

STUDENT EVALUATION OF UNIVERSITY TEACHING QUALITY: ANALYSIS OF A TEACHER'S RATING SCALE FOR A SAMPLE OF UNIVERSITY STUDENTS[#]

Mah Nazir Riaz
Department of Psychology
University of Peshawar
Peshawar, Pakistan

Students in the classroom observe and experience different types of teachers' behaviours including their interest and mastery of the subject, ability to stimulate intellectual curiosity, fairness in grading, student teacher communication, and teachers' attitude toward the students. Yet in Pakistan their evaluations are seldom sought in any systematic way for assessment of teaching quality. The present study was designed to devise an instrument named as Peshawar University Teachers' Rating Scale (PUTRS) primarily for use by teachers themselves for voluntary self-assessment aimed at self-improvement and also to evaluate students' perceptions of university teaching quality. It is a Likert-type scale consisting of 25 items. A total of 2038 post-graduate students of Peshawar University were included from 19 teaching departments. The number of teachers who were evaluated by this sample was 144. Construct validity was determined by factor-analysis and item-total techniques. Factor analysis revealed PUTRS as a uni-dimensional test. Internal consistency estimates of reliability (coefficient alpha) for the total scale is .95. The results demonstrated that PUTRS is a reliable and valid test to be used by teachers who want feedback that would pinpoint their specific strengths and weaknesses from students' point of view.

Evaluation is as much a part of education as is learning. In all educational institutions, from elementary school upto university, students' academic achievement is regularly evaluated by their teachers who communicate their appraisals in the form of marks, division, or grades. Students' reaction to these academic evaluations depends upon their attributions, i.e., students' inferences about the

[#] This project was completed through a grant by University of Peshawar, which is gratefully acknowledged.

Correspondence concerning this article should be addressed to Mah Nazir Riaz, National Institute of Psychology, Quaid-i-Azam University (New Campus) Islamabad, Pakistan.

causes of their performances and evaluations. Empirical studies show that Heider's (1958) classic foursome, namely, ability, effort, luck, and task difficulty are among the most frequently offered explanations of performance (Bar-Tal, Ravgad, & Zilberman, 1981; Elig & Frieze, 1979; Falbo & Beck, 1979), additional factors are also sometimes described as causes. For example, Forsyth and McMillan (1982) found that students attribute their examination results to good/faulty teaching; adequate/inadequate; preparation for test; high/low motivation; personal problems; knowledge; good/bad study habits; luck; support from friends; classroom atmosphere; good/bad textbook. These findings clearly demonstrate that majority of casual factors have been attributed to teacher characteristics. In other words, the students attribute their success or failure as well as academic excellence to teaching quality.

Although a universally agreed upon definition of teaching quality has not yet been attained, the concern for its formulation is strongly felt by educationists and policy makers. Within this context opinions of university students are being recognized as most important in determination of teaching excellence (Abrami, Apollonia, & Cohen, 1990; Marsh, 1987; Perry, 1990). The only group of individuals who routinely experiences all facts of a teacher's classroom behaviors are the students themselves. Numerous studies demonstrate that students can identify the most significant dimensions of effective and efficient teaching (Angulo, Fernandez, & Martinez, 1987; Feldman, 1976; Marsh, 1984).

The main objective of the present investigation was to develop a standardized test for the assessment of teaching quality. The items devised for this test aimed at assessment of variables relevant to good teaching by various researchers.

METHOD

Sample

The total number of students who participated in the present evaluation of university teachers was 2038 (men and women). These students were enrolled in M.A. previous and final year classes in 19 post-graduate departments of humanities and sciences faculties of Peshawar University. All students who were present in the departments on the day of test administration were included in the sample. The number of teachers who were evaluated was 144, belonging to the same 19 teaching departments of Peshawar University, selected at random for the present study.

Instrument

Development of Peshawar University Teachers Rating Scale (PUTRS)

The initial step taken in the development of the PUTRS involved the collection of short statements of specific behavior exhibited by the teachers in the classroom. Sources included experts in psychometrics, educationists, researchers, the instruction evaluation literature, and available evaluation scales. The Purdue Teacher Evaluation Scale (PTES) developed by Bentley and Starry (1970), and Complutense University Teachers Evaluation Questionnaire (CUTEQ) developed by (Angulo, Fernandez & Martinez, 1987) served as particularly useful sources in the development of PUTRS. Consequently a Likert-Type 5-point rating scale (always, usually, sometimes, rarely, never) comprising 25 items was devised. The test was named as Peshawar University Teacher Rating Scale (PUTRS). The scale in its final form includes Directions and Practice Exercises, printed on the first page of the test. Twenty five affirmative statements comprising PUTRS refer to teacher's mastery of the subject, ability to stimulate intellectual curiosity; fairness in grading; student-teacher communication; and teacher's attitude towards the students. There was a consensus between a representative sample of teachers and students of Peshawar university about the variables identified as basic determinants of teaching quality.

PUTRS was pilot tested on a sample of 100 student teachers (50 men and 50 women) enrolled in Institute of Education and Research, University of Peshawar, during the academic year 1996-97. Item analysis of the data was carried out. Qualitative analysis aimed at determining the content validity of the test items necessitated some changes. The experts in test construction and education/teaching unanimously approved 16 items; modified 5 items, and discarded 4 items, and added 4 new items about which there was 100% agreement among the judges. Finally a list of 25 items was approved for the new test under consideration by the team of experts.

Procedure

PURTS is a self-administering test. It was administered to 2038 post-graduate students of social sciences studying in 19 different departments of Peshawar University during November-December 1998. The administration of the PURTS was carried out by previously trained research assistants during class time. The students of each class, in each department, were asked to evaluate all of the teachers with whom they had studied that year and who offered lecture – format classes. The

teachers had been informed beforehand about the nature of the study explaining the main objectives of the study, and they were convinced that the Teacher Rating Scale has been devised for use by teachers themselves for voluntary improvement.

After distribution of PUTRS to students in a class, the following directions (printed on test) were read out by the examiner.

“Following is a list of qualities that, taken together, tend to determine the quality of teaching. Of course, no one is ideal in all of these qualities, but to provide information which may lead to the improvement of teaching quality, you are asked to rate each one of your teachers on the indicated qualities, by selecting one of the boxes at the right of each statement, most appropriate to your opinion about your professor.”

The students were asked to write the name of the department and the teacher whom they were evaluating in the blank spaces provided for this purpose. The students were asked to answer the two practice items before starting the test. It was followed by the following instructions:

“This rating is to be entirely impersonal. Give your honest opinion. Do not disclose your identity. The information will be kept confidential and will be used for academic purposes.”

Afterwards the students were asked to record their opinion about each one of their teachers on separate test forms.

Scoring

Each item score ranges between 5-1 with the following response categories: Always-5, Usually-4, Sometimes-3, Rarely-2, and Never-1.

Low scores on the test show poor quality of teaching whereas high scores demonstrate high quality of teaching. The minimum score on PURTS is 25 whereas the maximum possible score is 125.

Results

The following analyses were carried out to determine various psychometric properties of PURTS.

- i. Inter-item correlation,
- ii. Item-total correlation,

- iii. Factor analysis,
- iv. Analysis of variance, and
- v. Norms/percentile ranks.

Reliability

The most meaningful type of reliability estimate for a rating scale of this type is coefficient alpha, which represents the most widely used and most general form of internal consistency estimate (Murphy & Davidshofer, 1988). The alpha coefficient obtained for PURTS is .95 which clearly demonstrates that the test is reliable. These findings were further supported by highly significant ($p < .001$) intercorrelation coefficients among all the 25 items of the test (Table 1).

Validity

The following two procedures were used to determine the construct validity of PURTS.

Item – total Correlation

Item-total correlation indices indicate that each item measures the same construct that is being measured by the test. Our data presented in Table 1 show that each one of the 25 items comprising PURTS bear a statistically significant positive correlation with the total test score ($p < .001$).

Factor Analysis

To determine the construct validity of PURTS, factor analysis was carried out. A principal component analysis of the data was carried out. Examining the factor loading (Table 2) four factors were identified as: (a) teaching competence demonstrated, (b) motivational skills, (c) fairness in grading, and (d) teacher's attitude towards the students.

Table 2

Principle Component Analysis: Summary Data

Factors	Eigen Value	% of Var	Cum %
1	11.64	46.6	46.6
2	1.51	6.0	52.6
3	1.15	4.6	57.2
4	1.03	4.1	61.3

As evident from Table 2, Factor I has an Eigen value of 11.64 which explains 46.6 percent of the total variance. Other three extracted factors have minimum acceptable Eigen values and explain very little amount of variance.

Table 3

Factor Analysis (PUTRS): Factor Matrix

	Factor 1	Factor 2	Factor 3	Factor 4
Q1	.65301	-.36784	-.14837	.29398
Q10	.55593	.11682	-.02658	.33198
Q11	.67037	.33496	-.13964	-.27818
Q12	.67477	.12762	-.22542	.02376
Q13	.68145	-.14263	.02128	.34343
Q14	.69579	.05032	-.32228	.01267
Q15	.68161	.13797	.22213	.18170
Q16	.72528	-.43226	-.01007	.03177
Q17	.76128	-.20052	-.02617	-.05263
Q18	.60035	-.33679	.09582	-.33738
Q19	.76533	.06693	-.26101	-.02455
Q2	.46218	.38379	.24611	.29472
Q20	.68052	.36408	-.14662	-.30064
Q21	.69409	.13618	.26758	-.12128
Q22	.71698	-.25563	-.03457	.08621
Q23	.60831	.17689	-.037560	.09693
Q24	.70929	.33697	-.16804	-.11060
Q25	.81292	.04266	-.18455	-.04835
Q3	.67277	.22416	.39803	-.07724
Q4	.52911	-.32684	.19388	-.48213
Q5	.75347	-.29472	-.06921	.04622
Q6	.68042	.20716	.26147	-.01502
Q7	.58759	-.02592	.36015	.09521
Q8	.77913	.01803	.26443	.09344
Q9	.78262	-.20877	-.00743	-.03981

Moreover, all items have high correlations with Factor 1 (Table 3). These findings demonstrate construct validity of Peshawar University Teachers' Rating Scale, and as such it can safely be regarded as a one-factor test for assessment of teaching quality at university level. PURTS is a reliable and valid measure of teaching competence, motivational skills, fairness in grading, and teacher's attitude towards the students. All these variables taken together account for the teaching quality at university level.

Analysis of Variance

To investigate the possible variations among various departments *viz a viz* the quality of teaching, an analysis of variance was carried out. The results reveal that the quality of teaching in various departments is significantly different.

Table 4

One-way ANOVA: Summary Data Showing Department Effects on Quality of Teaching

Source of Variance	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Main Effects Department	94865.380	18	5270.299	13.503	.000
Residual	788015.831	2019	390.300		
Total	882881.211	2037	433.422		

Students' evaluation shows that Department of Journalism stands at the top, Departments of Physics and Electronics are ranked second, whereas Arabic comes next. On the other hand, Department of Fine Arts stands at the lowest rank, and Law comes at the second lowest rank.

Although the data concerning each individual teacher's evaluation is available, but the researcher does not consider it important to present these ratings in the present report. Suffice is to say that every teacher who has opted teaching as a profession and is determined to perform his/her duties in the interest of students, university, and the nation, can use the PUTRS for self-assessment and improvement, by administering it at the end of each semester/academic session.

Norms

The most common form of norms are percentile ranks, which represent the simplest method of presenting test data for comparative

purposes. Table 5 shows norm data as they are typically presented in a percentile rank format.

Table 5

Percentile norms: Peshawar University Teacher Rating Scale (PUTRS)

Percentiles	Total Score on PUTRS
95	124.00
90	122.00
95	120.00
80	118.00
75	116.00
70	113.15
65	112.00
60	110.00
55	107.00
50	105.00
45	102.00
40	99.00
35	96.00
30	92.00
25	86.00
20	82.00
15	75.00
10	66.00
5	56.95

The Table 5 can be used by the teachers to compare their item/scale ratings with those of the norm group. The teachers can prepare their own Profile Sheets to get maximum information on how their students perceive the specific behavioral characteristics they exhibit in the classroom.

CONCLUSION

Peshawar University Teachers' Rating Scale is a uni-dimensional test that can be used by the teachers themselves in order to identify specific strengths and weaknesses of their teaching quality. Reliability and validity estimates of PUTRS demonstrate that this instrument will

provide valuable information to the teacher concerned for the improvement of teaching quality.

Applied Significance

PUTRS, if used regularly by every teacher in the Teaching Departments of University, and affiliated colleges, will be instrumental in improving quality of teaching.

REFERENCE

- Abrami, P. C., D' Appllonia, S., & Cohen, P. A. (1990). Validity of student rating of instruction: What we know and what we do not. *Journal of Educational Psychology*, 82, 219-231.
- Angulo, F., Fernandez, J., & Martinez, M. R. (1987). *La evaluation' de la ensenanza universitari*. Madrid: Instituto de Ciencias de la Educacion.
- Bar-Tal, D., Ravgad, N., & Zilberman, D. (1981). Development of causal perception of success and failure. *Educational Psychology*, 1, 231-240.
- Bentley, R. R., & Starry, A. R. (1970). *Manual for the PURDUE teacher evaluation scale*. Indiana: PURDUE Research Foundation.
- Falbo, T., & Beck, R. C. (1979). Native psychology and the attributional model of achievement. *Journal of Personality*, 47, 185-195.
- Elig, T. W., & Frieze, I. H. (1979). Measuring causal attributions for success and failure. *Journal of Personality and Social Psychology*, 37, 621-634.
- Feldman, K. A. (1976). The superior college teacher from the student's view. *Research in Higher Education*, 5, 243-288.
- Forsyth, D. R., & McMillan, J. H. (1982). Reactions to educational outcomes: Some affective and attributional correlates. Paper presented at the annual meeting of American Psychological Association, Washington, DC.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

- Marsh, H. W. (1984). Validity of students' evaluations of college teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.
- Marsh, H. W. (1987). Students' evaluations of university: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.
- Murphy, K. R., & Davidshofer, C. O. (1988). *Psychological testing*. New Jersey: Englewood Cliffs.
- Perry, R. P. (1990). Introduction to the special section. *Journal of Educational Psychology*, 82, 183-188.

Received: July 02, 1999.