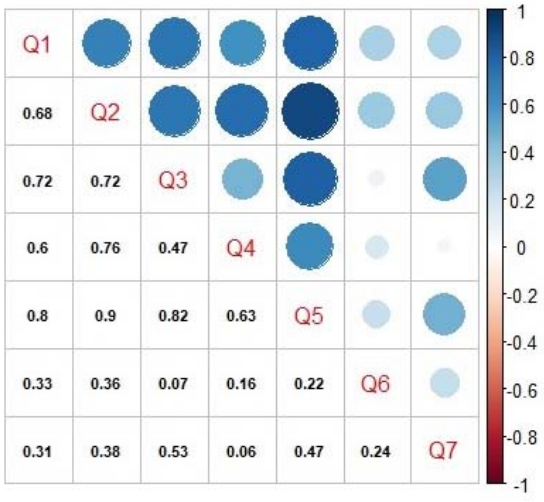
A. Correlation graphs

Results for correlation coefficients is computed through the R language for each correlation method. The graphs are generated through correlation matrix plots. The correlation graphs show the coefficient values for each method individually using the lecture and the lab course dataset.



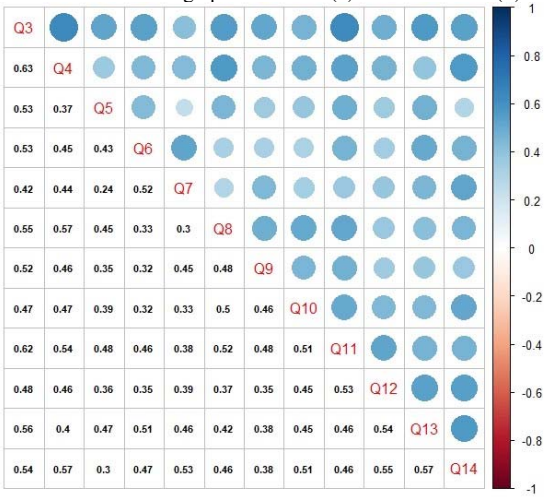
(a)

Fig 1. Spearman's rho correlation graph for lecture (a) and Lab courses (b).

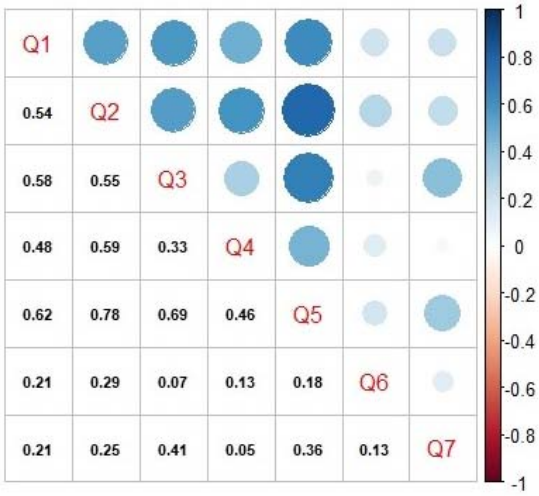


(b)

Fig 2. Kendall's Tou correlation graph for lecture (a) and lab courses (b)

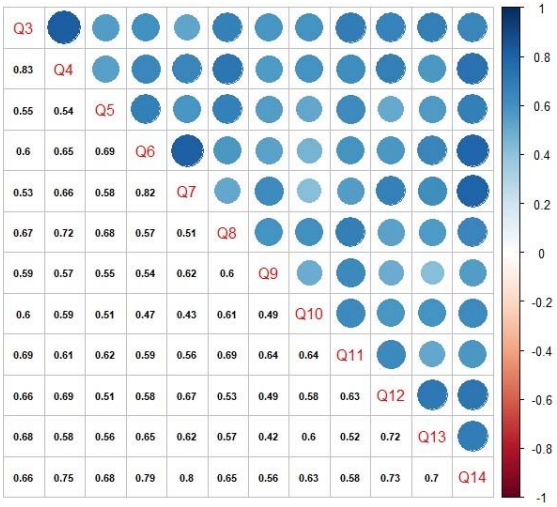


(a)

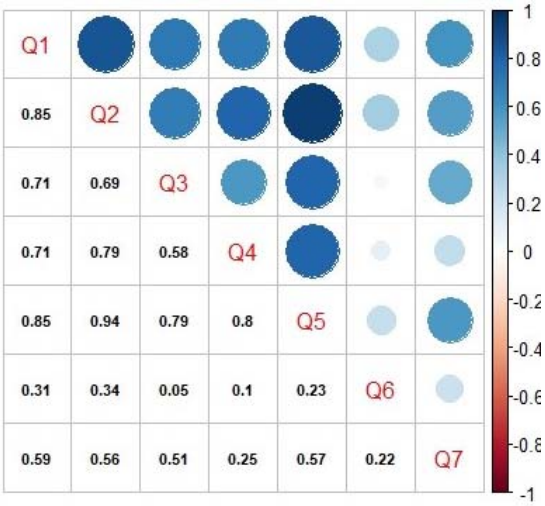


(b)

Fig 3. Pearson's r correlation graph for lecture (a) and lab courses (b)

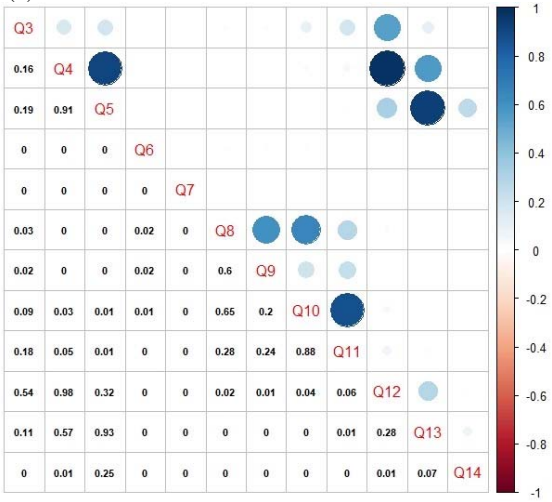


(a)

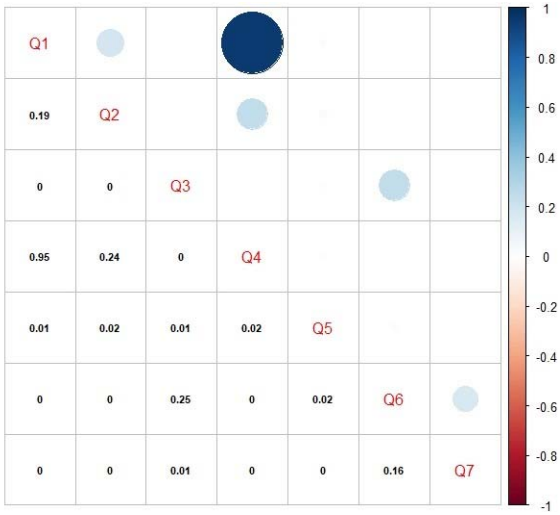


(b)

Fig 4 Wilcoxon's p-value correlation Graph for the lecture (a) and the lab courses (b)



(a)



(b)

This diagonal in the above graphs showing question numbers has the value 1 in each cell as self-correlation is always positive 1. The upper triangle of the matrix shows colored circled which different radius representing the magnitude of the correlation, and the lower triangle of the matrix gives the numerical values representing the correlation coefficient. The side scale represents the color gradient for a value ranging from -1 (warmest) to 1 (coolest). In all the graphs above we do not have a negative coefficient which suggests that each question is positively correlated with at least one question.

According to graphs, it is evident that Wilcoxon T-test shows the lowest number of correlations for both the lecture and the lab courses. Lecture case shows that almost all the questions are somehow positively correlated when it comes to instructor-based questions. This means that if one of the questions has a greater rank response from students, then all the other questions will tend to show a similar increase in response.

B. Summary of Correlation Matrix

When the results of Spearman correlation are compared with other methods, the Kendall correlation method shows a little drop in the correlation coefficient values, whereas Pearson's correlation method shows an increase in the correlation coefficient. Wilcoxon T-test has the most variating results than other results.

TABLE 3 SUMMARY FROM CORRELATION MATRIX FOR LECTURE COURSES

Question	Pearson	Spearman	Kendall	Wilcoxon
Lecture				
Q3	7	10	2	0
Q4	7	4	1	2
Q5	4	2	0	2
Q6	6	3	0	0
Q7	6	2	0	0

Q8	6	5	0	2
Q9	2	3	0	1
Q10	3	4	0	2
Q11	7	7	1	1
Q12	6	4	0	1
Q13	5	4	0	1
Q14	9	6	0	0
Lab				
Q1	4	4	1	1
Q2	4	4	1	0
Q3	3	3	1	0
Q4	3	3	0	1
Q5	4	4	3	0
Q6	0	0	0	0
Q7	0	0	0	0

The difference between the four methods is significant in Table 4. The Highest number of correlations is shown in the methods Pearson and Spearman whereas Kendall and Wilcoxon only show few correlations for some of the questions. The two methods, Pearson and Spearman illustrate similar results for data received from lab courses. Q6 and Q7 for the lab courses are not correlated to any of the other questions according to all the methods used.

C. Highest correlated questions for lecture and lab courses

TABLE 4 The highly correlated questions in each method for Lectures and Lab.

Evaluation Data	Spearman	Kendall	Pearson	Wilcoxon
lectures	Q3~Q11	Q3~Q4	Q3~Q4	Q4~Q12
lab	Q2~Q5	Q2~Q5	Q2~Q5	Q1~Q4

Table 4 illustrates only highly correlated questions from each method for both lecture and lab course datasets. It shows the irregularities between the four methods. Wilcoxon T-test results do not match the other three methods in both course datasets. Assuming by the similar results of other methods, it can be said that a null hypothesis is not true for the dataset, suggesting that there are similar values in multiple pairs of variable populations. Kendall's and Pearson's coefficients show similar results for both datasets, which suggest a suitability of the methods with each other and for two diverse populations. Spearman coefficient agrees with the two methods Kendall and Pearson when the lab dataset is analyzed.

When looking at similarities in the lab and lecture analysis, it is evident results from lab courses are more consistent than the lecture courses. It could be due to a small sample size of the lab course dataset when compared to the lecture course dataset. As the dataset sample size increase, the correlation coefficients show more conflicting results in each method. Both analyses of labs and lectures agree with the correlation results by Kendall and Pearson correlation method, even though one is non-parametric, and the other is parametric. Pearson follows the hypothesis of normal distribution in a variable population which is proven true due to similar results.

A realistic view of the correlated questions can be noted when looking at the questionnaire. These correlations show that Q3 and Q4 from the lecture course questionnaire can be used by the instructor. Looking at the similarities in them, both questions address the organization of course specified and followed by the instructor. It can be stated as, effective course objectives, grading policy, reference material, number of exams, and schedule helps the instructor to meet the defined objectives of the course.

The Q4 and Q12 are also correlated according to Wilcoxon Method. One another correlation according to Spearman is between Q3 and Q11. It can be anticipated that, the structure of course leads to effective student learning through assignments.

For lab courses, Q2 and Q5 are highly correlated. According to the questions, the defined objectives of the lab by the instructor would be related to the comparison of different lab instructor in students' views. It can be stated that when the instructor is compared with other instructor, students usually look at the lab objectives devised by the instructor. The Wilcoxon correlation suggests Q1 and Q4 as highly correlated. It shows that, the preparedness by the instructor for the class gives students a chance to have the consultation for a query with the instructor during lab and office hours.

V. CONCLUSION & LIMITATIONS

We have analyzed the panel data of the Student evaluations of lecture and lab courses taught by the same instructor at university for around 10 years. The correlated questions give the relationship between the two performance areas involved in teaching the course. The positively correlated questions indicate that teacher performing well in one aspect tends to have great performance in other aspects. This is helpful for the instructor to enhance his teaching performance by improving the highly correlated areas.

Even though the analysis reveals a better understanding of the correlation of various elements of student evaluation of teaching, there are some caveats for the results as shown in the present study. First, all student evaluations were collected from one instructor only. More data from multiple instructors might show different correlation patterns among the different teaching elements. Second, only the average scores of the entire class for each question are used for correlation analysis. The association evaluation based on individual student data could yield totally different results.

REFERENCES

[1] Athiyaman, Adee, "Linking student satisfaction and service quality perception: the case of university education", European Journal marketing 31,7, 1997, ISSN: 0309-0566.

[2] Menges, Robert J. & Brinko, Kathleen T., Effects of Student Evaluation Feedback: A meta-analysis of higher education research, 1986.

[3] Donald, Janet G. & D. Brian Denison, Quality Assessment of University Students, The Journal of Higher Education, vol. 72:4, pp. 478-502, 2001, DOI: 10.1080/00221546.2001.11777109.

[4] Peterson, Marwin W. & Marne K. Einarson. What are Colleges Doing about Student Assessment?, The Journal of Higher Education, vol. 72:6, 2001, pp. 629-669, DOI: 10.1080/00221546.2001.11777120.

[5] Uttl, Bob et al., Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. Studies in education evaluation, 2017.

[6] Gauthier, Thomas D., Detecting trends using spearman's rank correlation coefficient, Environmental forensics, 2001.

[7] Abdi, Herve, The Kendall rank correlation coefficient, 2007.

[8] Chok, Nian Shing, Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data, 2010.

[9] McKnight, Patrick E. & Julius Najab, Mann-whitney U test, wiley online library, 2010, DOI: 10.1002/9780470479216.

[10] Seldin, P., Using student feedback to improve teaching. D. DeZure (Ed.), To lmprove the Academy, Vol. 16, pp. 335-346, 1997.