

DEVELOPMENT AND VALIDATION OF TEACHERS' JOB PERFORMANCE SCALE[#]

Rubina Hanif & Seema Pervez

National Institute of Psychology

Centre of Excellence, Quaid-i-Azam University

Islamabad, Pakistan

The present article describes the process of development of an indigenous scale to measure the job performance of school teachers. The development and validation of this scale was carried out in different steps by using independent samples. The Teachers' Job Performance Scale (TJPS) comprises 25 items with 5-point rating scale. The teachers' performance was evaluated by their students. A sample of 180 students of secondary schools was taken to evaluate their 60 teachers. Each teacher was rated by three students. The factor analysis of items revealed four factors i.e., teaching skills, management skills, discipline and regularity, and interpersonal skills as the determinants of teachers' job performance. The psychometric properties of TJPS revealed that it has alpha coefficient of .94 for the entire scale and has sufficient internal consistency, and inter-rater reliability. The empirical evaluation also shows sufficient validity of scale.

The quality of an educational process and its product is unquestionably influenced by teachers' job performance. The entire edifice of education is shaky if the performance of teacher is weak and ineffective. Therefore, effective job performance of teachers is must for educational improvement, which we are striving hard to bring about. The definition of what constitutes best performance of teachers, is of course much more complicated than a simplistic listing of goals. It is much easier to list the rules of game than to coach someone to excel in performance. The mere creation and ratification of standards will never define good teaching at any level. There are many factors, which contribute to a teacher's performance i.e., effective teaching, time management, punctuality and regularity, effective communication with students etc.

There are two types of models to define job performance. First, there are several efforts outlining general models of job performance and the determinants of job performance. Campbell, McCloy, and

[#] Correspondence concerning this article should be addressed to Rubina Hanif, National Institute of Psychology, Centre of Excellence, Quaid-i-Azam University, Islamabad, Pakistan. <rubinahanif@hotmail.com>

Oppler (1993) proposed the view of job performance as multidimensional in nature, and comprised of eight factor latent structure (e.g., declarative knowledge, skill, and motivation). Waldman and Spangler (1989) developed a model of job performance focusing on characteristics of the individual (e.g., experience, ability); outcomes (e.g., feedback, job security); and the immediate work environment.

The second category of the models defining job performance move toward more flexible definitions of work roles and jobs, they viewed jobs as dynamic and more interchangeable and are defined with less precision. The focus is on the personal competencies required to perform various work roles and jobs rather than a narrow review of specific tasks and duties inherent in fixed jobs and work roles (Casccio, 1995; Ilgen & Hollenbeck, 1991).

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 (Ravegad & Zilberman, as cited in Arvey, 1998). 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, classroom atmosphere, etc. These findings clearly demonstrate that majority of causal factors have been attributed to teachers' characteristics. In other words, the students attribute their success or failure as well as academic excellence to quality of teaching.

Although a universally agreed upon definition of teaching performance has not yet been attained, the concern for its formulation is strongly felt by educationists and policy makers. Within this context, opinions of students are being recognized as most important in determination of teaching excellence (Abrami, Apollonia, & Cohen, 1990; Marsh, 1987; Perry, 1990).

Performance evaluation is a delicate issue. Performance evaluation is mostly used to meet the basic needs of any organization, to improve the work force, to provide certain administrative functions, to identify the strengths and weaknesses of individual employees, and to develop and evaluate human resource systems (Cleveland, Murphy, & Williams, as cited in Arvey, 1998). There are five main aspects involved in developing a system for the evaluation of teachers. The first is the purpose of the evaluation; the second is the target category of teachers to be assessed; the third is the conception of teachers' work that is adopted; the fourth concerns the dimensions of teaching quality

about which judgments are to be made; and the fifth is the approach to establishing the validity of the assessments.

A variety of techniques and instruments are used to conduct performance evaluations. The usual evaluator is a teachers' supervisor. Colleagues, pupils, subordinates, and self evaluations can also be used. Some inaccuracy in teachers' evaluation is due to personal and interpersonal factors. Researches have shown that ratings can be influenced by the gender and the race, personality traits of the rater, and the ratee (see, for example, Robbins & DeNisi, 1994).

Swartz, White, Stuck and Patterson (1990) derived 28 teaching practices for rating the teachers' job performance. Ferris, Bergin, and Wayne (1988) measured teachers' job performance from a synthesis of the process-product research on teaching, on seven performance dimensions.

Recently there has been strong interest in the student evaluation of teaching (SET) literature. The numbers of researches have focused on validity concerns with SETs (see, for example, Greenwald, 1997; Marsh & Roche, 1999). Particularly in the last 15 years, the study of student's evaluation has been one of the most frequently emphasized areas in American educational research. Thousands of papers have been written and reviewed (Aleamoni; Braskamp; Centra; Cohen; Costin; Doyle; Feldman; Kulik & Mckeachie; Murray; Remmers & Wolf, as cited in Marsh, 1984). It has been concluded through researches that students are in a unique position to assess a variety of aspects concerning effective instructions. Scriven (1996) identified several sources of validity for student ratings of instruction that include the students' ratings of their own increased knowledge and comprehension, the perceived changes in motivation toward the subject taught, a career associated with the subject, and the further learning in that subject area. Students get the opportunity to observe teacher behavior relevant to competent teaching, such as punctuality. This is also identification of teaching style indicators, such as enthusiasm. May also be able to give the feedback about the information that is not relevant to competent teaching, but important to other students, such as textbook cost, attendance policy, or homework.

In some studies conducted in Pakistan, i.e., Jahangir (1988) evaluates teachers' performance on a rating scale pertaining to the four broad categories of teaching behavior, i.e., intellect, personality, teaching techniques, and interaction with students. Riaz (2000) determined four factors as the measure of teachers' performance (a) teaching competence demonstrated, (b) motivational skills (c) fairness

in grading, and (d) teachers' attitude towards the students. These studies measured the job performance of university teachers as perceived by their students. There was no scale to measure the job performance of school teachers. So the need was felt to develop an indigenous scale to measure the job performance of school teachers. The main objectives of the present study are to develop a scale to measure school teachers' job performance and to determine psychometric properties of this scale.

METHOD

Development of teachers' Job Performance Scale (TJPS) was carried out in different steps.

Step 1: Item Generation

Sample

A sample of 60 students of 9th and 10th classes, 40 teachers and 20 headmistresses (both men and women) was randomly selected from different schools of the Islamabad, Rawalpindi, and Chakwal. The sample of school was comprised of both government and private schools.

Procedure

The first step to develop TJPS was the generation of items. An open ended questionnaire was given to the sample of 60 students, 40 teachers and 20 headmistresses to find out their views about the dimensions of job performance of a school teacher. They were asked to mention all those behaviors and qualities, which they would consider important for a school teachers for the good job performance.

The responses of the sample were carefully analyzed, and were converted into statements. These statements were arranged in frequency distributions. The statements with highest frequency were selected for making an item pool. In this way, a pool of 50 items was generated. These statements were carefully examined and scrutinized by the researchers with the help of literature review.

Step 2: Item Evaluation

In the next step, the items generated in the form of statements were evaluated by judges. The purpose was to clearly categorize the items into different categories and to check the inter-ratter reliability. Initially, these 50 statements pertaining to different categories were

given to 5 judges and they were asked to derive the categories of teachers' job performance. Based on the opinion of the judges, six dimensions of teachers' job performance were derived as categories, i.e., teaching quality, teaching style, subject knowledge, management skills, discipline and regularity, and interpersonal skills. After deriving these categories, 50 statements were given to a sample of 20 educationists and they were asked to categorize each item into their respective category of job performance of school teachers.

The criterion for the selection of the items for different categories was 75% consensus among the sample. The coefficient of concordance was also computed for the ratings of the sample. The judges also evaluated the face validity of the items. According to judges' opinion some items that were repeating concepts, and some that were not clearly relevant to the categories of job performance were excluded. Only 27 out of total 50 statements could be clearly categorized under the six categories of teachers' job performance. These statements were written with a five point rating scale "never", "sometimes", "often", "mostly", and "always". This 27-item scale was developed in two versions. One for the ratings by students, headmistresses, or colleagues, and other for the self-ratings of teachers.

Step 3: Empirical Evaluation

Sample

A sample of 180 students from 9th and 10th classes of government and private schools of Islamabad, Rawalpindi and Chakwal were selected. From each class, three students were randomly selected.

Procedure

In this phase, the scale comprising 27 items with 5 point rating scale, was given to 180 students. From each class, three students were randomly selected, and they were asked to rate their own teacher's performance individually. In this way, total 60 teachers were rated by their students. They were rated on five point scale as "never", "sometimes", "often" "mostly" and "always". The scores assigned to this scale were ranging from 1 to 4. The mean score of all three students for each teacher was computed.

Determination of Reliability and Validity

The reliability and validity of TJPS was determined by following statistical analysis:

1. Factor analysis
2. Item total correlation
3. Inter-scale correlations
4. Cronbach's Alpha Coefficient
5. Split Half reliability
6. Inter-rater reliability

Factor analysis

To find out the empirical value of the six categories of TJPS, responses of the scale were put to principle component factor analysis. The rotated extraction method was used to extract the factors. Loadings equal to or greater than .35 were considered as significant. The results indicated that eigen values greater than 1.00 supported only four factor solution, and accounted for 76 % of the variance.

Table 1

Factor loadings for the items of TJPS Obtained from the Principal Component Factor Analysis (N = 180) (items = 27)

Items	F1	F2	F3	F4
	TS	MS	DR	IS
	Items (6)	Items (5)	Items (7)	Items (7)
*1	.20	.11	.23	.13
2	.58	.41	.23	.31
*3	.11	.24	.31	.19
4	.58	.30	.40	.33
5	.82	.35	.33	.11
6	.41	.31	.12	.28
7	.50	.20	.34	.29
8	.42	.02	.25	.32
9	.32	.65	.34	.23

Continued...

Items	F1	F2	F3	F4
	TS	MS	DR	IS
	Items (6)	Items (5)	Items (7)	Items (7)
10	.10	<u>.73</u>	.42	.25
11	.23	<u>.74</u>	.16	.20
12	.10	<u>.70</u>	.16	.29
13	.30	.11	.21	<u>.81</u>
14	.22	<u>.52</u>	.25	.33
15	.25	.25	<u>.60</u>	.25
16	.34	.24	<u>.64</u>	.24
17	.42	.12	.91	.12
18	.16	.27	.77	.27
19	.16	.12	.87	.12
20	.21	.22	.91	.22
21	.25	.21	<u>.77</u>	<u>.84</u>
22	.26	.25	<u>.87</u>	<u>.51</u>
23	.23	.25	.12	<u>.91</u>
24	.24	.10	.27	<u>.86</u>
25	.13	-.12	.15	<u>.74</u>
26	.27	.17	.22	<u>.71</u>
27	.12	.12	<u>.81</u>	<u>.81</u>

* Item loadings < .35

Table 1 shows the loadings of the selected items of TJPS on four factors. These factors were labeled as TS (Teaching Skills), MS (Management Skills), DR (Discipline and Regularity) and IS (Interpersonal Skills). The loadings were obtained when principal component factor analysis was run to determine the factor structure of the scale. The criterion for the selection of items was loading of .35 and above. The item Nos. 1 and 3 were not falling on the criteria and did not show clear picture of the dimensions, so these two items were excluded from the scale and only 25 items were retained for the final scale.

Table 2

Eigen Values and Percentage Variances Explained by the Extracted Factors for the TJPS

Factor	Eigen values	PCT of variance	Cum percentages
F1	6.62	26.50	26.50
F2	5.60	22.30	48.80
F3	4.50	17.90	66.62
F4	2.22	8.90	75.52

Table 2 demonstrates the eigen values and percentages of variance explained by the four factors. It shows that F1 has an eigen value of 6.62 and explain 26.50 % of the total of the variance that is the highest value among four factors. All other factors have eigen values above 2.22 and total variance explained by the four factor is 76 per cent.

Item total correlation

To determine the internal consistency of scale and examine their relevance with the test, item total correlation was calculated.

Table 3

Item total correlation of TJPS (N = 180)

Items	Correlations	Items	Correlations
1	.41**	14	.63**
2	.60**	15	.70**
3	.43**	16	.61**
4	.62**	17	.65**
5	.70**	18	.74**
6	.73**	19	.70**
7	.82**	20	.90**
8	.64**	21	.73**
9	.70**	22	.70**
10	.80**	23	.50**
11	.80**	24	.47**
12	.80**	25	.50**
13	.80**		

** $p < .01$

Table 3 shows that all the items are significantly correlated with total score of TJPS. The correlation coefficient ranges from .41 to .90 for all the 25 items of scale. It shows that all the items are consistent with the total scores of scale. It determines the reliability and construct validity of scale as well.

Inter-scale correlation coefficient

The internal consistency was further determined by inter-correlation of scores on sub scales as well as with that of the total score on TJPS. All the correlations are found significant.

Table 4

Inter correlations for scores on TJPS

Subscales	I	II	III	IV
I. Teaching skills				
II. Management skills	.48**			
III. Interpersonal skills	.60**	.45**		
IV. Discipline and regularity	.28**	.80**	.40**	
Total	.70**	.84**	.82**	.80**

** $p < .01$

Table 4 shows inter-correlation of scores on subscales as well as with total scores for TJPS. The data indicates that all the subscales of TJPS have significant correlation with each other and with the total scores. It shows the internal consistency of scale. The highest correlation is found between subscales of management skills and discipline, and regularity (.80), and minimum correlation is between teaching skills and discipline, and regularity (.28).

Cronbach's Alpha Coefficients

Initial psychometric analysis, using Cronbach's alpha coefficient yielded an internal consistency coefficient of .94 for the whole 25-items, and ranges from .80 to .92 for the subscales.

Table 5

Alpha Reliability Coefficient of total and subscales of TJPS (N=180)

	Subscales	No. of items	Alpha Coefficient
I.	Teaching Skills	6	.80
II.	Management Skills	5	.90
III.	Discipline and Regularity	7	.92
IV.	Interpersonal Skills	7	.91
	Total	25	.94

Split Half reliability

For calculating the split half reliability coefficient, the TJPS was divided into two parts with 13 items in the first part and 12 items in the second part. The correlation coefficient between two parts was found .87. Split half reliability was also calculated for the subscales of TJPS.

Table 6

Split half reliability coefficient for scores on total and subscales of TJPS (N=180)

	Subscales	Items	Correlation
I.	Teaching skills	6	.67
II.	Management skills	5	.89
III.	Discipline and regularity	7	.90
IV.	Interpersonal skills	7	.84
	Total	25	.87

Inter rater reliability

To determine the reliability of performance measures inter-rater reliability is widely used technique. In this method, ratings are judged on another independent sample's ratings. For TJPS inter-correlation of ratings of students, headmistresses and teachers' self ratings were calculated. A sample of 30 teachers, 30 headmistresses and 30 students were selected for this purpose. The ratings of all three groups were correlated.

Table 7

Inter-correlations of the Ratings by Students (n=30), Headmistresses (n=30), and Teachers' (n=30) Self ratings on TJPS

Raters	Students	Headmistresses	Self
Students	-	.89**	.21
Headmistresses	-	-	.62**
Self	-	-	-

** $p < .01$

Table 7 indicates that students' ratings for their teachers' job performance has significant correlation with the ratings of these teachers' headmistresses. The students' ratings have positive but non significantly correlated with self ratings of teachers. The teachers' self ratings have significant correlation with the ratings by their headmistresses.

Cut-off points

Cut-off points for the scale can be determined through the percentile analysis, its frequency distribution of the scores and the scores corresponding to these percentiles. It is more favourable to determine the norms of a scale through percentiles on a large sample. But for the present sample the purpose was to determine the cut-off points to get the categories of the performance levels. For this purpose the frequency distribution for total sample of students can be used to locate cutting points for different levels of job performance in teachers.

Table 8

Percentile Ranks and Scores on TJPS (N = 180)

Percentile	Scores	Percentile	Scores
1	38	55	90
5	50	60	91
10	62	65	93
15	68	70	94
20	81	75	95
25	85	80	96
30	88	85	96
35	88	90	97
40	88	95	107
45	88	99	108
50	89		

Table 8 indicates the percentile scores for Teachers Job Performance Scale (TJPS). In this case a score of 85 falls on the 25th percentile, whereas, a score of 89 falls on the 50th percentile and a score of 95 falls at 75th percentile. Therefore, the cut off scores for three levels of performance in teachers i.e., poor, good, and excellent was determined as scores below 86 as indicative of teacher's poor performance, above 95 as excellent performance, and scores ranging above 86 to below 95 as indicative of average performance of teachers as rated by students on Teachers Job Performance Scale (TJPS).

DISCUSSION

The recognition of students' ratings of their teachers, performance has short history but has enormous vitality and strength (McKeachie, 1990) in construction of several tests. Most of these tests aim at assessment of variables considered relevant to good teaching (Doyle, 1994). Aleamoni (1981) supports the student ratings of teacher performance as students are the main source of information about learning, and classroom environment including teachers' ability, competency, and communication skills.

The Teachers' Job Performance Scale (TJPS) is designed for the evaluation of teacher's performance at their workplaces. It can help to identify the strengths, and weaknesses of teachers' performance at individual, and organizational level, and can help to improve the quality and effectiveness of teaching. The scale is developed through a standardized procedure. The empirical evaluation shows that it has sufficient reliability and internal consistency. The construct validity of scale was determined through factor analysis and item analysis. The factors determined in the present research were named as TS (teaching skills), MS (management skills), DR (discipline and regularity) and IS (interpersonal skills). Initially six factors of job performance were determined on the basis of literature review, and opinions of judges. These factors were categorized as teaching quality, teaching style, subject knowledge, management skills, discipline and regularity, and interpersonal skills but empirical analysis revealed only four factors. So through committee approach, the items pertaining to teaching quality, teaching style, and subject knowledge were merged into one factor and it was named as teaching skills. Two of the items which were not clustered on any factor (Item loadings < .35) were excluded from the scale, and finally, TJPS was retained with 25 items.

The four-factor solution is consistent with previous studies (see, for example, Jahangir, 1988; Riaz, 2000). Although these studies were conducted on the sample of university students. The empirical basis of scale development procedure and ratings by judges also reveal the content and face validity of this scale. The TJPS has revealed the alpha coefficient for entire TJPS as .94, and for subscales, it ranges from .80 to .92. The values obtained were highly significant indicating the scale as a reliable and internally consistent measure. The split half reliability is also quite satisfactory and it has strengthened our results. Item total correlations and inter scale correlation also showed significant results and these results suggests that TJPS is a reliable and valid scale. As some researchers suggested that internal consistency may also be taken as evidence of validity (see, for example, Cronbach & Meehl, 1955).

Moreover, inter rater reliability was also determined to cross validate the ratings of students for their teachers, and it was observed that students' ratings are significantly correlated with headmistresses' ratings. The correlation with self-ratings of teachers were also computed, although it was non significant but positive correlation was existing. The correlation of teachers' self ratings and ratings by headmistresses was also significant.

To differentiate different categories of teachers' job performance i.e., poor, average, excellent, cut off points were determined through percentile scores. The range of cut off points enabled us to interpret the scores of the children against the sample studied. Although these cut off points could not be estimated as definite points. The assessment based on cut off scores has to be repeatedly validated in various groups of sample.

It may be concluded that TJPS is a reliable and valid scale, which is easy in administration and scoring procedure. It can be used in the schools of the region to evaluate the teachers' performance, wherever Urdu language is a means of communication.

REFERENCES

- Abrami, P. C. D., Apollonia S., & Cohen, P. A. (1990). Validity of student ratings of instructions: What we know and what we do not. *Journal of Educational Psychology*, 82, 219-231.
- Aleamoni, L. M. (1981). *Student evaluation of teacher performance*. Retrieved on 15 July 2003 from <http://ericae.net/edo/ED289887.htm>.

- Arvey, R. D. (1998). Performance evaluation in work settings. *Annual Review of Psychology*. Retrieved on 10 July, 2003 from www.findarticles.com/cf_0/m0961/v49/21036265/print.html.
- Campbell, J. P., McCloy, R. A., & Oppler, S. H. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations*. San Francisco: Jossey-Bass.
- Cascio, W. F. (1995). Whither industrial and organizational psychology in a changing world of work. *American Psychologist*, 50(11), 928-939.
- Cronback, L. J., & Meehl, P. E. (1955). Construct validity in Psychological tests. *Psychological Bulletin*, 52, 281-302.
- D'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198-1208.
- Doyle, W. (1994). Academic tasks in classroom. *Curriculum Inquiry*, 14(2), 129-149.
- Ferris, G. R., Bergin, T.G., & Wayne, S. J. (1988). Personal characteristics, Job performance and absenteeism of public school teachers. *Journal of Applied Social Psychology*, 18(7), 552-563.
- Forstyth, D. R., & McMillan, J. H. (1982). *Reactions to educational outcomes: Some effective and attributional correlates*. Paper presented at the annual meeting of American Psychological Association, Washington, D. C.
- Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52(11), 1182-1186.
- Heider, F. (1958). *The Psychology of interpersonal relations*. New York: Wiley.
- Ilggen, D. R., & Hollenback, J. H. (1991). The role of job satisfaction in absence behavior. *Organizational Behavior and Human Performance*, 19, 148-161.
- Jahangir, S. F. (1988). Student's Evaluation of their teachers' performance. *Pakistan Psychological Studies*, 6, 27-30.
- Marsh, H. W. (1984). Students' evaluation of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.
- Marsh, H. W. (1987). Student's Evaluation of the University teachers: Research findings, Methodological issues, and directions for the future research. *International Journal of Educational Research*, 11, 253-388.

- Marsh, H. W., & Roche, L. A. (1999).** Rely upon SET research. *American Psychologist, 54*(7), 517-518.
- McKeachie, W. J. (1990).** Research on college teaching: The historical background. *Journal of Educational Psychology, 82*, 189-200.
- Perry, R. P. (1990).** Introduction to the special section. *Journal Of Educational Psychology, 82*, 183-188.
- Riaz, M. N. (2000).** Student evaluation of the university teaching quality: Analysis of a teacher's rating scale for a sample of University students. *Pakistan Journal of Psychological Research, 15*(3-4), 107-117.
- Robbins, T. L., & DeNisi, A. S. (1994).** A closer look at interpersonal affects as a distinct influence on cognitive processing in performance appraisal. *Journal of Applied Psychology, 79*(3), 341-353.
- Scriven, M. (1996).** Assessment in teacher education: Getting clear on the concept. *Teaching and Teacher Education, 12*, 443-450.
- Swartz, C.W., White, K. P., Stuck G. B., & Patterson, T. (1990).** The factorial structure of the North Carolina teaching performance appraisal instruments. *Educational and Psychological Measurement, 50*, 175-182.
- Waldman, D. A., & Spangler, W. D. (1989).** Putting together the pieces: A closer look at the determinants of job performance. *Human Performance, 2*(1), 29-59.

Received: April 19, 2004.