A Statistical Study to Develop a Reliable Scale to Evaluate Instructors within Higher Institutions

SAID TAAN EL HAJJAR
Mathematical Sciences
Ahlia University
P. O. Box 10878
KINGDOM OF BAHRAIN
Saidth2000@yahoo.com

Abstract: It is not clear yet to many scientists in different fields that using any scale to test a certain data must have this scale at least reliable and valid. The purpose of this study is to develop a reliable and valid scale that will promote the measurement of the level of instructor evaluation within higher institutions. To satisfy this measurement, Cronbach's alpha is used to obtain a thorough understanding of remarkable questions/constraints that is initiated in the instructor evaluation form to evaluate the instructor properly. Discussions during focus groups of students and interviews recorded. Then, a model is developed, and collected data classified based on the questions arising during the discussions with the participants. Accordingly, the collected qualitative data are analyzed using a thematic approach in order to identify the basic constructs/variables of the model. There are five main constructs or antecedents representing the independent variables that have great influence on the instructor's assessment. They are Teacher's Behaviour (TB), Type of Courses (TC), Level of Students (LS), Mood of Students (MS), and Accuracy of Faculty's Evaluation (AFE). Each variable is explained by six to eight constraints. Such a scale is being tested through a pilot study in terms of reliability and content validity is being purified further through the collection of quantitative data. To further investigate and verify our theoretical framework, the study employs ordinary least square (OLS) test. OLS technique as a linear regression model is implemented in this study to examine the extent to which the independent variables influence the dependant variable (Faculty Evaluation (FE)), and the results of the regression analysis reveals that the model is significant.

Key–Words: Instructor evaluation, correlation, performance, reliability, validity, construct, scale

1 Introduction

Cronbach's alpha is a measure of internal consistency to indicate how closely related a group of items to each other. A high value of alpha is frequently used as indication that the items measure a latent construct. Cronbach's alpha is not a statistical test - it is a coefficient of reliability or consistency. One of the most software that is used to do this job is SPSS. If Cronbach's alpha is smaller than 0.60, then the researcher has to remove the items that have lowest " Item-Total Correlation" and/or the highest value of alpha if " Item Deleted". However, if alpha stays lower than 0.60 after repeated elimination, then a scale formed by the remaining items cannot be constructed [1]. This means that if alpha is greater than or equal to 0.6, then the result is good.

Although most of the scales employed to measure the level of the faculty performance have been purified and tested, it is still a foggy process of evaluation since most of the scientists doubt the ability of students to assessment. Since most of the universities around the world do the faculty evaluation, then it is necessary to create or develop a reliable scale of questions to satisfy the goal of the required evaluation.

Some universities prefer online faculty evaluation while others do not. This study does not focus on the mechanical way of the faculty evaluation; it focuses on the questions that are needed to be provided for students to reply whether the process of evaluation is done online or not. The qualitative research that has been done over 98 students at a reputable university in the Kingdom of Bahrain has developed a scale model that consists of five major independent variables that are needed to describe the performance of the instructor evaluation , and moreover, a quantitative research has tested the reliability and content validity of this scale through a pilot study over 25 students of this university to recognize the valid constraints that are required to be used in the survey questionnaire of instructor evaluation. The output of the regression analysis test used in this study proves the inaccuracy of the instructor evaluation at higher institutions based

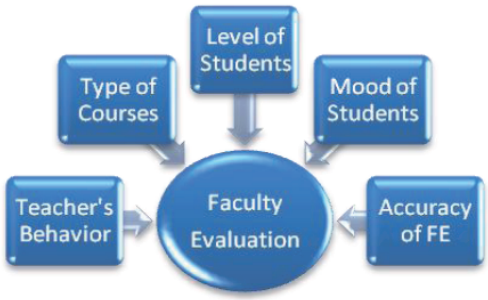

Figure 1: The Study Framework

on the students replies.

Questions are often used in different ways by different universities; the goal of this study is to overcome the difficulty in initiating valuable questions that each university will interpret in the same way. A good question should be short and straightforward. A questionnaire should not be too long, use plain English and the question should not be difficult to answer. Only through careful writing, editing, review, and rewriting can you make a good questionnaire.

In fact, this study is very important to both instructor and administrative to have a wide knowledge to have a reliable scale for faculty evaluation and to direct the aim of the output of this scale through the right track. If this happens in a proper way, then instructors are not going to be scared anymore about their results' evaluations obtained from their students through this survey, and students will be aware that their revenge from their instructors through instructor's evaluation is useless.

The current study attempts to consider various issues surrounding faculty evaluation and categorizes the independent variables under five categories: 1- Teacher's Behaviour (TB), 2- Type of Courses(TC), 3-Level of Students (LS), 4- Mood of Students (MS), and 5- Accuracy of Faculty Evaluation (AFE). The dependent for the five variables is "Faculty Evaluation (FE)." Fig.1 explains the framework of the study.

Hypothesis 1: Teacher's behaviour will have an effect on faculty evaluation.

Hypothesis 2: Type of courses will have an effect on faculty evaluation.

Hypothesis 3: Level of students will have an effect on faculty evaluation.

Hypothesis 4: Mood of students will have an effect on faculty evaluation.

Hypothesis 5: Faculty evaluation is not accurate and has an effect on faculty evaluation.

The hypothesized relationships among features and constructs will be tested through the model (Fig.1). Paths, in the model, indicate the positive effect of each feature on its corresponding construct.

2 Literature Review

A significant teacher evaluation structure should reveal a set of major certainties about good instruction. Unfortunately, most evaluations converse an overwhelming message-that all teachers are almost the same, and that the main purpose of evaluation is to recognize and remove a little number of teachers who are critic nastily unskilled. The classic evaluation form proposes that good teaching consists of performing an ordinary set of schedules that are largely unrelated to student engagement or learning. The standards, which are suggested for Teacher Evaluation 2.0, are founded on a far different set of core principles about the power of great instructors and the critical role evaluations play in developing them. A teacher's main professional responsibility is to guarantee that students learn. Therefore, measures of student learning should play a primary role in teacher evaluations. This does not mean that teacher evaluations should be based exclusively on the results of standardized tests, or based on the results of any single assessment. But it does mean that teachers should be responsible for helping students make measurable progress against ambitious learning standards [2].

Governor Christie recommends basing teacher evaluation significantly on student test scores. However, Christopher Cerf recommended that teacher tenure and dismissal, as well as compensation decisions, should be based largely on student assessment data [3]. Therefore, the goal is to estimate the extent to which a particular instructor contributes to the learning achievement of a group of students assigned to that teacher in a given semester. Unfortunately, while this all sounds good, it just does not work, at least not well enough to even begin considering using it for making high- bases decisions about teacher occupancy, firing or reimbursement.

Critical to instructors' enduring development throughout the semester, discussion was not confined to instructor and evaluator. The substitution between instructor and evaluator regularly enlarged to include the major content developer and academic department chair when curricular issues arose. Additionally, the promotion of formative substitutions highlighted the expectation that teachers would reply with elasticity and innovation to more efficiently address their students' needs. The dialogic nature of the formative assessments was paralleled in the summative assessment, which also included several perspectives (teacher, evaluator, and student, departmental admin-

istrator) to obtain a comprehensive and balanced assessment of the teacher's performance [4]. Instructor's evaluation could be done online. In writing about the unique constraints of evaluating faculty online, Tobin identifies as a major marker of traditional models the singularity of evaluator's interaction with individual faculty member [5]. Apparently, a customary evaluator only requires one visit to a face-to-face classroom because the markers of effective classroom teaching are easily observable to someone who is also teaching on-ground. In contrast, in the online environment, academic administrators, many of whom only teach face-to-face, are tasked with the evaluation of online faculty. Some scientists argue in favour of instructor evaluation websites. These websites sites are a simple way to converse precious information to fellow students about instructors. This information may help students create an educated decision in choosing a good instructor. Instructors are in a position of power, and students are not. This is one way that students may fight back against tough or serious instructors who abuse their power, and offer positive feedback about flexible instructors. In addition, these websites hold teachers more liable. No one wants negative comments written about him or her on the web, so teachers might think twice about how they are treating their classes. They may be less likely to try to instruct their students. The ambiguity can encourage students to evaluate instructors more honestly. Traditional instructor evaluations are unidentified too, but teachers can sometimes recognize handwriting or otherwise identify a student. As these websites turn out to be more popular, more evaluations are available about each instructor. This provides students a better idea about how the instructor is distinguished in general, as opposed to how the instructor is distinguished by one or two students [6]. On the other hand, some scientists argue against instructor evaluation websites. Some feel this is an assault of solitude. Certainly, instructors have the authority to grade their students every semester, but they do not have the authority to post those grades on the website for anybody in the world to see. These evaluations are not essentially precise. Students may be more likely to post if they have either a very positive or a very negative experience with his/her instructor. While many students post reasonable evaluations on these sites, some students use these websites as revenge for a very low grade and an easy prospect to blaze a teacher. This is not supportive to students, who are not getting an accurate review of a teacher, and it is definitely not supportive to the teacher. There is no warranty that the evaluations are from students. Anyone who might be angry at the instructor or a jealous colleague, can post comments and pretend to be a student. These websites are not considered seriously by anyone who composes a difference in education. If you want your opinion about an instructor to matter, take the time to contact the department or the dean. Instructors have feelings too. How would you feel if someone posted sarcastic comments about you on the web?

Evaluations should improve the teacher performance, but in most cases, they do not. "Our system of teacher evaluationfrustrated teachers who feel that their good work goes unrecognized and ignores other teachers who would benefit from additional support." [7].

In fact, Evaluation skills are too hard to be created by anyone. "As an undergraduate, it can be easy to be intimidated by academic discourse. The fear is that you will simply be parroting a standard line or that your inability to deconstruct an argument will be apparent. Developing critical thinking and evaluation skills is definitely a learned quality that needs to be practiced." [8]. Certainly, evaluation involves giving reasons for beliefs and decisions and choosing how to act, criticizing ideas constructively, and modifying ideas in response to criticism [9]. "Even with the best models and data, teacher ratings are highly inconsistent from year to year, and have very high rates of misclassification" [2]. A recent study shows that there is a 35% chance of identifying an average teacher as poor, given one year of data, and a 25 % chance given three years. Getting a good rating is a statistical crapshoot. Rating the same instructor with the same students, but with two different tests in the same subject, we absolutely obtain different results. University of California at Berkeley economist Jesse Rothstein, re-evaluating the findings of a much-touted Gates Foundation study, noted that more than 40% of instructors who put in the bottom quarter on one test were in the top half when using an alternative test. So teacher ratings based on the state assessment were only slightly better than a coin toss at identifying which teachers did well using the alternative assessment.

The Quality Assurance Agency (QAA) who examines the quality of the higher institutions insists on having Instructor Evaluation at those institutions. So whether we believe in the effectiveness of Instructor Evaluation at these institutions or not, Faculty Evaluation is a major document to be implemented and placed at those institutions' profiles where these higher institutions struggle to get full confidence from the QAA. Hence, a reliable scale to estimate -as close as possible- the performance of the instructor in the universities should be produced. With the rapid change in all types of faculty evaluation forms (hard copies or online), there is a need to know the effect of these evaluations on the faculty's performance at higher institutions. There are many criticisms about

its process not only from faculty members but also from students' comments. Students fill the faculty evaluation form quickly and inaccurately, they believe that faculty evaluation is useless, students assure that the faculty evaluation has a negative effect on faculties' performance, students use faculty evaluation as revenge for a bad grade. As we see, these comments are remarkable and should be examined to know their validations. The theoretical model (Figure1) will be tested in the present paper, as an initial step in developing and revising the model for further research investigations.

3 Research Methodology

This paper occupied quantitative research methodology. If a problem involves identifying factors that affect an outcome, then a quantitative approach is best [10]. A quantitative pilot survey is intended at identifying the factors influencing the adoption of faculty evaluation at higher institutions. The primary data collected is quantitative in nature. The population for this study is defined as all students who are enrolled at a reputable university in the Kingdom of Bahrain. The samples for the qualitative research and the pilot survey are selected randomly from students at this university. 98 respondents are interviewed. 25 responses are selected randomly from the pilot survey and a scale for an instructor evaluation is initiated. Once the scale is developed and the pilot survey data is collected, this research used SPSS software to establish the reliability of the various variables and factors included in the scale. This test is about the influence of factors affecting positively or negatively students' behaviours toward their instructor evaluation and his/her performance, and their overall opinions in the faculty evaluation process. The data obtained are analyzed by using SPSS V.18, various analytical tools, such as Correlation test and Regression Analysis, implement to interpret data. SPSS is a powerful tool that provides statistical results to answer the research questions and hypothesis. Primary, the current study attempts to use correlation and regression analysis to investigate the relation, cause and effects between variables.

To accomplish the previous mentioned research questions, the data for this study is collected through self-administered questionnaires. Due to high reliability and validity, the questionnaire of the current study is developed from a qualitative research, and a pilot survey is selected randomly from a reputable university in the Kingdom of Bahrain along which reliability and content validity are to be tested. The survey of this study is a 3-page self-administered questionnaire that consists of five independent variables. Using 5-point Likert-type scale, respondents rate their agreement from (1) 'Strongly Disagree', to (5) 'Strongly Agree'. At the end of the questionnaire, respondents are asked to reveal any other comments in the space provided to express their opinion. Table1 summarizes the variables and the number of questions/items of each variable.

Table 1: Summary of the variables and number of question/items

Variables	Questions	Type of Variable (2)
1. Teacher's Behaviour (*TB*)	*8* questions	*X*
2. Type of Courses(*TC*)	*7* questions	*X*
3. Level of Students (*LS*)	*6* questions	*X*
4. Mood of Students (*MS*),	*6* questions	*X*
5. Accuracy of Faculty Evaluation (*AFE*).	*6* questions	*X*
6. Faculty Evaluation (*FE*)	-	*Y*

Notes: Items designated are scored 1,2,3,4 and 5, respectively for responses strongly Disagree, and Disagree, Not Sure, Agree, Strongly Agree

4 Reliability and Validity Results

In order to make sure that the used survey is reliable and has content validity, the study conducted pilot study, by sending the questioners to twenty-five respondents randomly, and then assessed their responses. To find out whether the survey is reliable or not, we will measure the internal consistency, which is a useful method for testing reliability. Reliability has been identified as the similarity of results supplied by independent but comparable measures of the same object, trait, or construct [11]. On the other hand, Validity is directed to know the degree of measuring the constructs along which they are intended to measure. To guarantee a significant reliability the study implements Cronbach's alpha test, which is a known statistical measure used for testing the reliability of constraints. Consequently, Crombach's Alpha computed and the results were compared with the 0.7 level recommended as a cut-off point, which showed that scales were reliable [12]. However, a reliability between 0.5 and 0.6 is considered suffice by [13].

When testing Reliability, the option "Scale if item deleted" should be chosen. This indicator will result in the mean and the variance for the original dimen-

sion after removal of the relevant statement, in the correlation between the respective statement and one factor, consisting of the other statements, and Cronbach's Alpha for the scale after the elimination of the relevant statement.

The data analysis method has started by coding the collected data. The responses obtained were analyzed using SPSS V.18. The use of SPSS computer program will assist in the coding and analysis process of the data collected from the survey questionnaires. The study will use the following statistical tools:

- Cronbach's alpha test was used to check the reliability and the internal consistency of the questionnaire items. Methodologist suggested that Alpha value of 0.70 is deemed a desirable threshold, while a value of 0.60 is acceptable as the minimum cut-off point [12].
- Personal Correlation was implemented to test the relationship between the variables of the study. It is very important to test the relationship between the study variables before conducting regression analysis [14].

Linear Regression Analysis was used in this study to determine the prediction power between the dependent and the independent variable.

Reliability is considered as the internal consistency of a scale that measures the degree to which the items are constant and homogeneous. After the normality assessment, studies insisted on that a study should measure the reliability of a construct before performing statistical analysis such as correlation and regression test. Similarly, Cronbach's and Meehl, (1955) suggested that a study should conduct various tests to make sure that the expanded scales/items of the questionnaire are consistent in what were intended to measure before measuring the correlation between variables and regression tests. Therefore, to test the internal consistency of the study variables, Cronbach's alpha test was involved.

The coefficient varies from 0 to 1, but according to Pallant (2001), only a coefficient of scale above 0.7 is a construct with valid measurement. Similarly, other studies suggested that coefficient near 0.9 represent highly consistent scales, while those nearing 0.7 reflects a more moderate level of consistency whilst alpha values below 0.3 indicates that the items have little in common. Therefore, the current study considered coefficient of 0.6 as the minimum cut-off point to the construct reliability.

Studies suggested that to enhance the coefficients α, a study should delete some items (that scored negatively) when a variable scored value less than 0.6, however, if the coefficients value did not reach the minimum cut-of-point then it should be omitted [15].

Table 2: Teacher's Behavior (TB) Output

	Item-Total Statistics		Reliability Statistics
TB	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Cronbach's Alpha
TB1	*0.402*	*0.534*	
TB2	*0.349*	*0.560*	*0.601*
TB3	*0.765*	*0.418*	
TB4	*0.451*	*0.517*	
TB5	*-0.218*	*0.717*	
TB6	*0.323*	*0.562*	
TB7	*0.224*	*0.589*	
TB8	*0.307*	*0.566*	

The hypothesized relationships and reliability described in the model were tested using SPSS software for the type of internal consistency reliability and the results were as follows:

1. Teacher's behavior will have an effect on faculty evaluation (TB)

This variable is explained by eight independent variables or factors.

(i) I give a very friendly teacher high rank in FE (TB1).

(ii) I give an easygoing teacher high rank in FE (TB2).

(iii) If I get high grades in my exams I will give the teacher high rank in FE (TB3).

(iv) If I get low grades in my exams I will give the teacher low rank in FE (TB4).

(v) I give a high rank in FE to the teacher who allows me to cheat during exams (TB5).

(vi) A teacher who does not blame me when I use my I Phone or mobile during his lecture I will give him high rank in FE (TB6).

(vii) A faculty who makes revision before the exam gets high rank on FE (TB7).

(viii) A serious teacher would have a low rank in FE (TB8).

To see how much TB variable is explained by these factors is represented in Table 2 below.

Overall alpha is 0.601, which indicates good internal consistency among the eight Teacher's Behaviour items. However, the correlation between TB5 and the sum of the other items is r = -0.218. This means that there is a weak negative correlation between the scoreTB5 and the combined score of the other seven items. This result assesses how bad TB5

Table 3: Type of Courses (TC) Output

	Item-Total Statistics		Reliability Statistics
TC	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Cronbach's Alpha
TC1	*0.499*	*0.482*	
TC2	*0.601*	*0.466*	*0.593*
TC3	*0.546*	*0.473*	
TC4	*0.513*	*0.486*	
TC5	*0.099*	*0.633*	
TC6	*0.094*	*0.635*	
TC7	*0.022*	*0.652*	

score is internally consistent with composite scores from all other items that remain. Moreover, Cronbach's alpha would increase from 0.601 to 0.717 if item TB5 were deleted. Therefore, since alpha is increased by a large degree from deleting TB 5, and since TB 5 does not correlate very well with the composite score from other items, then there is a statistical reason to drop item TB5. Although there is a fair correlation between some of the other items and the sum of its other corresponding items, the Cronbach's alpha, if either of the other items were deleted, would drop from the overall 0.601 to a number less than it would be. Thus, the remaining seven items appear to be useful and contribute to the overall reliability of Teacher's Behaviour.

2. Type of courses will have an effect on faculty evaluation (TC).

This variable is explained by seven independent variables or factors.

(i) Courses based on knowledge have positive effect on FE (TC1).

(ii) Courses based on applications have positive effect on FE (TC2).

(iii) Courses which require a lot of analysis have negative effect on FE (TC3).

(iv) Courses which require a lot of synthesis have negative effect on FE (TC4).

(v) Courses which require lab instructions have negative effect on FE (TC5).

(vi) Hard courses have negative effect on FE (TC6).

(vii) Easy courses have positive effect on FE (TC7).

To see how much TC variable is explained by these factors is represented in Table 3 below.

Overall alpha is 0.593, which indicates good internal consistency among the seven Type of Courses items. However, the correlation of either of TC5, TC6, or TC7 with the sum of the other items is 0.099, 0.094, and 0.022 respectively. This means that there is a very weak correlation between either of the score TC5, TC6, or TC7 with the combined score of the other six items. This result assesses how bad TC5, TC6 and TC7 scores are internally consistent with composite scores from all other items that remain. Moreover, Cronbach's alpha would increase from 0.593 to 0.633, 0.635, and 0.652 if one of the items TC5, TC6, or TC7 were deleted. Therefore, there is a statistical reason to drop the three items TC5, TC6, and TC7. Although there is a satisfied or weak correlation between some of the other items and the sum of its other corresponding items, the Cronbach's alpha, if either of the other items were deleted, would drop from the overall 0.593 to a number less than it would be. Thus, the remaining four items appear to be useful and contribute to the overall reliability of Type of Courses

3. Level of students will have an effect on faculty evaluation (LS).

This variable is explained by six independent variables or factors.

(i) A sophomore student is ready to evaluate a teacher (LS1).

(ii) A junior student is ready to evaluate a teacher (LS2).

(iii) A senior student is ready to evaluate a teacher (LS3).

(iv) No matter the level of student, he is ready for FE (LS4).

(v) No matter the level of student, he is not ready for FE (LS5).

(vi) As a student, I have enough idea about FE (LS6).

To see how much LS variable is explained by these factors is represented in Table 4 below.

Overall alpha is 0.505, which indicates satisfied internal consistency among the six levels of students' items. However, there is a statistical reason to drop the two items LS1and LS4. Although there is a weak correlation between some of the other items and the sum of its other corresponding items, the Cronbach's alpha, if either of the other items were deleted, would drop from the overall 0.505 to a number less than it would be. Thus, the remaining four items appear to be useful and contribute to the overall reliability of Students' Levels.

4. Mood of students will have an effect on faculty evaluation (MS).

This variable is explained by six independent variables or factors.

Table 4: Level of Students (LS) Output

	Item-Total Statistics		Reliability Statistics
LS	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Cronbach's Alpha
LS1	*0.108*	*0.529*	*0.505*
LS2	*0.563*	*0.302*	
LS3	*0.445*	*0.366*	
LS4	*0.017*	*0.583*	
LS5	*0.336*	*0.417*	
LS6	*0.189*	*0.499*	

Table 5: Mood of Students (MS) Output

	Item-Total Statistics		Reliability Statistics
MS	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Cronbach's Alpha
MS1	*0.381*	*0.593*	*0.640*
MS2	*-0.097*	*0.748*	
MS3	*0.785*	*0.408*	
MS4	*0.785*	*0.408*	
MS5	*0.559*	*0.525*	
MS6	*-0.016*	*0.722*	

(i) I usually feel well when I evaluate a teacher (MS1).

(ii) When I am in good mood, I give good rank in FE (MS2).

(iii) When I am in bad mood, I give bad rank in FE (MS3).

(iv) The time is not suitable to fill the FE form just before the final exam (MS4).

(v) I prefer to do FE online because I will fill the form when I am in good mood (MS5).

(vii) No matter my mood is, I evaluate the teacher properly (MS6).

To see how much MS variable is explained by these factors is represented in Table 5 below.

Overall alpha is 0.640, which indicates good internal consistency among the six student's Mood items. However, there is a statistical reason to drop the two items MS2and MS6. Moreover, there is a significant correlation between some of the other items and the sum of its other corresponding items, the Cronbach's alpha, if either of the other items were deleted, would drop from the overall 0.640 to a number less than it would be. Thus, the remaining four items appear to be useful and contribute to the overall reliability of Student's Mood.

Table 6: Accuracy of Faculty Evaluation (AFE) Output

	Item-Total Statistics		Reliability Statistics
AFE	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Cronbach's Alpha
AFE1	*-0.149*	*0.751*	*0.612*
AFE2	*0.435*	*0.532*	
AFE3	*0.499*	*0.508*	
AFE4	*0.659*	*0.420*	
AFE5	*0.065*	*0.665*	
AFE6	*0.754*	*0.369*	

5. Faculty evaluation is not accurate (AFE).

This variable is explained by six independent variables or factors.

(i) I usually evaluate my teacher seriously (AFE1).

(ii) I just fill the FE form because I am obliged to do that (AFE2).

(iii) I take enough time to read the question in FE form before I give my rank (AFE3).

(iv) I usually give ranks in a FE form randomly (AFE4).

(v) The ideas of my classmates about my teacher affect me when I give a rank for my teacher in a FE form (AFE5).

(vi) Overall, faculty evaluation is not accurate (AFE6).

To see how much FEA variable is explained by these factors is represented in Table 6 below.

Overall alpha is 0.612, which indicates good internal consistency among the six Accuracy of Faculty of Evaluation items. However, there is a statistical reason to drop the two items AFE1and AFE5. Although there is a significant or weak correlation between some of the other items and the sum of its other corresponding items, the Cronbach's alpha, if either of the other items were deleted, would drop from the overall 0.612 to a number less than it would be. Thus, the remaining four items appear to be useful and contribute to the overall reliability of Accuracy of Faculty

Table 7: Reliability Results

Variables	Questions	Type of Variable (2)
7. Teacher's Behaviour (*TB*)	8 questions	*X*
8. Type of Courses(*TC*)	7 questions	*X*
9. Level of Students (*LS*)	6 questions	*X*
10. Mood of Students (*MS*),	6 questions	*X*
11. Accuracy of Faculty Evaluation (*AFE*).	6 questions	*X*
12. Faculty Evaluation (*FE*)	-	*Y*

Evaluation. Hence, the model constraints of the major variables will be reduced as follows:

We can say that the remaining items/variables used in this study were highly reliable and acceptable, and the content validity is satisfied.

5 Regression

To further investigate and verify our theoretical framework, the study employed ordinary least square (OLS) test. OLS technique as a linear regression model has implemented in this study to examine the extent to which the independent variables influence the dependant variable. The regression analysis uses to examine the influence of independents variables (Teacher's Behaviour (TB), Type of Courses (TC), Level of Students (LS), Mood of Students (MS), and Accuracy of Faculty's Evaluation (AFE) on the dependent variable Faculty Evaluation (FE). Tables 8, 9, and 10 summarized the results of the OLS analysis. The results of the Regression analysis reveals that the model is significant (F= 8.297; p ¡ 0.05) and the correlation between the independent variables and the dependent variable is strong (R = 0.966). The coefficient of determination (R2) for the regression is (0.933) indicating that (93.3%) of the variation in the dependent variable (Faculty Evaluation) was explained by the independent variables included in the regression model. This mean that 6.7% (100% - 93.3%) of variance are not included in the model.

5.1 ANOVA Test

The Analysis Of Variance (or ANOVA) is a powerful and common statistical procedure in the social sciences that compares the means of different variables. The following table presents the findings of ANOVA Test.

Table 8: Regression 1, ANOVA

Model		Sum of Squares	d.f.	Mean Square	F	Sig.
1	Regression	17.681	15	1.179	8.297	.002[a]
	Residual	1.279	9	.142		
	Total	18.960	24			

a. Predictors: (Constant), Teacher's Behavior, Type of Courses, Level of Students, Mood of Students, Accuracy of Faculty Evaluation.
b. Dependent Variable: Faculty Evaluation

Table 9: Regression 1 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.966[a]	.933	.820	.37693

a. Predictors: (Constant), Teacher's Behavior, Type of Courses, Level of Students, Mood of Students, Accuracy of Faculty Evaluation

Table 10: Regression 1

Model	Under standardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-1.466	.611		-2.401	.027
Teacher's Behavior	.409	.133	.338	3.082	.006
Type of Courses	.235	.087	.311	2.699	.014
Level of Students	.224	.123	.215	1.821	.044
Mood of Students	.246	.087	.292	2.835	.011
Accuracy of Faculty's Evaluation	.295	.122	.262	2.426	.025

a. Dependent Variable: Faculty Evaluation

Table 11: Summary of hypotheses testing

#	Hypothesis	Rejected/ supported
H1	*H1: Teacher's behaviour will have an effect on faculty evaluation.*	**Supported**
H2	*H2: Type of courses will have an effect on faculty evaluation.*	**Supported**
H3	*H3: Level of students will have an effect on faculty evaluation.*	**Supported**
H4	*H4: Mood of students will have an effect on faculty evaluation.*	**Supported**
H5	*H5: Faculty evaluation is not accurate and has an effect on faculty evaluation.*	**Supported**

The results of regression indicated that Faculty Evaluation is explained by the five variables: Teacher's Behaviour, Type of Courses, Level of Students, Mood of Students, Accuracy of Faculty Evaluation. The effect of the five antecedents influence on Faculty Evaluation was assessed by linear regression analysis by estimating the following equation:

$$\begin{aligned}\text{Faculty Evaluation} &= \beta_1 \text{Teacher's Behaviour} \\ &+\beta_2 \text{Type of Courses} + \beta_3 \text{Level of Students} \\ &+\beta_4 \text{Mood of Students} \\ &+\beta_5 \text{Accuracy of Faculty's Evaluation} + \varepsilon\end{aligned} \tag{1}$$

Linear regression is used to model the value of a dependent scale variable based on its linear relationship to one or more predictors. This table shows the coefficients of the regression line. It states that the five antecedents influence the Faculty Evaluation is equal to:

$$\begin{aligned}&0.409 \text{ Teacher's Behaviour} \\ &+0.235 \text{ Type of Courses} \\ &+0.224 \text{ Level of Students} \\ &+0.246 \text{ Mood of Students} \\ &+0.295 \text{ Accuracy of Faculty's Evaluation} \\ &-1.466\end{aligned} \tag{2}$$

As shown in Fig. 2, the results of the regression line indicated that the variance in the Faculty Evaluation was explained by the five variables.

Finally, the results of hypothesis testing are presented in Table 11 below.

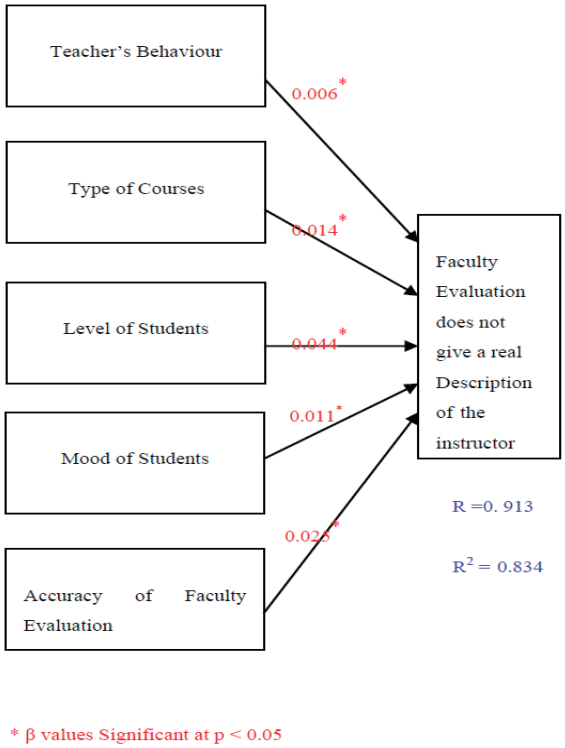

Figure 2: Theoretical Framework related to the Regression Analysis

6 Conclusions and Recommendations

This study has initiated a reliable with content validity scale on faculty evaluation that consists of five constraints. The constraint Teacher's Behaviour (TB) is explained by seven factors and each one of the other four constraints Type of Courses (TC), Level of Students (LS), Mood of Students (MS), and Accuracy of Faculty's Evaluation (AFE) is explained by four factors. Moreover, these constraints prove that there is a significant influence of these constraints on Faculty Evaluation. This effect could be positive or negative. It depends on the teacher's behaviour, whether he/she is very friendly, easy going, grades donation, makes revision before exams, or he/she is a serious teacher; on the type of courses, whether it is based on knowledge, applications, analysis or synthesis; on the Level of the student, whether he/she is a junior or a senior student; on the mood of the student, whether he/she is feeling well or bad, and whether the process is done immediately before the exam or online; and finally on the point of view of students towards the accuracy of Faculty Evaluation, whether students are obliged to fill the evaluation form or not, they take their times to read the questions or otherwise they do it randomly. At last but not least, the results of this study recommend not involving the output of Faculty Evaluation in any serious issue related to the corresponding faculty, such as faculty promotion, increments, etc. In fact, no one can evaluate anyone; the only one who may evaluate the instructor is the instructor himself. He is the one who knows the level of the students and the type of the course. Although Faculty Eval-

uation is not accurate, the instructor is able to filter the students' comments towards his performance and then he can modify his behaviour accordingly. This scale now is ready and purified to us at higher institution. It recommended to other researchers to make an exploratory and confirmatory factor analysis for this scale so that this study could be improved by developing a structure equation model for this scale.

References:

[1] W. Janssens, K. Wijnen, P. D. Pelsmacker, P. V. Kenhove, *Marketing Research with SPSS*, Prentice Hall, 2008.

[2] Teacher Evaluation 2.0, http://www.tntp.org; info@tntp.org. 2010

[3] B. Bruce, Seven reasons why teacher evaluations won't work. http://www.northjersey.com/news/.../evaluation_031311.html. 2011.

[4] Jean B. Mandernach, A Faculty Evaluation Model for Online Instructors: Mentoring and Evaluation in the Online Classroom, 2005.

[5] T. Tobin, Best Practices for Administrative Evaluation of Online Faculty, *Online Journal of Distance Learning Administration*, Vol. 7, No. 2, 2004.

[6] Naomi R. Gladen, *Are Teacher Evaluation Sites Like Rate My Professors.com a Good Idea?* College Professor Ratings Website. http://www.suite101.com/content/college professor -ratings -websitea21713#ixzz1H8LK 3lio. 2007.

[7] A. Duncan, *Teacher Evaluation 2.0.*, http://www.tntp.org; info@tntp.org. 2010.

[8] Griffith Graduate Project. Critical Evaluation Toolkit, Griffith Institute of Higher Education. http://www.griffith.edu.au/centre/gihe/griffith _graduate.2003.

[9] M. Lipman, *Higher order thinking in Institutions of higher learning. Thinking in Education: Cambridge University Press*, Cited in Slade, C. (1995). Unicorn Vol. 21, No. 1.1991. pp.39.

[10] J. Creswell, *Research design: Quantitative, qualitative and mixed methods approach*, Sage Publications, USA, 2003.

[11] G. Churchill and D. Lacobucci, *Marketing Research: Methodological Foundations*, 8th Ed. U.S.A.: South-Western Thomson Learning, 2002.

[12] J. C. Nunnally and I. H. Bernstein, *Psychometric Theory*, (3 ed.). New York : McGraw - Hill, 1994.

[13] R. A. Peterson, A Meta-Analysis of Cronbach's Coefficient alpha, *Journal of Consumer Research*, Vol. 21, 1994, No. 2, pp. 381-391.

[14] W . G . Zikmund, *Business research methods*, 6th ed. Orlando, US : Dryden Press. 2000.

[15] J. F. Hair, R. E. Anderson, R. L. Tatham, and W . C. Black, *Multivariate Data Analysis*, 5 th Edition, Prentice Hall: Upper Saddle River, NJ. 2006