A Comparison of Academic Staff Performance Evaluations by Head of Department and Their Students

Nazera Khalil Dakhil, and Amenah Abdullateef Hameed
Faculty of of Mathematics and Computer Sciences-Kufa University
e-mail: nazeera.dakhil@Uokufa.edu.iq

Abstract

Teaching staff performance assessment are vital for career progress and mentoring. Faculties at University of Kufaevaluate the performance of their academic staff on a yearly basis by Head of Department and the Students.The evaluation instrument are two different forms that contains set of questionnaires, one is usually completed by Head of Department (HE) and other by the students (S). It is therefore the main aim of this research is to assess the association between HE and SE.

The data was collected for 24 teaching staff from Department of Quality and Performance at the Faculty of Mathematics and Computer Science by means of questionnaires. The problem is that the form that is completed by student are not measuring what is suppose to measure. It is therefore new version of questionnaires are designed and the students asked to complete both version in order to compare them. The results shows the staffs grades as a result of HE are much higher than grades result from theSE, and that new version of questionnaires are more reliable than the old version.

Introduction

Systems for evaluating teaching and course quality in higher education have long beenestablished in the US, the UK and Australia and are also common in many othercountries.

Whilst there is a large number of possible sources of feedback and evaluation data onboth teaching and course quality (including, for example, course documentation,progression rates, curriculum design processes, teaching committees, etc.) the mostcommon source of input to teaching evaluation is feedback from students. Indeed, the collation of student feedback forms (forexample) is routine practice in most institutions and causes little concern or debate in these countries.

In the Iraqi universities, theDepartment of Quality Assurance and Academic performance(QAP) is responsible for the use of a system of evaluation in higher education.However, systematic collection and appropriate processing of such feedback is not well established in most Iraqi universities.

It is important to note that because the main reason for teaching evaluation is to improvethe quality of course delivery and to provide direct feedback to teaching staff, the QAP recommended that the feedback be kept confidential.However, it can be that there is a need for a more standardized approach to thisprocess; one that is informed by a review of best practice models and relevant researchfindings. Whilst it is thought highly unlikely to be possible to recommend onequestionnaire or system that will suit all purposes at the University.

Literature Review

Researchers generally consider student evaluations of instructors to be highlyreliable and at least moderately valid.[1,2,3]

Other methods of evaluation (such as evaluations by colleagues or trained observers)have not been found to be reliable and therefore not valid.[1]

Student ratings of instructors have been found to be related to ratings of instructor'sskills in course organization, rapport with students, and fair grading; variance inorganizational skill (having an organized course plan and clearly identifying whatstudents need to do) explained most variance in student evaluations.[4].

Although grades do have some effect on how students rate instructors,[5] its effect isfairly low [6] and can be statistically adjusted for.[7] Grades do not have as large of aneffect as how much they felt stimulatedby the class,[8] and whether the class was appropriately difficult (courses are ratedlower for being too easy or too difficult).[9]

Contrary to the, students who do poorly in a class are equally orless likely than those who do well to complete course evaluations.[10].

Surveys of students typically indicate that students believe faculty and administrators don't take their evaluations seriously.[11] This may be justified, as some studies have found that instructors do not view student evaluations as valuable for improving instruction[12] and very few report making changes to their courses as a result of course evaluations.[13]

Students are more likely to complete course evaluations if they see value in them (e.g., understand how they are being used, believe that their opinions have an effect).[14]. The SET results influence mainly the content of the course, its structure, teaching style and methods employed.[3].

The student evaluations of instructors to be highlyreliable and at least moderately valid.[2].

Why evaluate?

According to the literature teaching and/or course evaluations can be used for fourdifferent purposes, including:[9,10]

(1) as a evaluative and diagnostic feedback technique, in efforts to improve teaching and courses); (2)as a summative feedback technique , in personnel and administrative decision-making); (3) as a source of information for prospective students when selecting course units and lecturers; and (4) as a source of data for research on teaching.

The aim of teaching valuations aims to promote good quality teaching and supportacademic staff in their challenging and complex role as teachers, and emphasise the importance of undertaking such a process in a non-threatening andsupportive atmosphere. However, for many new staff,there may be a suspicion or even a fear that such feedback could be used for otherpurposes. It is, therefore, vitally important that the system that is developed is usedappropriately and that there is agreement on overall purpose. Important "political"decisions may have to be made regarding who sees the responses, how the feedback isresponded to and how it integrates with other sources of information regarding thecourses. Some authors discuss in some detail the problematic issue of developing fundamental policy for student feedback before providing a list of recommendations

for systems thatare more likely to prove effective in supporting teaching practice. Particular emphasis isplaced on trust, issues of controlled access/confidentiality and cross comparison withother sources of information regarding, for example, demographics within the groups.[11]

Other authors have also stressed the value inensuring that such feedback does not focus solely on the "performance" aspects of lecturedelivery, for example, but also asks strong questions about the students' level ofengagement, commitment and interest in their programmes of study. [12]

What are we evaluating?

Evaluation in higher education can take place at various levels: at the level of theindividual lecturer, the course unit, the course module, the semester, year of study, thesubject, the entire programme (e.g. the B.A., the B.Sc.), the academic department, thefaculty or even at the level of the institution .[13]

The level atwhich one should collect feedback is dependent upon the purpose(s) of the evaluation. Ifthe purpose is to improve teaching within a particular module, then feedback should besought on teachers and their activities within that module.

In terms of evaluating teaching quality, it is first necessary to define what is meant by"good" teaching. Individuals may hold different conceptions of what good teachingactually is. In the literature relating to student learning, key commentators have, in recentyears, defined good teaching as that which actively facilitates student learning (through afocus on conceptual change) as opposed to the (possibly stereotyped) educational "chalk andtalk" transmission of information .[14]

Why involve students?

Student evaluations of teaching (SET) are the most common source used in theevaluation of teaching in higher education. [20] However, aspreviously noted, students are not the only sources from which feedback may be obtainedon teaching and course quality. Colleagues (through peer observation, review ofcurricula, etc), Head of Department, self-reflection, assessment performance, attendance,comparison with other courses or other institutions, etc., are all valuable and an integratedapproach is necessary for a complete and detailed picture. It should be remembered that students are indeednot competent to evaluate teaching roles such as those involving course design(objectives, content, methods and assessment), delivery methods or grading practice inassessment. Individual lecturers, their colleagues, curriculum design experts and Heads ofDepartment are best placed to provide feedback on such matters. However, in terms of the quality of the delivery of instruction, it is generally agreed in the literature that only students are in a position to provide feedback. ([19] ,[18]).

How should student feedback be collected?

Although questionnaires are most often used to collect student feedback, but it is not the only method. Student feedback can also be obtained through studentrepresentation, groupdiscussions and/or focus groups, one to one student interviews, e-mail, bulletin boards,students' diaries and/or log books and informal comments.

Again, the method selectedwill be dependent upon the purpose(s), level(s) and context of the evaluation. Generally itis recommended that a combination of

techniques be used to gather feedback fromstudents .[13,9,7]. All Iraqi Universities provide at least acentralisedclosed ended questionnaire-based system whereby student feedback is collected andprocessed. The particular office or administrative centre that deals with this work is QAP. The questionnaires items may be open-ended in nature and thus gather morequalitative data or may be closed-ended and thus aim at collecting more quantitative-typedata focused on specific, pre-determined issues. Whilst rich, revealing and veryinformative, the analysis of qualitative data can nonetheless be extremely time-and labour-intensive, and is, therefore, not often used for "routine" course monitoring incentralised feedback systems. When surveying large numbers of students, it is morecommon to use closed-ended questions in standardised questionnaires.

In terms of delivery and collection methods, the commonest is to use forms that are issued in class (or by post, resulting usually in a lowreturn rate). This can be quite effective if done in class provided the attendance levels arereasonable. However, providing such a service requires appropriate equipment andstaffing.An alternative, which is being used in many institutions in the US, in particular, is onlinesurveys. These can be quite flexible in format, and relatively easy to organise. Thedisadvantage is that return rates can be very low since students are required to findadditional time in which they should log on in order to work through the survey. Manysystems can track which students have not completed the forms and send reminder emails.

The Evaluation Instruments

Staff Performance evaluations are essential for career development and mentoring and to ensure accountability and equity across the Faculty. Faculties at University ofKufa evaluate the performance of their academic staff on a yearly basis. Academic performance is to be assessed in relation to academic duties. Staff have the opportunity to complete form.The evaluation instrument includes questions related to weekly teaching hours for undergraduate and/or postgraduate, number of published papers, others activity like presenting conference paper or seminar; attending conference, workshops, symposium; provide training courses, etc. students have the opportunity to complete a Course Evaluation" form for each course in which they are enrolled. The evaluation instrument includes questions related to the quality of the educational experience and the faculty member's performance in teaching the course.

The Problem

Staff Performance evaluations are essential. Faculties at University of Kufa evaluate the performance of their academic staff on a yearly basis by Head of Department and the Students.A copies of evaluation form for Professor, Assistant Professor, Lecture, and Lecture Assistant. , and a copy of student evaluation form are attached.

The evaluation instrument are two different forms that contains set of questionnaires, one is usually completed by Head of Department (HE) and other by the students (SE).

Since both questionnaires measure the same things, "performance of academic staff". The outcomes of both of them are related. Unfortunately our study of data shows a lake of association. The first hypothesis is that the form that is completed by student is not measuring what is suppose to measure. It is therefore new version of questionnaires (NQ) are designed and a copy of this form is attached.

the students asked to complete both version in order to compare them. The second hypothesis is that new version (NQ) of questionnaires are more reliable than the old version (OQ).

Aims of the research

1. Compare evaluation of the HE with the SE.
2. Asses the agreement between HE and SE using Cohen*k* Kappa.
3. Asses (OQ) that was provided by the QAP.
4. Design new version of questionnaires (NQ)
5. Compare UQ with NQ using paired t-teat .

Correlation Coefficient [18]

The linear correlation coefficient r, Which is a numerical measure of the strength of the association between two variables representing quantitative
data. Using paired sample data, the value of r is use to conclude that there is (or is not) a linear correlation between the two variables.

A correlation exists between two variables when the values of one variableare somehow associated with the values of the other variable.

Before doing any formal statistical analyses, first plot a scatter-plot to explore thedata visually. We can examine the scatterplot for any distinct patterns and for anyoutliers. If the plotted pointsshow a distinct pattern, we can conclude that there is a correlation between the twovariables in a sample of paired data.

If tthe *x*-values increase, and the corresponding*y*-values increase, then there isa *positive* correlation between *x* and *y*. If the *x*-values increase, and the corresponding *y*-values decrease, then there is a *negative* correlation between *x* and *y*. If there isno distinct pattern, then there is no correlation between *x* and *y*.

Linear Correlation Coefficient

The linear correlation coefficient r measures the strength of the linearcorrelation between the paired quantitative x- and y-values in a

sample(Its value is computed by using Formula 1or Formula 2,

this is referred to as the Pearson product moment correlation coefficient

in honor of Karl Pearson (1857–1936), who originally developed it.

Because the linear correlation coefficient r is calculated using sample

data, it is a sample statistic used to measure the strength of the linear

correlation between x and y. If we had every pair of population values for

xand y, the result of Formula 1orFormula 2 would be a population

parameter, represented by ρ (Greek letter rho).

$$r = \frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{n\ (\sum x^2)-(\sum x)^2}\sqrt{n(\sum y^2)-(\sum y)^2}} \qquad (1)$$

This shortcut formula simplifies manual calculations, but *r* is usually calculated with computer software or a calculator.

$$r = \frac{\sum(z_x z_y)}{n-1} \qquad (2)$$

where$z_x$is the *z* score for the sample value *x* and $z_y$is the *z* score for the
sample value *y*.

If the *P*-value computed from *r* is less than or equal to the significance level, conclude that there is a linear correlation. Otherwise, there is not sufficient evidence to support the conclusion of a linear correlation.

The methods of this section apply to a *linear* correlation. If you concludethat there does not appear to be linear correlation, know that it is possible that theremight be some other association that is not linear.

The Properties of the Linear Correlation Coefficient rare : (a)The value of r is always between -1 and 1 inclusive. That is,$-1 \leq r \leq 1$; (b)If all values of either variable are converted to a different scale, the value of r does not change; (c) The value of r is not affected by the choice of x or y. Interchange all x- and y-values and the value of *r* will not change; (d) *r* measures the strength of a linear relationship. It is not designed to measure the strength of a relationship that is not linear; (e) *r* is very sensitive to outliers in the sense that a single outlier can dramatically affect its value.

Measuring Agreement[22]

There are many occasions when you need to determine the agreement between two raters. For example, the dean of a college might want to determine whether head of department and students agree on evaluations of their lecturer. The level of agreement between the two evaluations is analysed using Cohen's kappa.

There are different assumptions that your data must meet in order for a Cohen's kappa to give you a valid result. Cohen's kappa has five assumptions that must be met. If these assumptions are not met, you cannot use a Cohen's kappa, but may be able to use another statistical test instead. Therefore, in order to run a Cohen's kappa, you need to check that your study design meets the following five assumptions: (a) he response (e.g., judgement) that is made by your two raters is measured on a nominal scale and the categories need to be mutually exclusive, which means that no categories overlap;(b)The response data are paired observations of the same phenomenon, meaning that both raters assess the same observations; (c) Each response variable must have the same number of categories and the cross tabulation must be symmetric (i.e., "square") (e.g., a 2x2, 3x3, 4x4, etc.); (d) The two raters are independent (i.e., one rater's judgement does not affect the other rater's judgement).; (e) The two raters are fixed, meaning that they are specifically selected to take part in the study.

Cohen's *k*coecient[21]

Cohen's kappa coefficient is a statistical measure of inter-rater agreement or inter-annotator agreement [Carletta, Jean. (1996] for qualitative (categorical) items. It is generally thought to be a more robust measure than simple percent agreement calculation

since κ takes into account the agreement occurring by chance.

Cohen's *k* measures pairwise agreement among a set of raters making

Categorical judgments, correcting for expected chance agreement

Table 1. Contingency Table

	Rater 1		
Rater 2	a	B	a + b
	C	D	c+ d
	a + c	b +d	n = a+b+d+c

A general expression for *k* is

$K = \frac{P_0 - P_e}{1 - P_e}$(4)

Where p_o=(a+d)/n, p_e=((a+c)(a+b)+(b+d)(c+d))/n^2

where P_0 is the observed probability of agreement between the two raters

and P_e is the expected probability of agreement under the assumption of independentrating by the two raters. As is obvious from the dentition, *k*

must be less thanor equal to 1 and its lower bound depends on P_e, but will

be less than zero. If the raters are in complete agreement then κ = 1. If there is no agreement among the raters , κ = 0.

Table 2. Kappa Interpretation

Kappa	**Interpretation**
< 0	Poor agreement
0.0 – 0.020	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

Table 2 shows that Kappa has a range from 0-1.00, with larger values indicating better reliability. Generally, a Kappa > .70 isconsidered satisfactory

Linear Regression [18]

Given a collection of paired sample data, the regression equation

$\hat{Y} = b_0 + b_1 x$

algebraically describes the relationship between the two variables *x* and *y*.

The graph of the regression equation is called the regression line(or *line ofbest fit,* or *least-squares line*).

The regression equation expresses a relationship between *x* (called the explanatory variable, or predictor variable, or independent variable) and (called the response variable, or dependent variable). The preceding definition shows that in statistics,the typical equation of a straight line*y = mx + b* is expressed in the form

$\hat{\mathbf{y}} = \boldsymbol{b_0} + \boldsymbol{b_1 x}$ (5)

whereb_0 is the *y*-intercept and b_1 is the slope.

The slope b_0 and *y*-intercept b_1 can also be found using the following formulas.

$$\boldsymbol{b_1} = \frac{\boldsymbol{n}(\sum \boldsymbol{xy}) - (\sum \boldsymbol{x})(\sum \boldsymbol{y})}{\boldsymbol{n}(\sum \boldsymbol{x^2}) - (\sum \boldsymbol{x})^2}$$

$$\boldsymbol{b_0} = \frac{((\sum \boldsymbol{y})(\sum \boldsymbol{x^2}) - (\sum \boldsymbol{x})\ (\sum \boldsymbol{xy})}{\boldsymbol{n}\ (\sum \boldsymbol{x^2}) - (\sum \boldsymbol{x})^2}$$

The values of b_1 and b_0 can be easily found by using any one of the many computerprograms and calculators designed to provide those values. Once we have evaluated *b*1 and *b*0, we can identify theequation of the estimated regression line, which has the following special property:

The regression line fits the sample points best. (The specific criterion used to determine which line fits "best" is the least-squares property, which will be described later.); (1)The sample of paired (*x*, *y*) data is a *random* sample of quantitative data; (2)Visual examination of the scatter-plot shows that the points approximate a straight-line pattern; (3) Outliers can have a strong effect on the regression equation, so remove any outliers if they are known to be errors. Consider the effects of any outliers that are not known errors.

Notation for the Equation of a Regression Line

Regression Equation	Population Parameter	Sample Statistic
y-intercept	ß0	b_0
Slope	ß1	b_1
Equation	$y= ß_0 + ß_1x$	$y= b_0 +b_1x$

$$\hat{y}=b_0+b_1x \quad (6)$$

Slope: $$b_1= \frac{s_y}{s_x} \quad (7)$$

y-intercept: $$b_0 = \bar{y} - b_1\bar{x} \quad (8)$$

where r is the linear correlation coefficient,s_y is the standard deviation of the y values, and s_x is the standard deviation of the x values

The null hypothesis: the mean difference equals 0 can be tested by determining whether the confidence interval includes 0.

There are no exact procedures for dealing with dependent samples, but the *t* distributionserves as a reasonably good approximation., so the following methods arecommonly used. The basic requirementsare:(1)The sample data are dependent; (2)The samples are simple random samples; (3)Either or both of these conditions is satisfied: The number of pairs of sample data is large (n >30) or the pairs of values have differences that are from a population having a distribution that is approximately normal. (These methods are robust against departures for normality, so for small samples, the normality requirement is loose in the sense that the procedures perform well as long as there are no outliers and departures from normality are not too extreme.

Let Assume that:

d =individual difference between the two values in a single matched pair

μ_d= mean value of the differences *d* for the *population*of all pairs of data

$\bar{d}$=mean value of the differences *d* for the paired

s_d=standard deviation of the differences *d* for thepaired sampledata

n=number of paired sample data

Hypothesis Test Statistic for Dependent Samples

$$t = \frac{\bar{d}-\mu d}{\frac{s_d}{\sqrt{n}}} \quad (11)$$

where degrees of freedom = $n-1$.

P-values and Critical values can be obtained from t distribution table

Confidence Intervals for Dependent Samples

$$\bar{d} - \mathrm{E} < \mu d + E \qquad (12)$$

$$here E = t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

Critical values of $t_{\alpha}/2$: with $n - 1$ degrees of freedomresults.

Data Description

The data is collected from section of quality and university performance. The data consists of performance evaluation of 21 academic staff members at faculty of mathematics and computer science for both department of mathematics and department of computer science for year 2011-2012. The data contains three evaluation of each academic member of staff; to include head of department evaluation, and two students evaluation for two different subjects for each member of academic staff.

The Results And Discussion

Kappa(k) is used to measure the strength of agreement between head of department evaluation (HE) and students' evaluation (SE) for a member of staff for first subject (SEF), second subject (SES), and average of both (SEA). Kappa is lying between −1 and +1 . Where k = 1 for perfect agreement, $k = 0$ for no agreement in the sense of no relationship.The evaluations

grades are categorized into two categories < 80, and ≥ 80, as according to promotion instructions in Iraqi universities, teaching staff must get grade 70% to promote to lecturer and Assistant professor and 80% to promote to professor.

Kappa shows that fair agreement between head's evaluation and students' evaluation for the first subject, slight agreement between head's evaluation and students' evaluation for the second subject, and fair between head's evaluation and students' for average evaluation of the two subjects.

Table (1): Kappa between evaluation of head students for 1st subject

Evaluation			Students' Evaluation for Subject One		Total	Kappa(k)	
			< 80	≥ 80		Value	Interpretation
Head's Evaluation	**< 80**	Observed	9	0	9	0.300	Fair agreement
		Expected	7.3	1.7	9		
	≥ 80	Observed	8	4	12		
		Expected	9.7	2.3	12		
Total		Observed	17	4	21		
		Expected	17	4	21		

Table (2): Kappa between evaluation of head students for 2nd subject

Evaluation			Average Students' Evaluation for Subject Two		Total	Kappa(k)	
			< 80	≥ 80		Value	Interpretation
Head's Evaluation	**< 80**	Observed	8	1	9	0.125	Slight agreement
		Expected	7.3	1.7	9		
	≥ 80	Observed	9	3	12		
		Expected	9.7	2.3	12		
Total		Observed	17	4	21		
		Expected	17	4	21		

Pearson's correlation coefficient is used to measure a linear relationship between two evaluation data.. The highest correlation is (0.868) between the SEF and the SES. This is expected as the students are evaluating the same member of staff. There is a fairpositive relationship (0.300) between the HE and the SEF, and (0.222) between the HE and the SEA . There is a slighter weaker relationship (0.125) between HE and the SES.

Table (3): Kappa between evaluation of head students average

Evaluation			Average Students' Evaluation		Total	Kappa(k)	
			< 80	≥ 80		Value	Interpretation
Head's Evaluation	**< 80**	Observed	9	0	9	0.222	Fair Agreement
		Expected	7.7	1.3	9		
	≥ 80	Observed	9	3	12		
		Expected	10.3	1.7	12		
Total		Observed	18	3	21		
		Expected	18	3	21		

Table (4) Pearson's Correlations Coefficients

Variable (Evaluation)	*Head of Department*	*Students of 1st Subjects*	*Students of 2nd Subjects*
Head of Department	1		
Students of 1st Subject	0.432	1	
Students of 2nd Subject	0.388	0.868	1

A simple regression is performed on predictors the HE, and dependent variable the SEF. R square value indentifies the proportion of variance in HE accounts for by the SEF. In this case around 20%, of the variance of the HE is explained by the SEF, or by the SES, or by the SEA.

The analysis showed a significant (p = 0.051 > 0.05) difference between the mean of HE (81.90) and the mean of SEF (74.21). The regression coefficients for constant is not statistically significant (p=0.620 > 0.05) at 5% level of significance and the HE is also not

Table (5) Regression Coefficients[a] to Predict HE from SEF

Model	*Coefficients*		*t-Test*	*Sig.*
	B	*Std. Error*		
(Constant)	14.455	28.713	.503	.620
Head' evaluation	.730	.350	2.086	.051

(p=0.620 >0.05) at 5% level of significance and the HE is also not significant (p= 0.051 > 0.05) at 5% of significance(see table 6).

The regression coefficients for constant is not statistically significant (p=0.639 >0.05) at 5% level of significance and the HE is also not significant (p= 0.068 > 0.05) at 5% of significance(see table 8).

Table (6) Regression Coefficients[a] to Predict HE from SES

Model	*Coefficients*			
	B	*Std. Error*	*t-Test*	*Sig.*
(Constant)	-21.173	44.376	-.477	.639
Head' evaluation	1.046	.540	1.936	.068

a. Dependent Variable: students' evaluation of second subject

Table (7) Regression Coefficients[a] to Predict HE from SEA

Model	*Coefficients*			
	B	*Std. Error*	*t-Test*	*Sig.*
(Constant)	-3.359	35.123	-.096	.925
Head' evaluation	.888	.428	2.076	.052

a. Dependent Variable: students' average evaluation of second subject

The regression coefficients for constant is not statistically significant

(p=0.925 >0.05) at 5% level of significance and the HE is also not significant (p= 0.052 > 0.05) at 5% of significance(see table 10)

Table 8 illustrates that it is very obvious to note that the prediction of HE from SE to include SEF, SES, and SEA are not feasible.

Table (8) Regression Equation *to Predict the HE from the SE*

Degree of fit (R^2)	*Variable*		*Regression Equation*
	Independent (predictor)	*Dependent*	
0.186	*HE*	*SEF*	HE = 14.455 + 0.730 SEF
0.151	*HE*	*SES*	HE = -21.173 + 1.046 SES
0.185	*HE*	*SEA*	HE = -3.359 + 0.888 SEA

Assessment of New Questionnaire

To overcome the inconsistently between HE and SE. A New set of questioner are designed and the students are asked to evaluate a member of teaching staff using the new questionnaires (NQ) as well as the old questionnaire (OQ)provided by the department of quality and performance of the university .The paired t-test are used to compare the mean of NQ with OQ, and test the following hypothesis

H_0 : the mean of NQ = the mean of OQ OR H_0 : 81.2281= 46.4386

H_1 : the mean of NQ ≠ the mean of OQ OR H_1 : 81.2281 ≠ 46.4386

We examine the requirement to carry out t-test. Firstly the samples the samples data are paired (dependent) . The samples are simple random samples. The number of pairs of sample data is large ($n=57$ >30) the results of t-test will continue to be essentially reliable.Since the samples sizes are greater than 30, the results of t-test will continue to be essentially reliable, in addition to the pairs of values have differences that are approximately normal.

Table (9) Descriptive of difference

Evaluation	*Mean*	*N*	*Std. Deviation*	*Std. Error Mean*
NQ	81.2281	57	16.53245	2.18978
OQ	46.4386	57	18.72643	2.48038

Table (10) Dependent t-test to compare NQ with OQ

Paired Differences					*Paired t- test*		
		Std.	*95% C.I of the Difference*				
Mean	*Std. Deviation*	*Error Mean*	*Lower*	*Upper*	*T*	*d.f.*	*P-value*
34.789	15.692	2.078	30.626	38.953	16.738	56	.000

A paired t-test was used to determine whether there was a statistically significant difference in SE between the mean of the NQ (81.2281) and the mean of OQ (46.4386). The 95% of the NQ (76.8414, 85.6147)and the OQ (41.4698,51.4074) are clear.

A paired t-test is used on a sample of 57 SE to determine whether there is a statistically significant mean difference between the SE when NQ compared to OQ. There is a highly significant difference at ($t(56)=16.738$, $p< 0.000$) at 1% level of significance with (95% CI, 30.626 to 38.953).

In conclusion that there are a week relationship between the evaluation of the Head of Department and students evaluation. The NQ is more reliable than the OQ to evaluate the teaching staff.

References

1) Centra, John. 1993. Reflective faculty evaluation: enhancing teaching and determining faculty effectiveness (Jossey Bass higher and adult education series). San Francisco: Jossey-Bass Inc Pub.
2) Hobson, Suzanne M., and Donna M. Talbot. "Understanding Student Evaluations: What All Faculty Should Know." College Teaching 49, no. 1 (2001): 26-31.
3) Aleamoni, Lawrence M. "Student Rating Myths Versus Research Facts from 1924 to 1998." Journal of Personnel Evaluation in Education 13, no. 2 (1999): 153-166.
4) Jirovec, Ronald L., Chathapuram S. Ramanathan, and Ann Rosegrant-Alvarez. "Course Evaluations: What are Social Work Students Telling Us About Teaching Effectiveness?" Journal of Social Work Education 34, no. 2(1998): 229-236.
5) Johnson, Valen E. "Teacher Course Evaluations and Student Grades: An Academic Tango." Chance 15, no. 3 (2002): 9-16.
6) Gigliotti, Richard J., and Foster S. Buchtel. "Attributional Bias and Course Evaluations." Journal of Educational Psychology 82, no. 2 (1990): 341-351.
7) Greenwald, Anthony G., and Gerald M. Gillmore. "Grading leniency is a removable contaminant of student ratings." American Psychologist 52, no. 11 (1997): 1209-1217.
8) Remedios, Richard, and David A. Lieberman. "I liked your course because you taught me well:

the influenceof grades, workload, expectations and goals on students' evaluations of teaching." British Educational Research Journal 34, no. 1 (2008): 91-115.

9) Centra, John A. "Will Teachers Receive Higher Student Work? " Evaluations by Giving Higher Grades and Less Course. Research in Higher Education 44, no. 5 (2003): 495-518.
10) Liegle, J O and D S McDonald. Lessons Learned From Online vs. Paper-based Computer Information Students' Evaluation System. In The Proceedings of the Information Systems Education Conference 2004, v 21(Newport): §2214. ISSN: 1542-7382.
11) Spencer, Karin J., and LioraPedhazurSchmelkin. "Student Perspectives on Teaching and its Evaluation."Assessment & Evaluation in Higher Education 27, no. 5 (2002): 397-409.
12) Nasser, Fadia, and Barbara Fresko. "Faculty Views of Student Evaluation of College Teaching." Assessment & Evaluation in Higher Education 27.2 (2002): 187-198.
13) Beran, Tanya N., and Jennifer L. Rokosh. "Instructors' perspectives on the utility of student ratings of instruction." Instructional Science 37.2 (2009): 171-184.
14) Gaillard, Franklin D., Sonja P. Mitchell, and VahwereKavota. "Students, Faculty, AndAdministrators'Perception Of Students' Evaluations Of Faculty In Higher Education Business Schools." Journal of College Teaching & Learning 3, no. 8 (2006): 77-90.
15) TülaySaracbasi. Agreement Plus Disagreement Model for Agreement Data. Hacettepe Journal of Mathematics and Statistics. Vol. 40 (4) (2011), 609 – 616.
16) Anthony J. Viera, and Joanne M. Garrett. Understanding InterobserverAgreement:The Kappa Statistic. 360 May 2005 Family Medicine. Research Series
17) L. L. Lekena and A. Bayaga. Quality Assurance in Education: Student Evaluation of Teaching (SET). Int J EduSci, 4(3): 271-274 (2012).
18) Triola M.F. (2010) Elementary Statistics, Eleventh edition. Addison –Wesley.
19) Wayne W. D. (2010) Biostatistics. John Wiley & Sons, INC.
20) Hobson, Suzanne M., and Donna M. Talbot. "Understanding Student Evaluations: What All Faculty ShouldKnow." College Teaching 49, no. 1 (2001): 26-31.
21) Carletta, Jean. (1996) Assessing agreement on classification tasks: The kappa statistic. Computational Linguistics, 22(2), pp. 249–254.
22) Smeeton, N.C. (1985). "Early History of the Kappa Statistic". *Biometrics* 41: 795.

ملخص

تقييم أداء التدريسيين أمر أساسي للتقدم وتحسين الاداء الوظيفي. جميع الكليات في جامعة الكوفة تقيم أداء التدريسيين سنوي اويتم التقييم من قبل رئيس قسم والطلبه. بناءا عليه فأن الهدف الرئيسي للبحث هو ايجاد العلاقة بين تقييم التدريسيين من قبل رئيس القسم مع تقييم الطلبه.

وقد تم جمع البيانات الى 24 تدريسي من قسم الجودة والأداء في كلية الرياضيات وعلوم الحاسوب عن طريق.المشكلة هي أن النموذج الاستبيان الذي يتم ملئه من قبل الطلبه لايقيس مايفترض أن يقيس . ولهذا السبب تم تصميم نموذج جديد من الاستبيان،وطلب من الطلاب

أظهرت نتائج أن معدل تقييم التدريسيين من قبل رئيس القسم هو أعلى من تقييم الطلبه. وان النموذج الاستبيان الجديد هو أكثر موثوق من النموذج السابق.