Texila International Journal of Academic Research
Volume 5, Issue 1, May 2018

Two Important Predictors of Quality Instruction: Teacher Ratings and Student Outcomes: Correspondence and Interpretation

Article by Nusrat Ara Begum
Faculty of College of Engineering; Senior lecturer of Mathematics and Statistics, Effat University Jeddah
E-mail: nbegum@effatuniversity.edu.sa

Abstract

This quantitative study determined the coefficient of correlation between teacher ratings and students' grades in the final exam of an elementary statistics course GSTA 140 and their reliabilities (overall consistency of a measure) at Effat University in Jeddah Saudi Arabia. The study was conducted with a cluster convenience sample of 558 students, registered in 22 GSTA 140 courses over a period of 3 years (fall 2014 to fall 2017). To avoid the effect of variability of different teaching styles on ratings, only those courses were collected, which are taught by the same instructor. The study, designed and inferences obtained, is based on Pearson product-moment Statistical correlation test, Normality test and Empirical rule. The results showed that there is a very weak positive correlation between teacher ratings and student achievement, however; their independent standard deviations (a measure of variation) are approximately equal. Normality tests showed that the data set for the students' grades is well modeled by a normal distribution as compared to the data set of teacher ratings. Results also indicated that the application of the empirical rule is the optimum approach to interpret the reliability of the data of teacher ratings surveys and students' final grades and hence prediction can be made on the bases of these values. The findings of this study could prove useful to university, regional, and kingdom, implementing similar formal teacher evaluation framework as Effat University.

Keywords: *Teacher Ratings, Student outcomes, Normal distribution, Coefficient of linear correlation, Standard deviation, Empirical rule.*

Background and purpose of the study

Public and private universities in Saudi Arabia are striving hard to modernize their educational system to compete with international standards, further they hope to meet Saudi Arabia's vision 2030 which states "[1]We will continue investing in education and training so that our young men and women are equipped for the jobs of the future. We want Saudi children, wherever they live, to enjoy high quality, multi-faceted education. We will invest particularly in developing early childhood education, refining our national curriculum and training our teachers and educational leaders".

Effat University is one of the best universities in Saudi region striving hard to provide high standard education to its students. This university accommodates four colleges, Engineering, Architecture, Business and Humanities, which further subdivided into many departments. All departments 'students must study 42 credit hours of the general education program. The general education program aims to help the students acquire introductory background information in a variety of disciplines ranging from scientific, cultural, reasoning, to humanity and health wellbeing. In addition, it targets at equipping the students with shared values and ethics.

One of these disciplines is "Introduction to Statistics" the purpose of teaching this course is to prepare the students for the further research works to find the solutions of their problems scientifically.

[1] http://vision2030.gov.sa/en/node/8

DOI: 10.21522/TIJAR.2014.05.01.Art004
ISSN: 2520-3088

To monitor, check and maintained the quality of curriculum and teaching process regularly, at the end of every semester, the students asked to complete a course evaluation survey for all the taught courses. These surveys consist of three parts. The first is concerned with the quality of the course, while the second part refers to the faculty performance. In this part, the students asked to assess the faculty's readiness, knowledge, delivery, methods of assessment and ways of treating the students. The rating is categorized by " strongly agree, agree, disagree, strongly disagree and no response", and third section of the survey concludes with an open-ended section where the students are requested to respond at their own will to questions on how to improve the course or what they like/dislike about it.

Al though these surveys are significant to both the instructors and the human resources department. As they are considered an important, measure of effective teaching and determines the sustainability of teacher job.

Unfortunately, there is no objective way to draw conclusions from these surveys with 100% certainty and hence, their interpretation has become a challenge. Often, the nature of student feedback is assumed relative to the instructor's teaching skills, the higher the positive rating, the better the teacher.

This research makes an effort to find the relationship between a teacher's evaluation rating and student achievement in terms of their final exam grades. Furthermore, it compares the stability and dependability of these two variables statistically.

Previous researches and their findings

Teacher ratings (also known as student evaluation of teaching or course evaluations) are one of the most studied topics in higher education, with several thousand research articles and books addressing various aspects of this topic over the past 100 years.

Despite of ambiguity, misperceptions about ratings and problematic use of their result, course evaluation surveys are the most commonly used methods for evaluating and getting feedback on teaching, in all over the world.

Several techniques have been applied to determine the extent to which ratings measure effective teaching. One set of studies has examined that relationship between ratings and student achievement in a course. The results of this approach have been mixed. In the 80s and 90s, some meta-analyses indicated a moderate positive correlation between teacher ratings and student outcomes as measured by exam grades. In general, these findings show that overall ratings of the course or instructor ('Overall this is an outstanding instructor' 'Overall this is an admirable course') show a consistent positive relationship to achievement (i.e., in classes where the achievement level is high, instructors tend to receive better ratings). However, a more recent meta-analysis showed a fluctuating positive relationship. Other studies have compared student ratings results with the ratings of peers and of experienced faculty, and, in general, these studies have shown positive correlations. Investigators have also instituted that rating of overall instructional excellence are best explained by ratings of certain approaches to teaching (e.g., arranging the class, appreciating involvement, and establishing rapport) rather than unnecessary factors. Finally, a recent research suggesting that an instructor's favorable rating in one course do not necessarily predict their performance in subsequent courses. This could mean that teachers with good ratings may simply be "teaching to the exam" but not approaching deep learning, or it could show a weakness of curricular alignment between the instructor's course and subsequent courses.

Some very recent research and google trend pictures are given as follows.

2 Ingrid Yvonne Williams Medlock. (December 2017). Teacher evaluation ratings and student achievement: what's the connection?

[3]Neelie B. Parker. (May 2017). Framework for teacher evaluation: examining the relationship between teacher performance and student achievement.

[2] https://search.proquest.com/docview/1946167905?pq-origsite=gscholar.

[3] https://search.proquest.com/docview/1952703543/fulltextPDF/366C986A3D454A58PQ/1?accountid=130572.

[4]Thomas P. Scanlan and Joliet, llinois. (2016). A Correlational Study of Teacher Ratings Established by the Charlotte Danielson Framework for Teachers and Student Achievement in Reading and Math.

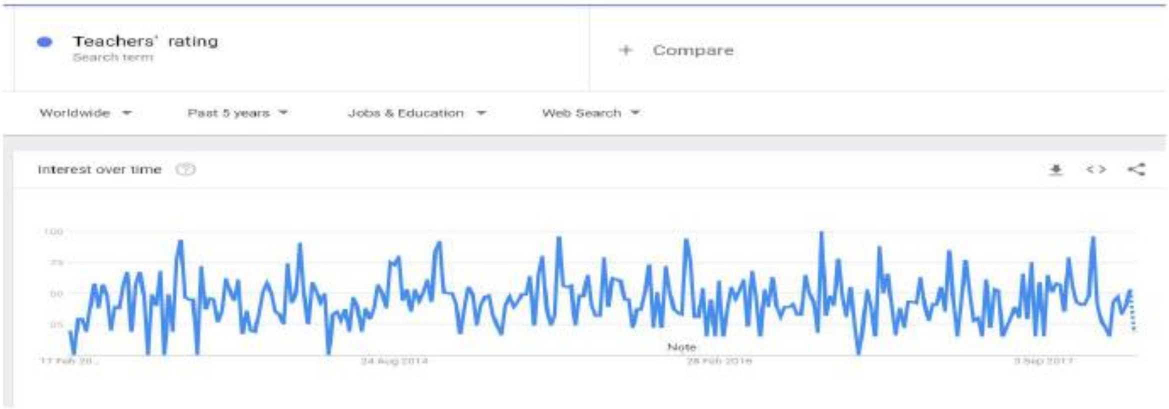


Figure 1. Google trends search: overall trends in 'teachers 'rating' over past 5- years.

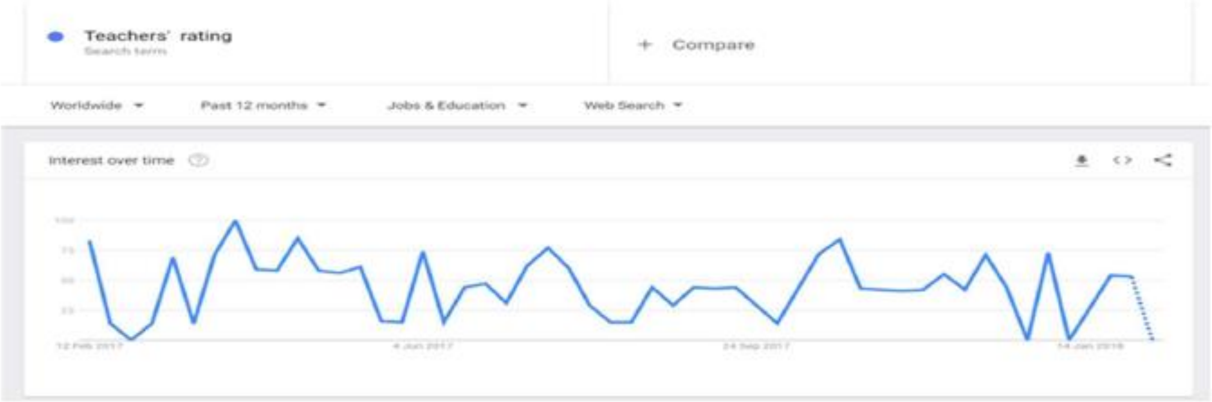


Figure. 2. Google trends search: overall trends in 'teachers 'rating' in 2017

Innovation

Teacher's evaluation rating and its interpretation is a very challenging and important educational topic and many researches have been done in this field but most of them are descriptive in nature. This research is a quantitative scientific research and the application of empirical rule and checking the strength of the data by normality test are innovations for this research, Also a big sample size of 558 students spanned over the period of three years is used which increases it reliability and validity.

Research questions

This study designed to answer the following three research questions.

Q1. What type of correlation is there, between students achieved average grade in their final exams and average course evaluation's satisfaction rating in the GSTA 140?

Q2. How to check the consistency of student's grades and teacher ratings by using measures of dispersion and empirical rule?

Q3. Which predictor is better to measure the level of educational instruction for a particular teacher, students' achievements or teacher ratings?

[4] https://search.proquest.com/docview/1854892927?pq-origsite=gscholar

DOI: 10.21522/TIJAR.2014.05.01.Art004
ISSN: 2520-3088

Literature review

Thomas P. Scanlan and Joliet, Illinois write in their very recent research about the correlation of teacher rating and students' success in the exam that "[5]the purpose of faculty evaluation tools is to identify weaknesses and strengths, promote professional development and immediate or, some cases, terminate ineffective faculty members".

A brief history of teacher's evaluation

Robert J. Marzano, Tony Frontier, and David Livingston describe the history of teacher evaluation in their book named "Effective Supervision" that the early days of supervision and evaluation began in the 1700s and lasted until the mid-1800s. They characterized by a reliance on clergy to provide guidance to and supervision of teachers. As school systems became more complex, the need for more specialized guidance for teachers gave rise to the principal teacher as leader and a growing awareness of the importance of pedagogy. The era of scientific management, from the late 1800s until right before World War II, characterized by two competing views of education. One was the view that the purpose of education was the promotion of democratic ideals. The other was the view that schools function best when approached from the perspective of scientific management. Throughout this era, the scientific approach gained strength and acceptance. The period after World War II saw a swing away from the scientific approach to an emphasis on developing the teacher as an individual. This period also saw a proliferation of the responsibilities of the supervisor.

The next era, lasting from the late 1960s to the early 1970s, saw the phenomenon of clinical supervision—one of the most influential movements in supervision and evaluation. The Hunter model combined with clinical supervision to produce a widely used, but often time's prescriptive approach to supervision. Developmental/reflective models that were much less prescriptive followed this period. The RAND study provided a realistic look at the actual practice of supervision and evaluation in districts and schools and concluded that teachers preferred specific as opposed to general feedback.

The mid-1990s saw the introduction of the Danielson model to teacher supervision and evaluation. It was widely applied to K–12 education. Finally, the first decade of the 21st century witnessed heavy criticisms of current evaluation practices calling for major changes in tenure and compensation.

Value-added analysis and value-added assessment

[6]The most significant trend to link teacher evaluation with student achievement is to use the exam grade and value-added modeling (also known as value-added analysis and value-added assessment). This method of teacher evaluation measures the teacher's contribution in a given year by comparing the current exam grade of his students in the grade of students in previous school years and so on. In this manner, value-added modeling seeks to isolate the contribution, or value added, that each teacher provides in a given year, which can be compared to the performance measures of other teachers. VAMs (value added models) are considered to be fairer than simply comparing student achievement scores or gain scores without considering potentially confounding context variables like past performance (Newton, Darling Hammond, Hearten, & Thomas, 2010).

Research design

The study is scientific, quantitative in nature based on specific, narrow questions to obtain measurable and observable data on variables x and y.

This study involves statistical analyses; conducted on numerical continues data to make generalizations and predictions about the population. The statistical test used for the inferences is "Pearson Statistical correlation test", to find the strength of the association between two continuous variables x and y.

[5] https://files.eric.ed.gov/fulltext/ED532775.pdf

[6] https://faculty.smu.edu/millimet/classes/eco7321/papers/koedel%20et%20al%202015.pdf

"Empirical rule" (also known as the three- sigma rule or the 68-95- 99.7 rule) is applied to provide a quick estimate of the spread of data in a normal distribution with the help of measured mean and standard deviation.

Sampling technique

Cluster convenience-sampling techniques is used for the selection of the sample, of 22 elementary statistics (GSTA 140) course grades and their teaching ratings.

Data collection

Observational data collection method (surveys) is used to collect the teacher rating's evaluation by the students; this is the method where the researcher has no control on the change or modify environment to get the desired results.

For students' achievement, the software known as "Blackboard Learn" gathered records (grades).

Blackboard Learn (previously the Blackboard Learning Management System), is a virtual learning environment and course management system developed by Blackboard Inc. It is a Web-based server software, which features course management, customizable open architecture, and scalable design that allows integration with student information systems and authentication protocols.

Students were asked to respond the following questions in the formal surveys of teacher ratings at EFTA University.

Table 1. Survey questionnaire

SR.NO	Questions	Strongly Agree%	Agree%	Disagree%	Strongly Disagree %	Nonresponse
1	The instructor was prepared					
2	The instructor was available for help outside of class					
3	The instructor returned assignments/tests timely and with feedback					
4	The instructor was enthusiastic about the subject					
5	The instructor stimulated interest in the course subject					
6	The instructors explanation was clear					
7	The instructor treated students with respect					
8	I would like to take another course with the same instructor					
9	The course generally followed the syllabus					
10	The course materials were up to date and useful					

11	The course resources I needed were available when I needed them					
12	The course materials were made available on Blackboard					
13	The course was intellectually challenging					
14	The assessment tasks and the assessment criteria were made clear to me					
15	The class activities, assignments, laboratories helped me in understanding the course					
16	The effort required for this course was appropriate					
17	Overall, I was satisfied with the quality of this course					
18	The lab sessions were organized and clear instructions were given to me					
19	The lab activities were well integrated with the lectures.					

Data analysis

Students’ average grades considered as variable “x” and average teacher rating considered as “Y”.

There are three measures have been calculated in this research, First is the coefficient of linear correlation between x and y, secondly value of standard deviation for both x and y and third is mean values of both above-mentioned variables.

A Pearson's coefficient of correlation “r” is used to analysis the data. The primary purpose of linear correlation analysis is to measure the strength of a linear relationship between two variables.

Some scatter diagrams given below that demonstrate different relationships between input, or independent variables, x, and output, or dependent variables, y.

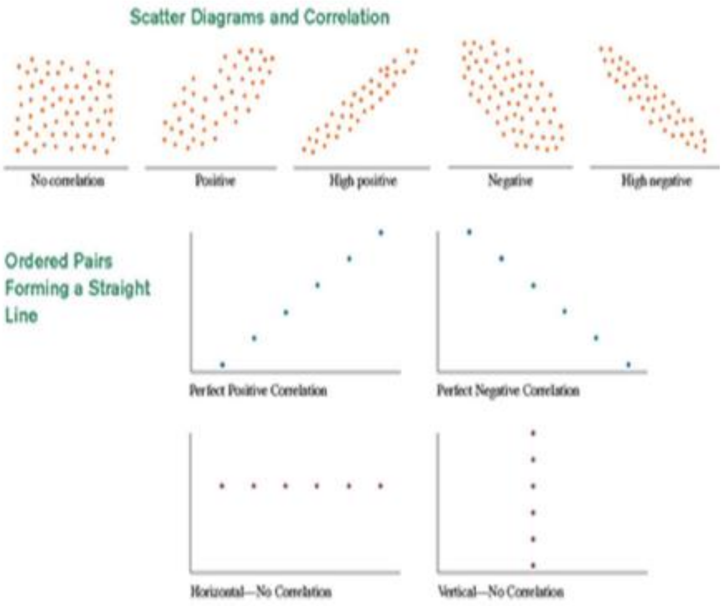

Scatter Diagrams and Correlation
No correlation
Positive
High positive
Negative
High negative
Ordered Pairs Forming a Straight Line
Perfect Positive Correlation
Perfect Negative Correlation
Horizontal—No Correlation
Vertical—No Correlation

Figure 3. Patterns of correlation (Robert johnson and patricia kuby (2012)).

If as x increases, there is no definite shift in the values of y, we say there is no correlation or no relationship between x and y. If as x increases, there is a shift in the values of y, then there is a correlation. The correlation is positive when y tends to increase and negative when y tends to decrease. If the ordered pairs (x, y) tend to follow a straight-line path, there is a linear correlation. The preciseness of the shift in y as x increases determines the strength of the linear correlation.

Perfect linear correlation occurs when all the points fall exactly along a straight line. The correlation can be either positive or negative, depending on whether y increases or decreases as x increases. If the data form a straight horizontal or vertical line, there is no correlation, because one variable has no effect on the other. Students' average grade in the final exam considered variable "X" and average satisfaction rating of the course evaluation surveys considered as the second variable "Y".

The following diagram shows the computing formulae to find the coefficient of linear correlation with the help of variables x and y.

7The standard deviation (SD, also represented by the Greek letter sigma σ (the Latin letter) or s, is a measure that is used to quantify the amount of variation or dispersion of a set of data values. A low standard deviation indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points spread out over a wider range of values.

The computing formula for standard deviation also uses the similar calculations with the help of variables x and y and the sample size n, as shown below.

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} \ and \ S = \sqrt{\frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n-1}}$$

[8]Empirical rule

If a variable is normally distributed, then (1) within 1 standard deviation of the mean, there will be approximately 68% of the data; (2) within 2 standard deviations of the mean, there will be approximately

[7] https://en.wikipedia.org/wiki/Standard_deviation#cite_note-StatNotes-1

[8] Robert Johnson and Patricia Kuby (2012)

95% of the data; and (3) within 3 standard deviations of the mean, there will be approximately 99.7% of the data.

[9] The Empirical Rule holds for normally distributed populations. In addition: The Empirical Rule also approximately holds for populations having single peaked; mound-shaped distributions that are not much skewed. In some situations, the skewness of a mound-shaped distribution can make it tricky to know whether to use the Empirical Rule. Hence, the class width is a decisional choice to have a realistic view of the distribution.

The empirical rule tells us that at least 95% of all sample means fall within about 2 standard deviations (SD) of the population mean, meaning that there is less than a 5% probability of obtaining a sample mean that is beyond 2 SD from the population mean.

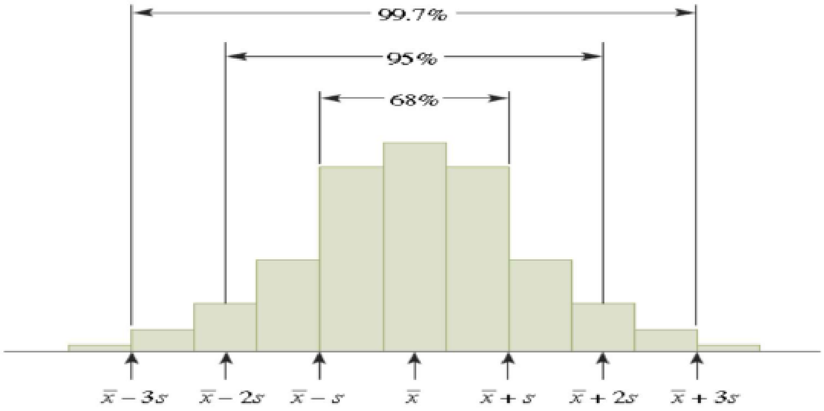


Figure 4. Empirical rule

Calculation

The following table is consisting five columns. The first column is showing the courses' sections and the times, when these courses taught, the second column is showing a number of students, third is showing grades out of 40. The teaching in Effat University is the type of continuous assessment system and weighted of the final exam is 40%, the fourth column is graded percentage and the last one is teacher rating out of 100.

Table. 2. Data collection

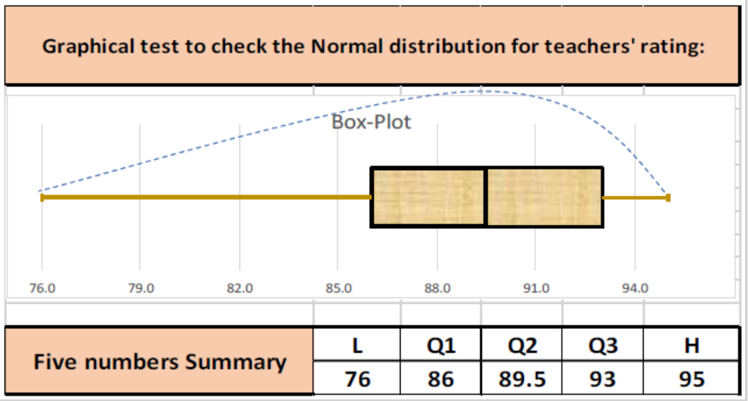
Courses	**Number of students**	**Average grade out of 40**	**Percentage of grades (X)**	**Percentage of Positive rating (Y)**
Fall 2017-GSTA 140 (1)	**26**	**27.85**	**69.625**	**89**
Fall 2017-GSTA 140 (2)	**26**	**29.41**	**73.525**	**95**
Fall 2017-GSTA 140 (6)	**20**	**27.88**	**69.7**	**94**
Spring 2017-GSTA 140 (2)	**33**	**29.02**	**72.55**	**93**
Spring 2017-GSTA 140 (4)	**33**	**28.79**	**71.975**	**92**
Spring 2017-GSTA 140 (5)	**29**	**26.9**	**67.25**	**95**

[9] http://www.academia.edu/18782753/Application_of_Empirical_Rule_on_Standard_Deviation_and_the_Chebyshev_s_Theorem_Quantitative_Aspects_of_Real_Estate_Market_Studies

Spring 2017-GSTA 140 (6)	**30**	**28.9**	**72.25**	**90**
Fall 2016-GSTA 140 (3)	**26**	**27.84**	**69.6**	**86**
Fall 2016-GSTA 140 (4)	**26**	**25.88**	**64.7**	**93**
Fall 2016-GSTA 140 (6)	**24**	**25.15**	**62.875**	**81**
Spring 2016-GSTA 140 (2)	**27**	**28.61**	**71.525**	**93**
Spring 2016-GSTA 140 (3)	**24**	**30.33**	**75.825**	**93**
Spring 2016-GSTA 140 (7)	**24**	**27.16**	**67.9**	**79**
Fall 2015-GSTA 140 (2)	**25**	**28.34**	**70.85**	**82**
Fall 2015-GSTA 140 (11)	**24**	**26.81**	**67.025**	**92**
Spring 2015-GSTA 140 (2)	**26**	**32.20**	**80.5**	**86**
Spring 2015-GSTA 140 (3)	**22**	**32.32**	**80.8**	**89**
Spring 2015-GSTA 140 (4)	**23**	**30.66**	**76.65**	**87**
Fall 2014-GSTA 140 (1)	**28**	**28.63**	**71.575**	**93**
Fall 2014-GSTA 140 (2)	**18**	**28.86**	**72.15**	**76**
Fall 2014-GSTA 140 (3)	**19**	**26.7**	**66.75**	**86**
Fall 2014-GSTA 140 (9)	**25**	**31.41**	**78.525**	**84**
Total Number of students	**558**			

Normality test

To compare the sample distribution with the normal distribution, a graphical method is used and box plots (a simple way of representing statistical data on a plot in which a rectangle drawn to represent the second and third quartiles, usually with a vertical line inside to indicate the median value. The lower and upper quartiles are shown as horizontal lines either side of the rectangle) are drawn for both the variable x and y.



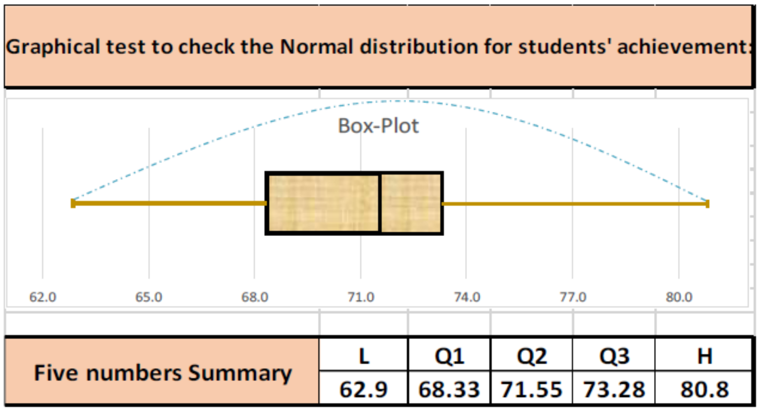
Graphical test to check the Normal distribution for teachers' rating:

Five numbers Summary	L	Q1	Q2	Q3	H
	76	86	89.5	93	95

Graph 1. Box plot for Y

Quartiles Q1, Q2, and Q3 are the values, which represent 25%, 50% and 75% of the data set.

Least and highest values in the data sets are representing by L and H.



Graphical test to check the Normal distribution for students' achievement:

Five numbers Summary	L	Q1	Q2	Q3	H
	62.9	68.33	71.55	73.28	80.8

Graph 2. Box plot for X

Students' achievements' graph is showing better normal distribution the teacher ratings.

The normal distribution is important in statistics and used in the natural and social sciences to represent real-valued random variables with unknown distributions.

The values for product x and y and squares of x and y are given below.

Table 3. Sum of variables, x, y, x^2, y^2 and xy

	Average grades % (X)	**Average Positive rating % (Y)**	**XY**	x^2	y^2
1.	69.63	89	6196.63	4847.64	7921
2.	73.53	95	6984.88	5405.93	9025
3.	69.70	94	6551.80	4858.09	8836
4.	72.55	93	6747.15	5263.50	8649
5.	71.98	92	6621.70	5180.40	8464
6.	67.25	95	6388.75	4522.56	9025
7.	72.25	90	6502.50	5220.06	8100
8.	69.60	86	5985.60	4844.16	7396
9.	64.70	93	6017.10	4186.09	8649
10.	62.88	81	5092.88	3953.27	6561
11.	71.53	93	6651.83	5115.83	8649
12.	75.83	93	7051.73	5749.43	8649
13.	67.90	79	5364.10	4610.41	6241
14.	70.85	82	5809.70	5019.72	6724
15.	67.03	92	6166.30	4492.35	8464
16.	80.50	86	6923.00	6480.25	7396
17	80.80	89	7191.20	6528.64	7921
18.	76.65	87	6668.55	5875.22	7569
19.	71.58	93	6656.48	5122.98	8649
20.	72.15	76	5483.40	5205.62	5776

21.	66.75	86	5740.50	4455.56	7396
22.	78.53	84	6596.10	6166.18	7056
	$\sum x = 1574.13$	$\sum y = 1948.00$	$\sum xy = 139391.85$	$\sum x^2 = 113103.89$	$\sum y^2 = 173116$

Sum other important calculations are given below.

Average of students' achievement

$$\bar{x} = \frac{\sum x}{n} = \frac{1574.13}{22} = 77.55\%$$

Average of teacher's rating

$$\bar{y} = \frac{\sum y}{n} = \frac{1948}{22} = 88.55\%$$

Correlation between students' achievement and teachers' rating:

$$ss(x) = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

$$=113103.89 - \frac{(1574.13)^2}{22} = 472.74$$

$$ss(y) = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$

$$=173116 - \frac{(1948)^2}{22} = 629.45$$

$$ss(xy) = \sum xy - \frac{\sum x \sum y}{n}$$

$$= 139391.85 - \frac{1574.13 \times 1948.00}{22}$$

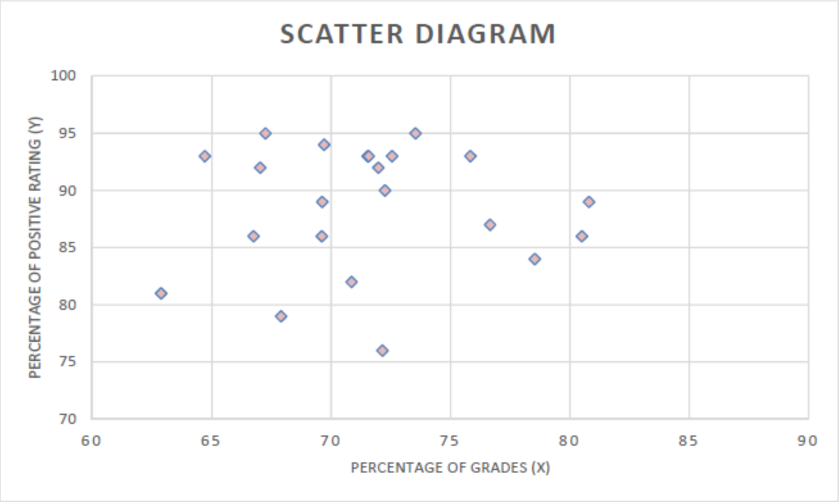
$$= 9.75$$

$$r = \frac{ss(xy)}{\sqrt{ss(x)ss(y)}}$$

$$= \frac{9.75}{\sqrt{472.74 \times 629.45}}$$

$$= \frac{9.75}{545.5}$$

$$r = 0.02$$



Graph 3. Scatter diagram

DOI: 10.21522/TIJAR.2014.05.01.Art004
ISSN: 2520-3088

The standard deviation of students' achievement

$$s = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}}$$

$$s = \sqrt{\frac{113103.89 - \frac{(1574.13)^2}{22}}{21}}$$

S=$\sqrt{22.51}$ =4.7≅ 5

The standard deviation of teachers' rating

$$s = \sqrt{\frac{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}{n-1}}$$

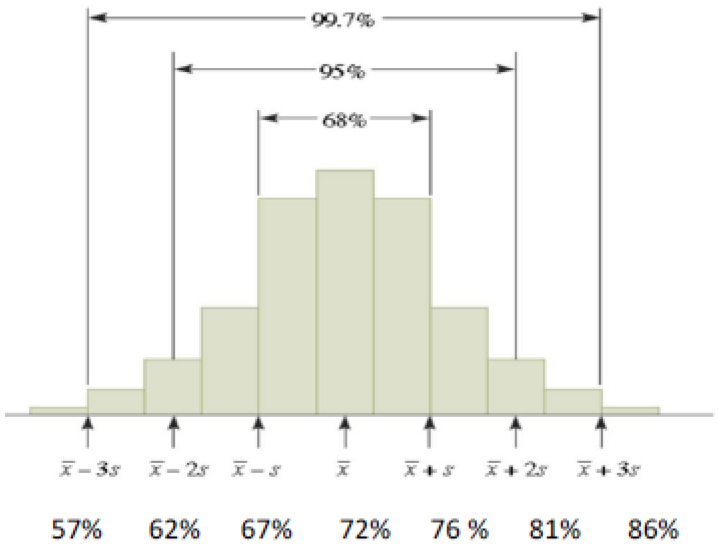
$$s = \sqrt{\frac{173116 - \frac{(1948)^2}{22}}{21}}$$

S=$\sqrt{29.97}$ =5.4≅ 5

Empirical Rule for students' achievement

Table 4. Three sigma values

$\bar{x} - s$	**67**
$\bar{x} - 2s$	**62**
$\bar{x} - 3s$	**57**
$\bar{x}$	**72**
$\bar{x} + s$	**76**
$\bar{x} + 2s$	**81**
$\bar{x} + 3s$	**86**

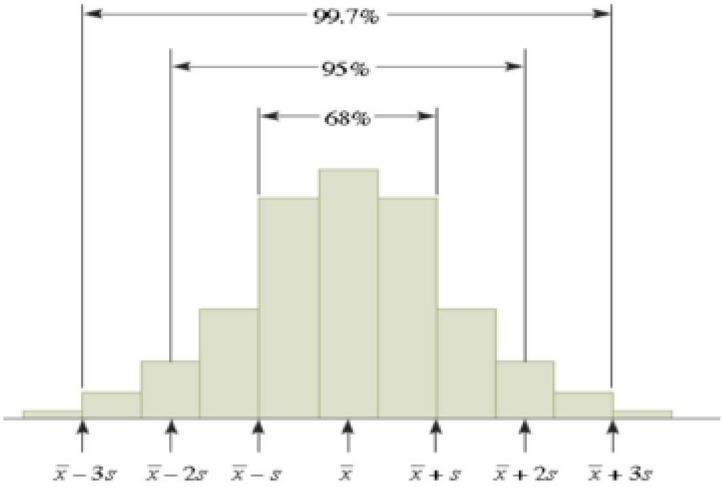


Graph 4. Empirical rule

Empirical rule for teachers' rating

Table 5. Three sigma values

$\bar{x}-s$	**83**
$\bar{x}-2s$	**78**
$\bar{x}-3s$	**72**
$\bar{x}$	**89**
$\bar{x}+s$	**94**
$\bar{x}+2s$	**99**
$\bar{x}+3s$	**105**



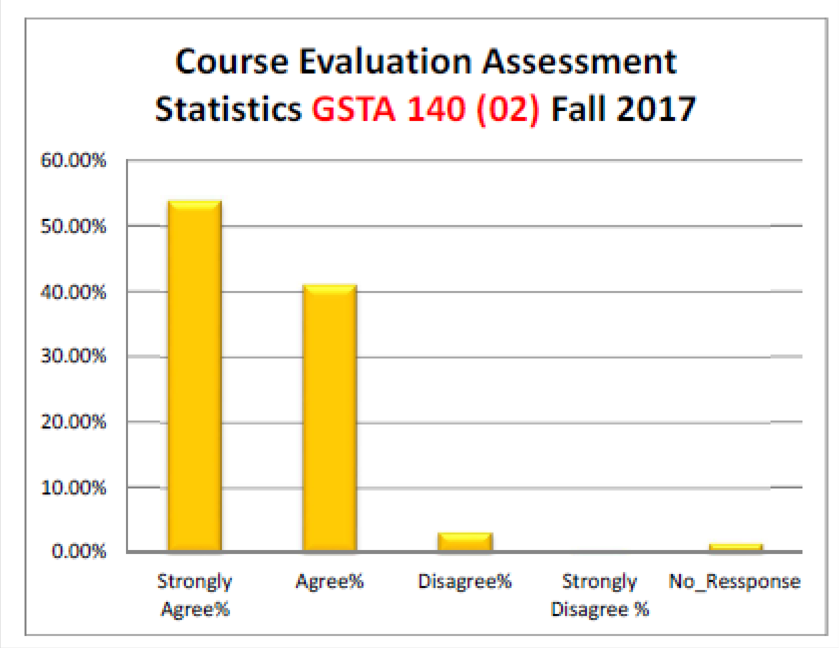
Graph 5. Empirical rule

Sample of formal teacher rating survey used in Effat University and its graphical interpretation

Table 6. Sample-teacher's evaluation survey

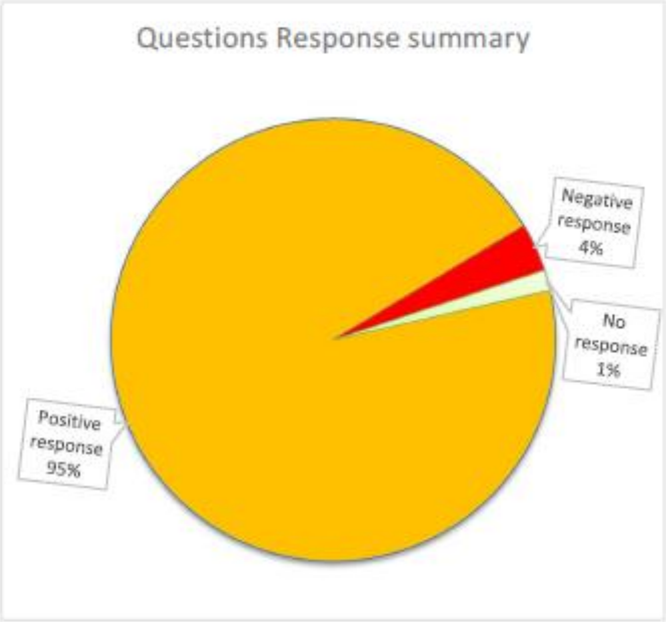
SR.NO	Questions	Strongly Agree%	Agree%	Disagree%	Strongly Disagree %	Nonresponse
1	The instructor was prepared	13	8	0	0	0
2	The instructor was available for help outside of class	12	7	2	0	0
3	The instructor returned assignments/tests timely and with feedback	12	8	1	0	0
4	The instructor was enthusiastic about the subject	12	8	1	0	0
5	The instructor stimulated interest in the course subject	11	9	1	0	0
6	The instructors explanation was clear	11	8	2	0	0
7	The instructor treated students with respect	14	7	0	0	0
8	I would like to take another course with the same instructor	11	8	1	1	0
1	The course generally followed the syllabus	11	10	0	0	0
2	The course materials were up to date and useful	12	9	0	0	0
3	The course resources I needed were available when I needed them	12	9	0	0	0
4	The course materials were made available on Blackboard	12	8	1	0	0
5	The course was intellectually challenging	9	11	1	0	0
6	The assessment tasks and the assessment criteria were made clear to me	12	8	1	0	0

7	The class activities, assignments, laboratories helped me in understanding the course	12	9	0	0	0
8	The effort required for this course was appropriate	11	10	0	0	0
9	Overall, I was satisfied with the quality of this course	12	9	0	0	0
10	The lab sessions were organized and clear instructions were given to me	8	9	1	0	3
11	The lab activities were well integrated with the lectures.	8	9	1	0	3
12	**Average**	53.88 %	41.10 %	3.26 %	0.25 %	1.50 %
13						
14	Student response	SR	95 %	Total Students		27
15	Response Rate	RR	78 %	N0. of Response		21



Graph 6. Sample-course evaluation assessment

DOI: 10.21522/TIJAR.2014.05.01.Art004
ISSN: 2520-3088



Graph 7. Sample-question response summary

Findings of the study

Q1. What type of correlation is there, between students achieved average grade in their final exams and average course evaluation's satisfaction rating in the GSTA 140?

Answer: There is a very weak positive correlation between students' achieved average final exam grade and course evaluation's satisfaction rating in the GSTA 140 final exam, the value of the coefficient of correlation is 0.02.

Q2. How to check the consistency of student's grades and teacher ratings by using measures of dispersion and empirical rule?

Answer: The standard deviation is approximately equal to 5 for both variables, which shows the same consistency.

The average Final grade of 558 students in 22 courses is 77.55%.

The average teacher's evaluation rating done by students in 22 courses is 88.55%.

By comparing above results, we conclude the ranges of reliability of these two variables by using empirical rule in 95% confidence interval, as follows:

$$73\% \leq \text{x} \leq 82\% \text{ and } 84\% \leq \text{Y} \leq 94\%$$

Q3. Which predictor is better to measure the level of educational instruction for a particular teacher?

Answer: Normality tests showed that the data set for the students' grades is well modeled by a normal distribution as compared to the data set of teacher ratings and hence the overall reliability of students' grades is better than teacher ratings and hence students' grades are the best indicator of educational instructions.

Conclusion

Teacher rating is not a true reflector of student outcomes as there is a very weak correlation between them. Less grades of the student does not mean that the teacher is not good. There are many factors effect on the grades; especially it is hard for the weak base students to get good grades. Similarly, teacher rating also badly effect if the course is a hard level or teacher is strict in grading or in maintaining discipline, etc., but again it does not reflect that the teacher is not qualified or not delivering the lessons properly.

The Empirical rule is the best way to find the real interpretation and reliability of these two predictors and their extreme limits.

Limitations

Apart from a good sample size of 22 similar courses taught by the same instructor, there are some limitations too, which can affect the findings of the research. For example,

- Difficulty level for different courses may affect the teacher' evaluation survey ratings.
- Instructors' teaching style can affect the consistency of the results.
- The sample distribution is not 100% normal distribution, which may affect the empirical rule's finding.
- Experience of the instructor also an important factor for the consistency of results.
- Feedback is gathered only by formal evaluation surveys.
- The participation rate of the students is less than 100% in all surveys.

Recommendations

Good teachers get large gains in student achievement (Hanushek, 2002) and there is strong agreement that good teaching may be the single most factor in student achievement (Haycock, 1998).

An effective teacher evaluation system can help all teachers improve as well as indicate areas where a teacher may need to grow (Wise et al., 1984). The most common method is for an evaluator to conduct frequent classroom visits (Peterson, 2009).

Teacher survey data showed that the majority of teachers used the evaluation information to improve teaching. The survey data also showed that the evaluation process led to improvements in student learning.

The most common and effective method of teacher evaluation is for the administrator to conduct classroom visits (Peterson, 2009).

There are some suggestions to improve the process of teachers' evaluation.

Administrators should not be considered the formal teacher's rating survey alone for judging the teaching skill for teachers. There are a number of approaches that can be used for this purpose, e.g., classroom observations (by managers, other teachers or external evaluators), value-added models that try to measure gains in student achievement, judgements made by the teacher's line manager or principal, teacher self-evaluation, teacher portfolios of work etc.

Results of surveys, should be discussed with the teacher and help her/him to overcome the weak areas by providing different types of professional development activities and workshops.

Value-added models or continuous assessments should be considered to appreciate/ awards /reward for the teacher.

Students should be given very clear and specific guidance about the feedback requested.

An environment of trust and support should be maintained for both students and teachers by the management.

Teacher's evaluation process or model should be followed.

DOI: 10.21522/TIJAR.2014.05.01.Art004
ISSN: 2520-3088

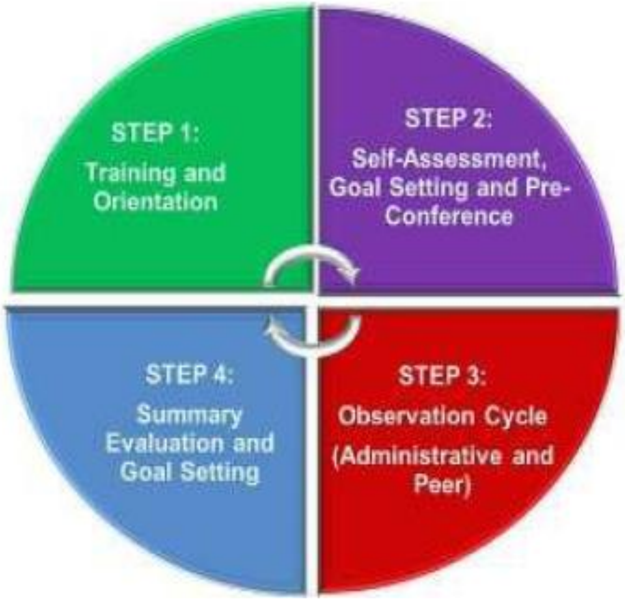

STEP 1:
Training and Orientation
STEP 2:
Self-Assessment, Goal Setting and Pre-Conference
STEP 3:
Observation Cycle (Administrative and Peer)
STEP 4:
Summary Evaluation and Goal Setting

Figure 4. Teacher evaluation process

References

[1]. Abrami, P. C., & Mizener, D. A. (1983). Does the attitude similarity of college professors and their students produce "bias" in course evaluations?. American Educational Research Journal, 20 (1), 123-136.

[2]. Alexander, K., & Morgan, S. L. (2016). The Coleman Report at Fifty: Its Legacy and Implications for Future Research on Equality of Opportunity. The Russell sage foundation journal of the social sciences. Retrieved from http://www.rsfjournal.org/doi/pdfplus/10.7758/RSF.2016.2.5.01

[3]. Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. Journal of Personality and Social Psychology, 64, 431–441.

[4]. Basow, S. A. (1998). Student evaluations: Gender bias and teaching styles. In L. H. Collins, J. C. Chrisler, & K. Quina (Eds.), Career strategies for women in academe: Arming Athena (pp. 135-156). Thousand Oaks, CA: SAGE.

[5]. Baslow, S. A. (2000). Best and worst professors: Gender patterns in students' choices. Sex Roles, 45 (5/6), 407-417.

[6]. Bill & Melinda Gates Foundation. (2013). ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study. Retrieved from http://files.eric.ed.gov/fulltext/ED540958.pdf

[7]. Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. Journal of educational psychology, 75 (5), 631.

[8]. Butin, D. W. (2010). The education dissertation: A guide for practitioner scholars. Corwin Press.

[9]. Chaskin, R. J., & Rauner, D. M. (1995). Youth and caring: An introduction. Phi Delta Kappan, 76 (9), 667.

[10]. Coleman, L. M., Jussim, L., & Isaac, J. L. (1991). Black students' reactions to feedback conveyed by white and black teachers. Journal of Applied Social Psychology, 21 (6), 460-481.

[11]. Cohen, D. K., & Ball, D. L. (2001). Making change: Instruction and its improvement.

[12]. Calkins, S., & Micari, M. (2010). Less-than-perfect judges: Evaluating student evaluations. Thought & Action: The NEA Higher Education Journal, Fall 2010, 7-22.

[13]. Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? Journal of Higher Education, 71 (1), 17-33.

[14]. Ingrid Yvonne Williams Medlock. (December 2017). Teacher evaluation ratings and student achievement: what's the connection? https://search.proquest.com/docview/1946167905?pq-origsite=gscholar

[15]. Kardia, D. B., & Wright, M. C. (2004). Instructor identity: The impact of gender and race on faculty experiences with teaching. CRLT Occasional Paper No. 19. Ann Arbor, MI: Center for Research on Learning and Teaching, University of Michigan.

[16]. Kogan, L. R., Schoenfeld-Tacher, R., & Hellyer, P.W. (2010). Student evaluations of teaching: perceptions of faculty based on gender, position, and rank. Teaching in Higher Education, 15 (6), 623-636.

[17]. LaVaque-Manty, M., & Cottrell, D. (2015, October). Course evaluations at Michigan: What do we know? Ann Arbor, MI: University of Michigan Learning Analytics Task Force.

[18]. Neelie B. Parker. (May 2017). Framework for teacher evaluation: examining the relationship between teacher performance and student achievement.

[19]. https://search.proquest.com/docview/1952703543/fulltextPDF/366C986A3D454A58PQ/1?accountid=130572

[20]. Nilson, L. B. (2012). Time to raise questions about student ratings. In J. E. Groccia & L. Cruz (Eds.), To improve the academy:

[21]. Resources for faculty, instructional, and organizational development, Vol. 31 (pp. 213-228). San Francisco, CA: Jossey-Bass.

[22]. Phi Delta Kappan, 83 (1), 73-77. doi:10.1177/003172170108300115.

[23]. Sprague, J., & Massoni, K. (2005). Student evaluations and gendered expectations: What we can’t count can hurt us. Sex Roles: A Journal of Research, 53 (11-12), 779-793.

[24]. Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. Review of Educational Research, 83 (4), 598–642.

[25]. Thomas P. Scanlan and Joliet, Illinois. (2016). A Correlational Study of Teacher Ratings Established by the Charlotte Danielson Framework for Teachers and Student Achievement in Reading and Math.

[26]. https://search.proquest.com/docview/1854892927?pq-origsite=gscholar