

## **Devising a Scale for the Appraisal of Performance of Prospective Teachers during Model Lessons**

Muhammad Sadiq

Associate Professor (Education), Govt. Post Graduate College, Chishtian  
[m.sadiqap@gmail.com](mailto:m.sadiqap@gmail.com)

Ghulam Mustafa

M.Phil (Scholar), UMT, Lahore  
[gmustafa5686@gmail.com](mailto:gmustafa5686@gmail.com)

### **Abstract**

The main objective of the study was to devise a scale to appraise the performance of prospective teachers during model lessons. The scale was developed after an intensive review of related literature. The criterion validity of the scale was determined statistically using a valid scale 'Pinellas Country Schools Teacher Performance Appraisal' developed by PCS (2008) as criterion after confirming its textual and structural resemblance with the devised scale. Split-half method and coefficient alpha with values 0.76 and 0.74, respectively were calculated to determine the internal consistency of the scale. The averaged out inter-raters' reliability of the scale was 0.71. The relevant descriptive statistics pertaining to the distribution of appraisal scores with nominal deviations from normality provided a base to declare the scale a trustworthy tool for the intended job. The replicative execution of the scale confirmed the claim greatly when statistics were close to normality once again. In the last, the researcher suggested for the further improvement of the scale to make it a more useful tool.

**Keywords:** Prospective teachers, Appraisal scale, Model lesson, Scoring rubrics, Inter-rater reliability, Internal consistency, Psychometric measures

## **Introduction**

Teaching is a complex and many-sided process that demands a variety of attributes and abilities to meet the diversity in the classroom (Kumar, 2010). Therefore, a teacher is expected to be equipped with sufficient subject knowledge, workable classroom skills and laudable attitude to play a multiform role within and beyond the classroom (Armstrong & Savage, 1983). Arthur, et al. (2003) bring forth that a teacher has to deliver and demonstrate in the class, keeping rapport with the learners. In this way, his practices in the class count much along with his knowledge and manners (Woolfolk, 1995). That is why, teacher training programs comprise theory and practice all over the world.

Kumar (2010) calls teaching practice, a way of direct experience necessary to acquire professional competence. Bansal (2007 a, p.81) endorses this view as “teaching practice constitutes conditions for genuinely optimizing learning in pre-service teacher education programs.” Such programs are meant to enable fresh-on-the-scene teachers to meet the classroom challenges (Bansal, 2007 b). The writer suggests pairing beginners and veteran teachers, and promoting collegial collaboration important for effective practicum. Really, teaching practice provides a chance to the pupil teachers to plan, perform and evaluate the appropriateness of their efforts in the classroom.

Discussing another aspect in teacher training, Malderez and Wedell (2007) say that good teachers are born not made. Hence, there is need to engage right people to be teachers. It means that personal belief and sufficient confidence turn a person into a good teacher with the additional edge of teacher preparation programs. In the same direction, McGee and Fraser (2005) think that teachers work at the crossroads of personal and pedagogical qualities and the same determines the dynamic milieu of teaching. Tileston (2005) says that a pupil teacher needs to be creative to pursue original ideas. Further, he needs to be committed with the belief that the current era is of tremendous information that change rapidly. Hence, teaching proficiency seems to emerge from tendencies as well as from acquired behaviors.

Appraisal and rating of prospective teachers' performance during teaching practice is a demanding job. For valid evaluation, Peterson and Peterson (2006) suggest to make sure that the evaluators are familiar with

the procedure and intentions behind to make teaching practice useful. Bansal (2007 b) also calls training of evaluators important prior to evaluation to have trustworthy results. Kane et al. (2014) consider bias awareness crucial to enhance the credibility of rating. External evaluators may safeguard against this pitfall to a considerable extent. Meanwhile, Cyril and Poster (1993) reveal that appraisal of new teachers is uncomfortable both for appraisers and appraisees. However, they can be facilitated by incorporating pre-determined consensual criteria for the same and providing orientation to pupil teachers about this. In the same context, Sadker and Sadker (1997) believe in some rigorous rating evidence of pupil-teachers performance otherwise, it would be hard to reach the standards.

Teaching practice, as an integral part of B.Ed. and B.S.Ed. programs elsewhere has also been included in 2<sup>nd</sup> and 3<sup>rd</sup> semesters under code numbers EDT 12806 and EDT 21802 at Islamia University Bahawalpur. Each code has a weightage of three credit hours. In this regard, Canadian International Development Agency (2010) bears the activities and responsibilities of prospective teachers in detail that demand a critical, dynamic and unusual role on the part of pupil teachers during teaching practice and afterwards as regular teachers. However, it is regrettable that this inseparable segment of teacher training is commonly overlooked and is just taken as formality. Prospective teachers' classroom practices are not minutely observed, criticized and reformed. Niazi (1994) rightly says that mostly, trainers' visits are instant and customary. Their comments are unorganized, incomplete, ill-judged and ill-sorted. Prospective teachers are put to do much at their own. Here, their inbuilt tendencies, reconsideration of usual practices, seeing actions in term of consequences and urge to do things differently may make them good teachers. In the mentioned scenario, it seems important to make teaching practice meaningful rather, a matter of routine.

The situation gets worse at the time when marks in model lessons by external evaluators happen to misfit in a logical testing scheme due to the non-availability of some unanimous criterion for awarding. Everywhere, self-perceptive and self-judged norms on the part of different panels of evaluators make the grade award awkward and unjust to a great extent. Luck may confer favor on the candidates of a particular college for

appearing before a lenient panel and in contrast, it may deprive candidates of another college from compatible award due to facing a panel of strict evaluators. In this way, positions and divisions are considerably disturbed and consequently, prospective teachers suffer a lot either for higher studies or at the time of recruitment. The researcher noted this false trend in many affiliated colleges for years and found pupil teachers and their trainers complaining about it. Probably, the cause behind this fluctuation was the conflict of norms set for awarding grades in the minds of external evaluators.

In the light of above scenario, the need for the development of prospective teachers' performance appraisal scale was seriously felt. The recruitment criteria in Punjab adds to the worth of the scale where marks in NTS test and academic-cum-professional career solely determine the status of the recruits while marks in interview are kept a matter of routine to avoid any risk of exploitation and malpractice. Probably, it is assumed that B.Ed. program, especially teaching practice has already rated the prospective teachers in term of their professional efficiency for teaching and hence, already been considered for defining merit. Such perception adds to the demand of a credible scale.

Basically emerged from the Latin word "Scala" that means a ladder, a scale bears a series of ordered options at fixed intervals to use as standard for any measurement (Somer and Somer, 1980). Later on, such options are quantified according to set rules. In real, measurement in social sciences differs from that of practical life where well-trusted instruments like meter rod, clock, balance, electric-meter and gas meter etc. exist to measure physical quantities. The same does not happen to the measurement of constructs, attributes and performance where tests and scales are developed and standardized to make these worthy of trust and consistent. In this regard, validity, reliability and internal consistency add to the credibility of measuring tools. The present scale is a rafter of the same chain.

Besides the psychometric properties of a scale, its honest and accurate execution is utmost important. For this, the researcher recommends well-organized training sessions for evaluators to encompass plenty of practice opportunities, discussion on typical errors, comments on atypical rating to recognize probable mistakes and conceiving scoring

rubrics for grading range. The researcher also suggests observational vigilance, reasonable inspecting time for the operationalization of scale parameters, assigning sections of the scale as per desire of different raters and conducive environment to ensure on-target and unflawed rating.

Suggesting more in the above context, Popham (1981) advises to safeguard against rating instrument's flaws, procedural flaws and raters' bias for valuable rating. Further, raters' reluctance to pass unkind remarks about other people, incapability to understand built-in evaluative characteristics in scale parameters, halo-effect and generosity-factor need to be noted and deal for genuine rating. Even then, in the presence of all these considerations, Thorndike and Hagen (1977) bring forth that inter rater reliability on rating procedures is mostly low as they summarized many studies to conclude that correlation coefficient of independent raters is about 0.55. That is why; Somer and Somer (1980) recommend that other available criteria of performance may be used to supplement scale-based rating. Apart from suggestions made so far, Thorndike and Hagen (1977) additionally think of rater's ability and willingness as important factors for valid rating.

In the context of rating, scoring rubrics are quite popular and are reported to add to performance appraisal. Teacher evaluation rubrics cover different aspects of teachers' job performance and are commonly used in rating scales for more accurate results. Terminologically, rubric refers to a scoring guide to evaluate ratees' performance. Commonly, rubrics encompass evaluative criteria, limits to rate performance and a scoring strategy to make rating uniform and appropriate (Popham, 1977). Therefore, it can be said that a scoring rubric comprises standards or criteria linked to learning outcomes. Rubrics are usually presented in grids and tables, and assist to delineate stable criteria for marking. Goodrich (1996) brings forth that a series of steps is required to create scoring rubrics. The writer considers focus on measuring dimensions, concise hierarchical categorization, users-friendly approach and revision of feedback analysis important to improve scoring rubrics. The current study proceeds with the use of scoring rubrics.

## **Method**

The population of the study was the evaluators nominated to appraise the performance of prospective teachers in model lessons of the B.Ed. examination under different universities. The intention of the study was to devise a prospective teachers' performance appraisal scale and determine its psychometric properties to convert it into a valid tool.

## **Development and validation of the tool**

The leading determinants of teachers' performance were covered under four major titles i.e. "Planning and Preparation", "Classroom Management", "Delivery of Instruction", and "Monitoring and Assessment." Each title has ten indicators each entry assumes a criterion with level-wise performance descriptions as scoring rubric to ascertain systematic awards. It tried to embed necessary competencies, attitudes, skills and other requisites in the scale searched from relevant literature and shared by experienced teachers. The judgemental validity of scoring rubrics was ensured by incorporating the opinion of experts. Rubrics-wise close resemblance between "Teacher Evaluation Rubrics" by Marshall (2013) and the currently devised scale also affirms its validity. The validity of the total scale was ascertained by administering the current scale and 'Pinellas Country Schools Teacher Performance Appraisal' developed by PCS (2008) on 102 examinees and then correlating scores on the both with an index 0.81.

## **Execution of the tool**

The researcher made a personal request to the evaluators to assist to appraise 102 prospective teachers locally available in the 3<sup>rd</sup> semester of B.Ed. program under Islamia University, Bahawalpur. Only three evaluators got ready to perform the assignment. The researcher himself performed as the fourth evaluator keeping in view the tetra-dimensional structure of the scale. All the three evaluators were trained by the researcher in two sessions prior to administer the redesigned scale and the criterion. All the ambiguities and expected hurdles were discussed and resolved till the satisfaction of the evaluators. Additionally, the scale was tested on 10 respondents to note and solve on-the-spot queries in advance.

Again in the light of the pilot testing, certain amendments in the scale were made as suggested by the evaluators.

In this scale, “praiseworthy” denotes performance meeting very demanding criteria and tightly aligned with high standards. “Needs improvement” stands for mediocre performance. In other words, there is no contentment for remaining at this level. “Below standards” indicates unacceptable and poor performance calling for intensive efforts and improvement plan or otherwise suspension. The same will be the limits for the judgement of all parameters of the performance appraisal scale. The final scale becomes as under:

Table 1  
*Planning and Preparation*

Parameter	Performance Limits for rating		
	Praiseworthy (3)	Needs Improvement (2)	Below Standards (1)
Subject knowledge	Strong evidence of expertise in the subject	Somewhat knowledge of the subject reflects	Nominal familiarity with the subject was seen
Text clarity	Explicit and vivid for purpose	Partially fuzzy and inappropriate	Mostly confusing and misleading
Examples & illustrations	Well-chosen	Partially-fitted, needs reconsideration	Mis-fitted and misleading
Lesson plan	Well-woven, convenient to practice	Evidence of mistakes in content, syntax and lay-out	Conventional, unserviceable
Materials selection	A.v. aids and gadgetry suitable both for purpose and cohort level	Can be replaced with more workable alternatives, unwisely opted	Just formality and decoration, not meeting intended purposes
Teacher personality	Professional dress and appearance	Acceptable dress and appearance	Ridiculous dress and appearance
Anticipation	Anticipate students' misconception and confusions, also devises strategies to overcome these	Luckily considers one or two aspects for students to get confused from text	Proceeds ignoring possible misconceptions and confusions of students

Differentiation	Difficulty-wise well graded contents and properly treated	Knows hard contents but do not try to handle these accordingly	No differentiation among contents, uniform policy for hard and soft contents
Class routines	Well-defined with high awareness among students, simple to follow	Ambiguity in practicing routines, so so awareness of these among students, difficult to follow	No evidence of observation of classroom routines, impossible to follow
Teaching Method	The best and justified method incorporated	Method is just appropriate to work with	Method is inappropriate and unwisely selected
Overall rating: _____			
Comments:			

Table 2  
*Classroom Management*

Parameter	<u>Performance Limits for rating</u>		
	Praiseworthy (3) Below Standards (1)	Needs Improvement (2)	
Seating arrangement	Height-wise adjustment, accommodating needy, typical	Demanding alterations for betterment	Quite jumbled, unreasonable layout
Materials vicinity	Easy-to-access, readily available	Need some effort to approach	Quite difficult to access, cause wastage of time to approach
Lecture proceeding	Keeps lecture scheduled	Tries to control rough transitions	Fails to overcome ragged transitions
Nimble approach	Full-blown use of the entire period, uses 95% or above period time profitably	Sometimes loses teaching moments, uses 80% to 94% of period time usefully	Wastages most of the period, uses up to 80% of period time passively
Class control	Nips discipline problems in the bud, no disruption at all	Tries to spot and prevent discipline problems, casual disruption seen	Fails to detect and control discipline problems, continuous



Establishing procedures	Strong evidence of observing procedures by the class	Partial violation of procedures by some students	disruption observed Procedures completely overlooked. Absence of classroom procedures
Classroom environment	Culture of regard, respect, encouragement, flexibility and motivation	Dignity, inspiration and reinforcement restricted to some students	Culture of insult, discouragement, demoralization, teasing remarks and rigidity
Efficiency	Technicality, sharpness and exactness observed in proceeding	Unintentional inefficiency due to lack of clarity, interruptions, lazy transitions and off-task divergence	Loses time in late starts, early closing and gossiping
Behaving	Warmth, caring, respect and fairness for all students	Warmth, caring and respect for some students but not for others	Apathetic, harsh and disrespectful for all students
Material storage	Neat and organized storing with convenience to reuse	Storing is proper to some extent but needs rearranging	Poor and improper storage with a chance of damage
Overall rating: _____			
Comments:			

Table 3  
*Delivery of instruction*

Parameter	<u>Performance Limits for rating</u>		
	Praiseworthy (3)	Needs Improvement (2)	Below Standards (1)
Resourcefulness	Repertoire, persevering	Just expedient	Helpless, not fertile, good for nothing
Expression	Fluency and spontaneity	Communication gaps, doubt of time wastage	Stammering, hesitation, frequent communication blockage

Pronunciation	Ideal	Rare mispronunciation	Devastation of pronunciation
Gestures	Meaningful, communicate desired message	Commonly lacks relevancy between message and action	No more than over-acting
A.V. Aids/Gadgetry	Operating proficiency, right-time use	Operating imprudence, untimely use	Operating clumsiness, wrong time use
Appreciation	Welcomes new ideas	Indifferent from new ideas	Discourage new ideas
Engagement	Engage the entire class	Engage only the volunteers	No engagement at all
Balancing	Balances leniency and tightening	Tries to restore imbalances	Extremist, shows either lenience or tightening
Rewarding	Offers praise to deserving students	Offers praise indiscriminately, uses it as needless word	Shows miserliness to offer praise
Closure	Students capable to sum up learned task and apply it in different context, materials internalized	Teacher himself summarize the material and asks students to think about applications	Ends the lesson without closure, mentions nothing about tasks application in other contexts
Overall rating: _____			
Comments: _____			

Table 4  
*Monitoring and assessment*

Parameter	<u>Performance Limits for rating</u>		
	Praiseworthy (3)	Needs Improvement (2)	Below Standards (1)
Homework	Highly engaging	Somewhat different from routine	Solely customary and usual, mostly based on end-exercises
Comparisons	Makes intra and inter student comparisons	Makes only inter students comparisons	Sees performance in isolation
Questioning	Questions spread to entire class	Questions limited to specified students	Absence of questioning,

coverage			silence in this regard
Questioning level	Covering advanced cognitive levels	Restricted to rote-learning and just comprehension	Based on rote-learning, stereotyped
Tenacity	Manages extra coaching and guidance for needy	Manages some additional time and re-takes	Indifferent from needy and rely on advices
Intentions	Focuses on acquiring knowledge by students, underlying intention is knowledge	Focuses on attaining only marks by students, underlying intention is marks	Focuses on using short-cuts to get students only passed, underlying intention is avoidance from poor results
Diagnosis	Diagnosis students' problems & makes adjustments based on conclusions	Quickly goes through "know, what to know" to diagnose deficiencies	Takes start without diagnosing existing knowledge and skills
Feedback analysis	Checks misconceptions and makes corrections	Succeeds partially to draw conclusion for remedy	Entangled to make clear judgements for feedback to improve learning No knowledge or consideration of different cohorts to manage especial tips
Classification	Has high, average and low achievers on finger tips	Knows well only a few high and low achievers	
Comments	Well-judged and sorted- out comments on students' performance	Comments beneficial for students but incompletely	ill-judged and poorly- sorted out comments making no way to improve

---

Overall rating: \_\_\_\_\_

Comments: \_\_\_\_\_

### Quantification of data

The marks/points allocated were 3 for ‘Praiseworthy’, 2 for ‘Needs Improvement’ and 1 for ‘Below Standards’ levels of performance on the scale. In this way, the minimum possible score on each of the 40 criteria was ‘1’ and the maximum ‘3’ with a mean of ‘2’. In the same way, each evaluator can award marks possibly in the range 10 to 30 with mean value 20. Following the same pattern, the minimum possible score on the total scale was 40 and the maximum 120 with a mean 80. The scores by four evaluators were added to determine the overall credit of each respondent.

### Psychometric properties of the tool

Following tables bear the psychometric properties of the scale emerging from the analyses of data.

Table 5

*Credibility indexes pertaining to performance appraisal scale*

Criterion Validity	Reliability (Internal Consistency)		Interrater Reliability (Averaged out)
	Split-half	Coefficient alpha ( $\alpha$ )	
0.81	0.76	0.74	0.71

Table 5 reveals that the criterion validity of the teachers’ performance appraisal scale was 0.81 when determined by correlating ratees’ scores on the currently devised scale with that on “Pinellas Country Schools Teacher Performance Appraisal” developed by PCS (2008). The criterion bears 25 statements each with four proficiency levels—E (Exceeds Expectation), M (Meets Expectation), I (Expectation in progress) and N (Expectation Not Evident). The resemblance of both scales was confirmed by the experts prior to use the later on prospective teachers as criterion to know the validity of currently devised scale.

The split-half reliability on odd and even basis for the full scale was 0.76. However, it was 0.74 when measured with Cronbach’s coefficient alpha. The averaged out interrater-reliability of the scale was 0.71 with a range from 0.53 to 0.87 on part of different raters.

Table 6  
*Normality statistics pertaining to performance appraisal scale*

Mean ( $\bar{X}$ )	Median ( $\tilde{X}$ )	Standard Deviation ( $\sigma$ )	Skewness (Sk)	Kurtosis (Ku)
83.34	82.70	16.72	0.11	0.277

Table 6 puts forward the measures of central tendency and dispersion with respect to teachers' performance appraisal scale. The calculated values of mean and median are 83.34 and 82.70, respectively. The standard deviation of 16.72 brings forth the dispersion of the distribution of scores earned by prospective teachers. The observed skewness is 0.11, while the kurtosis is 0.277.

The researchers were interested to note how closely the results match normative statistics of the scale in case of its execution on other prospective teachers by different evaluators in the absence of the researcher. A God-sent opportunity occurred when Allama Iqbal Open University scheduled its model lessons in the second decade of February, 2019. It was a good chance to replicate the study. For this, the researcher requested evaluators appointed by the university to appraise model lessons of their 87 prospective teachers using the devised scale. The panel was trained on the pattern already used at the time to train evaluators of Islamia University, Bahawalpur to affix the credibility indexes of the scale. Astonishingly, results were concordant with the normative data sufficiently as:

Table 7  
*Replication of the study to note its correspondence with normative data*

Mean ( $\bar{X}$ )	Median ( $\tilde{X}$ )	Standard Deviation ( $\sigma$ )	Skewness (Sk)	Kurtosis (Ku)
85.05	84.13	15.31	0.26	0.244

Table 7 reflects that this time mean and median were 85.05 and 84.13, respectively. The standard deviation was 15.31. The skewness and kurtosis were 0.26 and 0.244, respectively. The slight departures were

sufficiently ignorable to consider the distribution normal. The phenomena confirm that the scale is quite stable to measure the performance of perspective teachers. It can be used safely subject to the prior training of the evaluators.

### **Analysis and results**

The psychometric properties of a tool relate the collected data on the tool to measure how accurately and consistently it meets the intended purpose in statistical terms in addition to logical and judgmental decisions. The criterion validity of the performance appraisal scale using “Pinellas Country Schools Teacher Performance Appraisal” developed by PCS (2008) was 0.81. The reliability of devised scale in the form of internal consistency was measured through split-half method and co-efficient alpha with indexes 0.76 and 0.74, respectively. The reliability co-efficient in both cases crossed 0.70, the value established by Mugenda and Mugenda (1999) to confirm the reliability of a tool to collect data with confidence. Mangal (2010) also considers the correlation co-efficient range 0.71 to 0.90 as marked relationship or high correlation to use a tool safely. The averaged out inter-rater reliability was 0.71 to reflect that independent judges have a high degree of consensus or high concordance between ratings. Reciprocally, the measuring errors were very small. The validity and internal consistency also confirmed that the scoring rubrics were valid both in content and level.

The normative statistics related to the appraisal scale was encouraging to declare it a distinguished tool for the intended purpose i.e. rating the prospective teachers’ performance. The  $\bar{X}$  and  $\tilde{X}$  on the scale with values 83.34 and 82.70, respectively were close to the mean value 0.80. The SD of 16.72 revealed good dispersion of scores across the wide-range of the distribution. The nominal Sk 0.11 near to the ideal value 0 indicated that the distribution was positively skewed but quite slightly. The Kr 0.277 near the reference 0.263 made the distribution, a bit platykurtic. However, the nominal deviations from the references were ignorable to declare and treat the distribution as normal and in turn, have confidence in the appraisal scale. The distribution of scores of replicative study also happened quite similar to that of the distribution of normative statistics with ignorable fluctuation as evident from tables 2 and 3. However, this

time, kurtosis got reversed to make the distribution leptokurtic but almost with the same magnitude as in case of normative distribution. Even then, there is no risk to declare the distribution of replicative study normal and in turn, have confidence in the tool. The slight differences between two distributions count no more to declare the scale, a good tool to perform its purpose consistently.

### **Discussion and recommendations**

The researchers made an attempt to make the assessment of prospective teachers systematic and evidence-based during model lessons through the lens of proposed performance appraisal scale. The scale embeds all the psychometric properties of a standard measuring instrument for the intended purpose. It seems to rationalize previous pattern of awarding purely tied with the discretion of evaluators due to the non-availability of some unanimous criterion that put some prospective teachers to reap the benefit of unjustified awards due to generosity of evaluators and some others suffer a lot due to miserliness of a particular panel of examiners unluckily. Such phenomenon may deprive many prospective teachers of their due right at the time of recruitment and higher studies due to under-assessment. On the other side, over-assessment cause silent and unintentional violation of merit. Therefore, it is needed to note and remove flaws and bad happenings to make awards fair and performance-based. The present scale is the remedy for the same.

In the prevailing situation, the proposed appraisal scale may safeguard against the existing exploitation and provide relief to a considerable extent in this regard. However, the training of evaluators seems important for the proper execution of the appraisal scale to bring uniformity in the procedure by all panels of evaluators elsewhere. It looks logical to get the devised scale administered by a panel of four evaluators every time due to its tetra-dimensional nature. It may allow each evaluator to focus on one segment adequately. It is also suggested to restrict the number of appraisees as 10 per day to see the things minutely and in depth.

Generality of the appraisal scale broadens its spectrum and provides a base to use it in other prospective also. It may be equally beneficial to assess performance of prospective teachers at 40 days

teaching practice, to evaluate regular teachers' classroom practices and to observe the classroom proceeding of teachers to be appointed in private sector through demonstration. Although, the scale is developed for the prospective teachers enrolled in teacher training colleges affiliated with Islamia University, Bahawalpur. Yet, it may be used for trainees of sister institutes imparting teacher training due to its scope and common nature. The replicative execution of the scale with perspective teachers of Allama Iqbal Open University affirms this claim. The scale also permits inter-evaluator and inter-segment comparisons also. The idea can also be extended to link teacher training imparted under different systems.

In spite of all the above, the devised appraisal scale claims no perfection. It is an initiation to do better. It is said that something is better than nothing. The possibility of enrichment, refinement and reorganization of the scale is always there. Awarding limits may also be reconsidered to make these self-explanatory, explicit and definite. Experts are openly invited for in depth analysis of the scale to meet its intended objective. At present, it seems to control the undue fluctuation of awards sufficiently on the part of different panels of evaluators and even ensures inter-institution consistency of appraisal.

In spite of the above, there is another point that the researcher gave equal weightage to all the segments embedded in the appraisal scale irrespective of their worth or share in the task. Surely, the resourcefulness of a teacher counts more than his/her dress or appearance. The researcher leaves it on the subsequent researchers to define and determine the relative importance of various segments. Luckily, the scale is sufficiently flexible to accommodate such alterations. Additionally, the scale can also be incorporated to compare strong and weak componential areas of teacher training within or between the institutions through segment-wise analysis of different sections of the appraisal scale. In this way, it can add to strengthen current teacher training contents and practices.

### **References**

- Armstrong, D. G. & Savage, T. V. (1983). *Secondary Education: An Introduction*, London: Collier MacMillan Publishers
- Arthur, M., et al. (2003) *Classroom Management: Creating Positive Learning Environment*, Victoria: Thomson



- Bansal, H. (2007 a) *Modern Methods of Teacher Training*, New Delhi: APH Publishing Corporation
- Bansal, H. (2007 b) *Teacher Training Concepts*, New Delhi: APH Publishing Corporation
- Canadian International Development Agency, (2010). *Manual for Teaching Practice*, Lahore: CIDA
- Cyril and Poster, D. (1993) *Teacher Appraisal: Training and Implementation 2<sup>nd</sup> ed.*, London: Routledge
- Goodrich, H. (1996) "Understanding Rubrics" *Educational Leadership*, 54(4), 14-18
- Kane T. J., Kerr, K. A. and Pianta R. C. (2014). *Designing Teacher Evaluation Systems* (edit.), San Francisco: Jossey-Bass
- Kumar, T. P. (2010). *Teacher Education*, New Delhi: APH Publishing Corporation
- Malderez, A. and Wedell, M. (2007). *Teaching Teachers*, Wiltshire: The Cromwell Press
- Mangal, S.K. (2010). *Statistics in Psychology and Education 2<sup>nd</sup> ed.*, New Delhi: PHI Learning Private Limited
- Marshall, K. (2013) *Rethinking teacher supervision and evaluation 2<sup>nd</sup> ed.*, San Francisco: Jossey-Bass
- McGee, C. and Fraser, D. (2005). *The Professional Practice of Teaching 2<sup>nd</sup> ed.* (edit.), Victoria: Dunmore Press
- Mugenda, O.M. & Mugenda, A. G. (1999). *Research Methods: Qualitative and Quantitative Approach*, Nairobi: Act Press
- Niazi, M.A. (1994). *Educational Issues in Pakistan*, Lahore: Book Talk, 3-Temple Road.
- PCS (2008) "Pinellas Country Schools Teacher Performance Appraisal" *Teacher Performance Appraisal Manual*, Florida retrieved from [https://www.nctq.org/dmsView/18-07\\_6670](https://www.nctq.org/dmsView/18-07_6670)
- Peterson, K.D. and Peterson, C. A. (2006) *Effective Teacher Evaluation*, Thousand Oaks: Corwin Press
- Popham, J. (1977) "What's Wrong-and What's Right-With Rubrics" *Educational Leadership*. 55(2). 72-75(<http://eric.ed.gov/?id=EJ552014>) Retrieved on 08-02-2019
- Popham, W.J. (1981) *Modern Educational Measurement*, Englewood Cliffs, NJ: Prentice-Hall, Inc.

- Sadker, M.P. and Sadker, D. M. (1997). *Teachers, Schools and Society*, New York: The McGraw-Hill Company
- Sommer, R. & Sommer, B. B. (1980). *A Practical to Behavioral Research: Tools and Techniques*, New York: Oxford University Press
- Tileston, D.W. (2005) *10 Best Teaching Practices* 2<sup>nd</sup> ed., Thousand Oaks: Corwin Press
- Thorndike, R. L. & Hagen, E. P. (1977). *Measurement and Evaluation in Psychology and Education* 4<sup>th</sup> ed., New York: John Wiley and Sons.
- Woolfolk, A.E. (1995). *Educational Psychology*, Boston: Allyn and Bacon