

PERFORMANCE APPRAISAL SYSTEM OF BANGLADESH CIVIL SERVICE: AN ANALYSIS OF ITS EFFICACY

Mohammad Ashraf Haque

ABSTRACT

Performance appraisal plays a crucial role in maintaining a high quality and competitive workforce, but has been ignored so far from the civil service reform agenda in Bangladesh. At present, an Annual Confidential Report (ACR) that was introduced in the Pakistan period and retained with minor modification is used to measure performance. The prevailing gross inefficiency of the Bangladesh Civil Service gives the impression that this appraisal system is poorly functioning. This paper articulates the difficulties in designing a good appraisal system in the public sector and provides a theoretical framework arguing that the 'goodness' of a PA system should be seen in its capacity to meet its predetermined objectives instead of counting on the so called infallible psychometric properties. A certain degree of validity, reliability and perceived fairness are essential for an appraisal system to be efficacious. The paper then analyzes historical development of the current ACR of Bangladesh Civil Service and its components to measure its validity, reliability and perceived fairness. The paper found that though the current ACR holds some degree of fairness, it significantly lacks in validity as well as reliability, and therefore is deficient in meeting its stated objectives.

Keywords – annual confidential report, efficiency, perceived fairness, psychometrics, reliability, validity

INTRODUCTION

“Survival of the fittest” once advocated by the famous biologist Charles Darwin in his ‘Origin of Species (1859)’ eventually transcended the field of Biology and has become the guiding principle of all interactive processes in which competition determines ultimate fate. In the realm of organization, this principle operates in the guise of productivity, profitability, efficiency, and effectiveness. Organizational theorists from the classical period emphasized structural and technological perfection for survival and growth. Such mechanistic ideas of competitiveness started fading since 1930s onwards. The rise of behavioral and industrial humanism made way for ‘human capital’ to be considered central in the study of competitiveness. The measurement unit of the contribution made by human capital is termed ‘Performance’.

Copyright: © 2012 Haque. Copyright for this article is retained by the authors, with first publication rights granted to the International Public Management Review (IPMR). All journal content, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. By virtue of their appearance in this open-access journal, articles are free to use, with proper attribution, in educational and other non-commercial settings.

Corresponding Author: haqash.01@gmail.com

Performance Appraisal (PA) as a technique of measuring employee competitiveness is not a modern phenomenon. In managing a large scale labor force, the first notable initiative in systematic performance appraisal was introduced by Robert Owen in his Lanark textile mills during the 1800s. As part of the Owen's appraisal, a piece of wood was placed before each employee's machine each day marked with a different color, which conveyed the supervisor's opinion of that worker's performance the previous day (Newstrom and Davis, 1993). Thus Owen's appraisal system could aggregate annual assessment of employee performance and recorded it in a 'book of character'. However, that assessment process was subjective, which is one of the main criticisms that still haunts the performance appraisal systems of today.

In the early part of the twentieth century, Owen's ideas were superseded by those of 'scientific management' propounded by F.W. Taylor. This emphasized the use of quantitative methods to measure work performance (e.g., a fair days work and differential piece rate system), and led to the development of 'work study' techniques during the inter-war period.

Notwithstanding the private sector has been credited as the protagonist of the performance appraisal (PA) techniques and justifiably so, the modern psychometric approaches to performance appraisal have been first invented and practiced in the public sector (Daley, 1992). Due to its monopolistic and legalistic nature, public sector performance appraisal practices could not reach the standard achieved in the private sector, due mostly to the latter's sensitivity to market (both factor market and supply market) competition.

The PA as an efficiency tool was reinforced in the public sector under the wave of the liberal reform popularly known as the New Public Management (NPM) during the late 1970s (Daley, 1992). The assumption that the public sector can operate under market like forces and foster competition allows rethinking about how private sector techniques can be applied in the public sector. Consequently, performance management of public officials got momentum aiming to know 'who contributes what, why, and how' in order to create fair value for public money. Since the 1980s, many countries reviewed their PA systems in order to make the most out of their civil servants.

Bangladesh was not insulated from this wave of reforms. In fact, it started initiating NPM-type reforms under the rubric of the 'Structural Adjustment' in the early 1980s. Initially, the focus was on structural variables like reorganization of ministries/divisions, corporatization and privatization of state owned enterprises, and capacity building and right sizing of administrative agencies. But the apparent failure of structural transformation to produce expected result implies the importance of performance of the officials who man and run the structure and use the technology in producing the desired result. Needless is to say that the outcome of any reform effort depends ultimately on the ability and the willingness of officials to transform such reform opportunities into improved organizational performance. The importance of a sound performance appraisal system therefore cannot be ignored. Unfortunately, this vital aspect of administrative development has been so far neglected from the reform agenda of the country. This article will revisit the PA system of Bangladesh Civil Service in order to track its historical development and analyze its efficacy in measuring the capacity of the

officials. Two critical questions that will be explored to measure the intended efficacy are:

1. Is the current appraisal system valid, reliable and fair under the lens of the theoretical requirements?
2. If not, what factors are compromising the validity, reliability and perceived fairness?

THE SIGNIFICANCE OF REFORMING THE PA SYSTEM

The PA system is essential for maintaining human resources at the desired standard. However, an ineffective appraisal system can bring many problems including reduced morale, decreased employee productivity, and diminished sense of belonging (Somerrick, 1993). Therefore continuous improvement in the PA system is a precondition for high performing organizations. In Table 1, we have compiled some indicator based ranking of civil services of six countries. Three countries that are shaded, i.e., the UK, Singapore and Malaysia, significantly outperformed the rest, i.e. India, Pakistan and Bangladesh. The UK, Singapore and Malaysia are well known for their performance appraisal reforms and all of them introduced performance management systems (Demmke, 2007; Siddiquee, 2006). Even India and Pakistan also introduced more defined and segregated appraisal systems and also scored better than Bangladesh. We do not claim that such differences in performance are the function of the difference between their respective PA systems. However, we can arguably claim that having a better PA system is one important variable of the performance function.

Table 1: Comparative Efficiency of the Civil Service of Selected Six Countries

Country	Revenue as % of GDP	Registering Property (rank)	Paying tax (rank)	Enforcing contracts (rank)	Government Effectiveness (percentile rank)	Regulatory Quality (percentile rank)
Bangladesh	10.96	176	89	180	23	21
India	15	93	169	182	54	47
Pakistan	13.37	119	143	158	26	35
Malaysia	24.5	86	24	59	84	60
Singapore	21.73	16	5	13	100	100
UK	38.41	23	16	23	94	98

Source: The author compiled the latest data from World Development Indicators 2010, Doing Business Indicators 2010, and Governance Indicators 2010

EFFICACY OF THE PA SYSTEM: THEORETICAL FRAMEWORK

Whenever we talk about ‘performance’, we deliberately compare the accomplished task of an employee against some predetermined standard. When the case is about civil service, we are puzzled by the non-metric nature of accomplishments and standards. For example, it is easy to set a contraceptive distribution target for a family planning worker and make him/her accountable based on the target achievement. But is it at all possible to make him/her accountable for reduction in total fertility rate or crude birth rate, which is the final objective of this public program? There are many intermediate and independent variables that lie outside the control of the performer and the concerned organization. In other words, the activities of a civil servant being social and political in nature are difficult to measure, let alone measure it objectively.

Two specific inferences can be drawn from the principles of the Chicago School of Economics and the Public Choice theory. The former delivered the Principal-Agent problem saying that the contract between the principal (the executive politician) and the agent (the administrator) in the public sector is unquantifiable and highly susceptible to information asymmetry (Burgess et al., 2004). Likely consequences that follow are the subjective appraisal and the overrating aiming at safeguarding the bureau interest. The Public Choice School claims that the PA system is susceptible to political manipulation in ways that maximize the self-interest of the appraiser. Political manipulation could, if it becomes pervasive, conform to the *Gresham's Law*¹ in the appraisal process.

Prentice et al. (2003) found the following bottlenecks in measuring the performance of the civil servants:

1. The divergence between legalistic parameters and managerial discretion;
2. The probable impact of political influence;
3. Ill-defined organizational purpose; and
4. The absence of a measurable bottom-line.

This in no way implies that the performance appraisal is irrelevant or counterproductive in the public sector. Instead, the difficulty of performance measurement in the public sector demands for more cautious and knowledge based intervention in developing a PA system that is capable of distinguishing between the good and the bad performance. Here comes the question of the efficacy of the PA system.

Research findings confirm that performance measurement techniques developed so far are not free from biases (Cleveland & Murphy, 1992; Dewberry, 2001; Scott and Einstein, 2001; and Lam and Schaubroeck, 1999). Therefore, absolute accuracy in performance measurement is a near impossibility. Anderson (2002) proposed that absolute accuracy need not be an inalienable determinant of the efficacy of the PA system. Instead, a PA system can be considered effective if it produces the desired result despite weaknesses in the measurement. McGregor (1987) advocated three categories of functions that a PA system performs, viz. administrative functions (e.g. promotion, transfer, salary), informative functions (e.g. generate performance data for management and employees), and motivational functions (e.g. training, learning, career planning). Other objectives of a PA system include clear definition and segregation of tasks, assignment

of responsibility and recognition of accomplishments (Landy and Farr, 1983; Latham and Wexley, 1981; Lawler, Mohrman and Resnick, 1984; Cleveland and Murphy, 1995). This array of objectives of a PA system should be seen as more of a range of choices rather than the binding outcome of a PA system regardless of the context. The efficacy of a PA system therefore can be measured by how effectively it meets its pre-determined objectives.

PA systems vary in their emphasis on factors that ensure efficacy. In this respect, Folger et al. (1992) summarize three distinct categories of PA system. First among them is the test metaphor. PA systems of this category emphasize psychometric properties of employees and developing appropriate rating scales for measuring those properties. Popular psychometric tools can be either criterion-referenced or norm-referenced (Milkovich and Wigdor, 1991: 55). In norm-referenced appraisal, employees are ranked relative to one another based on some traits, behavior, or output measures. This procedure does not necessarily involve the use of a performance appraisal scale. Examples include Ranking, Critical Incident, Essay, Forced Choice method, etc. In criterion-referenced performance evaluation, the performance of each individual is judged against a standard defined on a rating scale. Examples include the Graphic Rating Scale (GRS) and the Behaviorally Anchored Rating Scale (BARS). Test metaphor PA systems count on the validity and the reliability of measurement for their efficacy. The assumptions are: 1) tasks can be easily distinguished, compared and measured either in absolute or relative terms; 2) raters have considerable information on ratee's task behavior and are capable to assess ratee's performance accurately; and 3) uniform criteria for measuring performance exist (Folger et al., 1992).

The second category of PA systems is classified under 'political metaphor'. Scholars argue that psychometric scales become less valid for jobs that are not routine and standardized. They are not wholly reliable due to the rater's cognitive limitations (DeNishi et al., 1984; Ilgen and Feldman, 1983). Politics is intertwined in the appraisal process as supervisors tend to use their power in resolving conflict over distribution of scarce resources (i.e. promotion, salary increase) among the employees. Use of power is also prominent in directing employee effort in a way that best achieves organizational goals (Pfeffer, 1981; Cascio, 1982). Such political maneuvering of the appraisal system could go with all psychometric techniques as well as with the Management By Objective (MBO). However, political manipulation of PA system for parochial interests instead of realizing organizational goal cannot be ruled out.

The due process metaphor, i.e. the third category advocates for introducing distributive and procedural justice in the PA process irrespective of their psychometric or political orientation. The assumption is that perceived fairness of the appraisal process by the raters and the ratees is critical to the legitimacy and the success of the PA system. Folger et al. (1992) suggests that the employee participation in formulating performance standards and open communication in the measurement process, fair hearing, and evidence based measurement create a sense of fairness among the raters and the ratees.

The foregoing discussion makes it clear that the validity, reliability and perceived fairness/justice are three critical determinants of the efficacy of a PA system. In the follow-

ing sections, we will discuss about the measurement process of these three determinants.

Validity of Measurement

When a tool measures what it is supposed to measure is called a valid tool. Cronbach (1990) argued for three kinds of validation.

1. Predictive Validity: Score in the present performance is consistent with the score of some future performance. This requires rating of some general abilities along with job specific rating.
2. Construct Validity: Score that can clearly and measurably differentiate between the relative performances is constructively valid. This requires proper definition of scales and standards. Definition should aim at increasing objectivity in measurement of performance across the job.
3. Content Validity: Factors that are to rated are representative of the job performance. This requires extensive job analysis and identification of performance factors therefrom. The measurement should be grounded on solid behavioral or outcome based evidence.

McDonald (2005) argued that incomplete definition of performance dimensions and the use of an inadequate measurement technique diminishes validity. Validity may be difficult to establish if performance criteria across various jobs are not clearly evident. This is more so for jobs that are non-routine and frequently changing. High level managerial jobs are of this kind. This is why managerial performance dimensions are often stated in non-behavioral terms and raters have to score such dimensions based purely on judgment that might not align with task behavior. In developing common performance dimensions across jobs, scholars advocate for the use of competency based assessment. 'Competency' is defined as the mix of knowledge, skills, abilities and behavior that can be universally applied to different jobs. For example, the superior, average or below average job knowledge can be defined for various jobs, and reasonably compared (Spangenberg, 1990; Spencer and Spencer, 1993).

Predictive validity requires inclusion of some general but valued performance dimensions that might accommodate unusually good contextual performance. For example, qualities like interpersonal skills in communication and team work, cooperation with coworkers, initiative in undertaking extra role and responsibility might not be the part and parcel of all jobs, but the presence of these qualities regardless of jobs help organization perform better (Oran, 1997).

Due to the great variety and complexity that surround public sector jobs, the PA systems of the public sector usually use the multivariate performance dimensions for ensuring the validity. For example, civil services of the EU use a mix of dimensions involving quantitative and qualitative indicators, social and psychological criteria, functional criteria, etc. ((Demmke, 2007). Segregation of the appraisal system across different job families can be applied in order to accommodate diversified jobs with appropriate performance dimensions and measurement scales.

Reliability of Measurement

When measurement is consistent under all conditions and by all raters, then it is reliable. This requires objective standardization of the performance. The validity of the measurement takes into account task related competencies alone. But many exogenous factors can also affect the performance significantly. This is why these factors should be considered during measurement (Armstrong and Baron, 1998). Quality of supervisory direction, support of colleagues, work environment and other contextual factors and even chance factors can affect performance in a way that an individual cannot control. Under such variable conditions, the same performance is subject to varied interpretation by different raters, and the same competencies and efforts are susceptible to getting varied score by the same rater.

Inter-rater reliability is another concern in the appraisal process. This may be a result of the difference in the cognitive processes of the raters. The common rating errors such as halo effect, horn effect, leniency, central tendency, recency effect, first impression, similarity bias, stereotypes and such other forms of subjectivity reduce inter-rater reliability. Murphy and Cleveland (1992) state that if raters have a high degree of job knowledge about the ratees, open communication exists between them, and ratees have an opportunity to appeal against the rating, inter-rater reliability is likely to increase. Extensive training for raters and the standardization of measurement also significantly contribute to reducing biasness in measurement (DeNishi, 1984).

Perceived Fairness in the Measurement

Employees' perceived fairness about the appraisal system is one critical aspect of the efficacy (Bretz, Milkovich and Read, 1992). It establishes legitimacy of the process and secures genuine effort from the raters and the ratees to make appraisal work for the organization. Skarlicki and Folger (1997) found the appraisal process a potent source of dissatisfaction when employees consider the system as biased and unequal. Equity theories of motivation argue that employees always compare between their own efforts and the rewards they get against those of other employees. Whenever they feel injustice, they reduce their effort and by doing so justify the rightness of a fallacious appraisal system.

Most scholars identify distributive and procedural justice as sufficient condition for a fair appraisal process (Byrne and Cropanzano, 2001). Distributive justice in the appraisal process refers that the right performance has been linked to right reward. As to this particular meaning, distributive justice is directly related to the validity of measurement. On the other hand, procedural justice emphasizes on the equitable performance measurement process. In other words, procedural justice requires that the appraisal process has adequate countervailing measures against discriminatory rating. Folger et al. (1992) articulated three essential factors for procedural justice: adequate notice, fair hearing and evidence based judgment. Scholars even advocate that employees will endure a certain amount of unfairness in the actual rating if they are convinced that the appraisal process that the system follows is fair.

THE EFFICACY OF THE PA SYSTEM AND EMPLOYEE MOTIVATION

Using Vroom's 'Expectancy Theory', Anderson (2002) has shown that the validity, reliability and equity in the PA system provide sufficient condition to motivate both the raters and the ratees in using the appraisal system for organizational success. The raters put high valence on the PA system if they perceive the measurement process as valid and reliable. Perceived validity and reliability of the PA system motivate the ratees to change behavior for improved performance. In other words, validity and reliability work as the instrumentation to the outcome of the appraisal system. Perceived fairness, i.e. equity in the appraisal system creates expectation among the raters and the ratees that each employee will be fairly evaluated and awarded with equitable rewards. Thus the efficacy of the PA system motivates the raters and the ratees to give due importance to the appraisal system that further fortifies the efficacy of the PA system.

In the following sections, we will provide a brief history of the appraisal system of Bangladesh Civil Service and then analyze its validity, reliability and perceived fairness.

Historical Development of the Performance Appraisal System of Bangladesh Civil Service (BCS)

Under the Company rule of the British India, a performance appraisal system was first introduced. In 1836, a quantitative measurement of works performed was introduced under the name of 'Annual Returns' (Misra, 1970). This was not confidential. When a superior officer was transferred from his work station, he used to write a note on the performance of his subordinates for the information of the new officer. This practice was continued till the end of Pakistan period and named as 'the Note to Successor'. Annual Confidential Report (ACR) as an appraisal technique was introduced in pursuance of a provision under the Board's Miscellaneous Rules 1934 (Misra, 1970). During this period, emphasis was given on measuring administrative competence, loyalty, temper, language skill, and public relation. Since then the ACR survived as the performance appraisal mechanism of Civil Services of Pakistan and Bangladesh.

During Pakistan period, Graphic Rating Scale (GRS) along with a pen picture was introduced to measure both the personality and the performance traits (BPATC, 1989: 21). After independence, the Administrative and Services Reorganization Committee (1973) reviewed the inherited ACR system and recommended to use a similar form for all types of employees for the sake of comparability and equality. The Pay and Service Commission (1977) recommended measuring the potential of an employee as to make better promotion decisions (BPATC, 1989: 22).

In 1982, the then government formed a Martial Law Committee to review the ACR. The committee was heavily influenced by the system of appraisal used in the defense services. Following its proposal, a new ACR was introduced in 1982, which was a significant break from the past (Hossain, 2004). The number of parts of ACR was increased from 4 to 10 to accommodate newer needs, e.g. a structured pen picture, a health check up, a job description, a pen picture part for inserting useful observation about performance, and a separate column for inserting additional remark were introduced. In the

pre-1982 ACR, employees used to be evaluated based on the overall impression on different factors as perceived by the Rater. The 1982 ACR introduced two dimensional measurements- some factors dealing with the personality traits and some other are related to job performance (Hossain, 2004). The objectives of the ACR also broadened to identify of employee training needs and suitability for placement, which were in the past limited to administrative control function only. Moreover, the system was made open in 1982 to give employees feedback about their performance. But the open system was discontinued in 1984. During the open appraisal phase, the rater had to fill part three to part five of the ACR, i.e. job description, personality rating and performance rating after consultation with the ratee. But after 1984 when the ACR again made confidential and onward, the ratee fills job description and the rater fills personality and performance factors and no consultation is required during filling up the ACR. The BPATC study (1989) argued that during the open phase, the rating was significantly inflated and tension between the rater and the ratee was high. This view is in sharp contrast with the theoretical construct of perceived fairness as discussed earlier.

In 1990, the ACR format was again changed. In the following table fundamental differences between the post-1982 and the post 1990 ACRs are highlighted.

Table 2: Fundamental Differences between ACR-1982 & ACR-1990

Criteria	ACR-1982	ACR-1990	Evaluation
Measurement Factors	Total 20 factors: 8 personality and 12 performance factors	Total 25 factors:13 personality and 12 performance factors	Personality has been given more emphasis which is difficult to measure
Rating Scale	Ranged from 1 to 5; each with generalized anchor definition, e.g. anchor 5 of punctuality trait was defined as Never late in office	Ranged from 1 to 4; no anchor definition	Without anchor definition, ACR became more prone to subjectivity
Summary Scale	Top 2 grades thresholds were 91 and 81	Top 2 grades thresholds were 95 and 85	Getting of top grades became stricter

Source: prepared by the author

A careful attention to the table raises skepticism about the aim of reforming the ACR in 1990. Renewed emphasis on the personality traits as was the case with the pre-1982 ACR gives the impression that the subjectivity is preferred to objectivity in the appraisal process. One argument in favor of this reform might be the desire to bring uniformity in the appraisal process among different cadres. In BCS, there are 28 functional cadres and the promotion to the deputy secretary level² from a pool of candidates from each cadre require good score in the ACR. As each cadre has unique functional nature, more emphasis on general personality traits instead of performance factors has been assumed conducive to bring about uniformity in inter-cadre evaluation. Another confusing change made in 1990 was the omission of the anchor definitions from the scale. Graphic

Rating Scale (GRS) requires definition of each anchor on the measurement scale (Snell and Bohlander, 2007). This change therefore was a departure from good practice.

For further evaluation, we can compare the measurement factors of the two ACRs (cf. Table 3 and 4).

Table 3: Comparison of Personality Factors Used in ACR

ACR 1982	ACR 1990	Remark
Sense of Discipline	Same as 1982	No change
Judgment and Sense of Proportion	Do	Do
Intelligence	Do	Do
Initiative	Do	Do
Public Relations	Do	Do
Cooperation	Do	Do
Personality	Do	Do
Security consciousness	Do	Do
-----	Punctuality	It was a performance factor in 1982
-----	Reliability	New addition
-----	Sense of Responsibility	Bifurcated from performance element of 1982
-----	Interest in Work	Transferred from performance factor of 1982
-----	Promptness in taking action and carrying out orders	Do

Source: Adapted from Hossain, 2004

These two tables reveal that there are no significant differences between the two. In some cases, factors were interchanged. This cast a serious doubt on the validity of the measurement criteria, viz. what guiding principles the designers of the ACR followed to differentiate between the personality and the performance factors. Some new factors were introduced, but most of those were only replication of an existent factor, except for three newly added performance factors. For example, it is difficult to differentiate ‘the reliability’ from the ‘sense of responsibility’ though both were newly introduced in 1990.

Table 4: Comparison of the Performance Factors Used in the ACR

ACR 1982	ACR 1990	Remark
Professional knowledge	Same as 1982	No Change
Quality of Work	Do	Do
Quantity of Work	Do	Do
Punctuality	-----	Transferred to Personality factors in 1990
Sense of responsibility and devotion to duty	Devotion to duty	Sense of responsibility was transferred to personality factors
Promptness in taking action and carrying out orders	-----	Do
Interest in work	-----	Do
Ability to supervise and guide	Same as 1982	No change
Relationship with colleagues	Do	Do
Capability to implement decision	Do	Do
Ability to express- Writing	Do	Do
Ability to Express-Oral	Do	Do
-----	Capacity to make decisions	New
-----	Interest and capacity to train subordinates	New
-----	Promptness in initiating and countersigning ACR	New

Source: Adapted from Hossain, 2004

In 1997, the ACR format was again changed. This time, the number of parts of the ACR reduced to 8 from 10. Job description was merged with the Bio-data and the Additional Remark part was abolished. Apart from this, the pen picture part was made unstructured. In the 1982 ACR, the pen picture part was structured into five components: personality, professional ability, loyalty and reliability, any other strengths/weaknesses that sharply distinguishes the ratee, and advices to the ratee. The structured pen picture provided more uniformity and managerial expediency in measurement and recording. This giving more flexibility to the rater was inconsistent with the uniformity principle that was the dominant theme of the post 1990 reform.

The following points can be summarized from historical development of the performance appraisal system of BCS:

- Except the 1982 reform, other reforms of the appraisal system were marginal and disjointed from a clear sense of administrative development purpose;
- The post 1990 reforms brought about more subjectivity in the appraisal process that might lower validity and reliability;
- Open communication in the appraisal process that was likely to increase perceived fairness was claimed to lower validity, reliability and legitimacy of the appraisal system in BCS;
- Inconsistency in the classification of the personality and the performance factors was obvious and therefore likely to lower validity and reliability; and
- Inconsistency was found in the attempt of bringing about uniformity in the appraisal process. Most importantly, uniformity was seen inversely proportional to objective measurement and directly proportional to subjective judgment in the appraisal process which is contradictory to the theoretical propositions.

In the following sections, we will analyze the validity, reliability and perceived fairness of the current appraisal system of BCS in order to measure its efficacy. The analysis will be within the bounds of the theoretical framework given earlier and wherever appropriate evidence from secondary sources will be cited to corroborate our analysis.

ANALYSIS OF VALIDITY, RELIABILITY AND PERCEIVED FAIRNESS OF THE PRESENT APPRAISAL SYSTEM

In accordance with the theoretical framework, we will examine the content, construct, and predictive validity of the current system. In order to expedite uniformity, we will also assess the generality of the evaluation factors. To crosscheck the reliability, our focus will be on the contextual and the chance factors that cause score and interpretation of the same performance to vary. To what extent raters enjoy the discretionary judgment in the evaluation process will also be considered. Perceived fairness will be examined in terms of procedural justice, i.e. the extent of adequate notice, right to participate and appeal, and evidence based rating.

Features of the Existing System

For the sake of expediency, we describe the features of the ACR by linking them to four stages of the appraisal process that are usually practiced in the BCS. In subsequent sections, we will explain each of the defining features and link them to validity, reliability and perceived fairness.

Table 5: Features of the Present ACR

Stage	Focused Area	Defining Features
1. Objectives of the PA System	1. Nature of the objectives	1. Static, One-dimensional, and Multiple, Control
2. Performance Planning	2.1 Job Description	2.1 Narrative, Disjointed
	2.2 Performance Standard	2.2 Discrete-qualitative
3. Performance Rating	3.1 Report Initiating Officer (RIO) or the rater	3.1 Superior Officer
	3.2 Nature of Rating	3.2 Uniform, Pseudo-objective
	3.3 Rating Period	3.3 Annual/ Continuous
	3.4 Criteria of Rating	3.4 Personality Trait and Performance Factors
	3.5 Method of Rating	3.5 GRS and Open Essay
	3.6 Summary Rating	3.6 Cumulative GRS Score
	3.7 Evaluation Process	3.7 Semi-closed, Non-participative, Confidential
	3.8 Rater's Disposition	3.8 Discretion
	3.9 Ratee's Right	3.9 Right to Prior Guidance and Appeal
4. Feedback and Action	4.1 Nature	4.1 Partial, Negative
	4.2 Core Action	4.2 Control

Source: prepared by the author

Stage 2: Performance Planning

The first item of the second part of the ACR provides with the job description (five relatively important tasks can be mentioned) upon which the official will be rated. The Manual says that five major duties and responsibilities will be mentioned in quantitative terms, if possible (Section 3.7). During the open phase period in between 1982 and 1984, the performance target used to be set based on the discussion between the rater and the ratee (Section 3.8) and the latter fill this item. After abolishment of the open appraisal phase, this part is not being filled by the ratee. This practice violates the reliability principle as the ratee and the rater might differ about the responsibility of the concerned ratee. Initial consultation for determining ratee's responsibilities for a particular period and the midterm review and revision of such responsibilities are likely to increase reliability by reducing cognitive bias. Periodic communication about job responsibilities is likely to increase perceived fairness as employees could know what are expected of them and how they could achieve a good rating. The supervisor also has the opportunity to know about the contextual and the chance factors that might accidentally affect the performance of an employee. Haque's study (2010) shows that 49% of the respondent civil servants are of the opinion that their ACRs are not based on perfor-

mance planning. 45% of them think that ACRs are partially based on performance planning. This gives the impression that ACRs are at present prepared on the rater's discretionary judgment about the performance instead of punctiliously validity and reliability.

Job description (JD) of the civil servants has vital role in this respect. To bring about uniformity in the rating process, the same position at different work stations should have similar JD, though the target might vary. In practice, JDs in BCS can be termed at best inconsistent set of narratives as there have hardly any attempts been made to systematically analyze jobs. There are many duplications and overlapping in JDs both among and between organizations. The author investigated the Ministry of Expatriate Welfare and Overseas Employment (MEWOE) and found duplications in JDs between the Joint Secretary (Ministry), the Deputy Secretary (Administration), the Senior Assistant Secretary (Administration). At the field level, one can easily mark duplication of responsibilities between the Deputy Commissioner and the Police Super, the UNO and the Thana Education Officer. This is a violation of content validity as it requires linking clearly between the assigned tasks and the performance criteria.

In addition, lack of goal clarity is pervasive in BCS. Therefore, it is very difficult to causally link each official's performance to the attainment of the organizational goals and thereby sort out relatively better and valuable performers. A direct corollary of such ambiguity in the organizational goal and the JD is the use of discrete-qualitative standards against which performance is measured. In the 4th part of the ACR, five point standards have been set: Excellent, Very Good, Good, Average, and Below Average. The Manual asks the rater to consider these standards while rating each factor (Section 3.17). In the absence of quantitative JD, each standard is devoid of setting measurable benchmark. Therefore, they became mere qualitative ascriptions. Moreover, it is difficult to gauge how much difference exists between an 'Excellent' and a 'Very Good' score over a particular performance criterion.

Stage 3: Performance Rating

1.1 The Rater/ Report Initiating Officer (RIO)

Among different techniques of enforcing report initiating authority, such as the self-appraisal, the superior's appraisal, the peer appraisal, the group appraisal, the 360⁰ appraisal, BCS uses the superior rating. This system has certain advantages:

1. It is simple;
2. It is less time-consuming;
3. It strengthens the chain of command.

Among possible disadvantages, the following are conspicuous:

1. It increases the likelihood of Bossism;
2. It creates sycophantic tendency among the ratees;
3. It is highly susceptible to rater's biases

To counteract irregular bossism and to minimize the rater's biases, the ACR provides with a review by the Countersigning Officer (CSO). In the 7th part of the ACR, the CSO

can write his own judgment about the rating recorded by the rater. He has the authority to rectify any score, if he deems necessary. Although this theoretically provides a good check for reliability and perceived fairness, in practice it rarely works. BPATC study (1989: 65-66) revealed that only in 4.36% cases, the CSO rectified raters' judgment despite the presence of noticeable inconsistency in the performance measurement. A recent study on the performance appraisal (Haque, 2010) found that 81% respondent civil servants believe that the CSOs do not play their due part in rectifying rating errors done by the rater. The reasons they enumerated are:

1. CSO is not familiar with the job responsibilities of the ratee;
2. Ratees cannot know about inconsistency in their rating score as the appraisal process is confidential and therefore they are unable to appeal to the CSO with evidence.

The first reason violates evidence based rating principle of perceived fairness and the second reason reveals inadequate noticing that reduces procedural justice.

1.2 The Nature of Rating

From procedural perspective, the PA system is class-wise uniform. Two standardized forms are used: Form No. 290 (D) for Class I and II officers and Form No. 290 (B) for all Class III and IV officials. Such class-wise segregated use of the PA is likely to increase rating consistency and reliability since:

1. Class-wise functional variation is significant in BCS;
2. Class I officials who constitute only 6% of the total public manpower contribute most toward the achievement of goals of the government and shoulder almost all responsibilities. Therefore, they must be rated rigorously and differently; and
3. Such differential rating reduces complexities and inefficiency in the rating process.

The same logic though applicable for Class I posts of twenty eight cadres, the same ACR form is used for all. The arguments for using the same form are:

1. It facilitates comparison;
2. It reduces disparity in rating between various cadres; and
3. Promotion to deputy secretary and above posts which have pool of candidates from all cadres is only possible if their performances are measured on the uniform criteria.

Uniform performance criteria for all cadre posts imply that these criteria are more focused on general aspects of performance while ignoring functional aspects. Indeed, except three criteria (professional knowledge, quantity and quality of work), the other twenty two criteria do not focus on the specific job functions. This nature of rating puts unbalanced weight on the predictive validity while lowering the content validity. India and Pakistan have similar cadre system but they introduced functional appraisal using similar form for all cadres. Indian performance appraisal form includes planned activity in item two and assessment of work output in item 5 of the form. General aspects of

performance are included in item six and seven (Central Government of India, 2007). The ACR form of Pakistan provides flexible insertion of performance factors according to functional nature (Establishment Division of Pakistan, 2004).

Not all factors are equally relevant for all cadres and that is why the use of the uniform criteria might reduce validity and reliability. For example, security consciousness is highly relevant for the Police Service, but not equally relevant for the Finance and the Economic Services. On the other hand, public relation is critical for the Administrative Service but indirectly applicable for the other two services. Analytical ability is very critical for the Economic Service, though there is no scope to measure this factor in the appraisal process.

From a technical viewpoint, the current PA system in paper seems partially objective in the sense that the ACR gives an impression of the objectivity by providing numerical measurement of performance. On the other hand, the system allows unfettered discretion of the rater, which makes the system more biased to subjectivity. The following table presents a summary of the objective and subjective elements of the ACR.

Table 6: Evaluation of Objectivity of the ACR

Objectivity	Quasi-objective	Subjectivity
1. Use of numerical scale	1. Inclusion of JD	1. Absence of defined performance standard
2. Summary rating based on cumulative score	2. Countersigning	2. Absence of anchor definition of the GRS
3. Consistency between third (personality traits), fourth (performance criteria), fifth (pen picture) and sixth (recommendation for promotion, training) part of the ACR		3. Unstructured Pen Picture of ACR
		4. No definition of factors

Source: prepared by the author

In practice, however, subjectivity predominates in the appraisal process. Inconsistency is rife among the third, fourth, fifth and sixth parts of the ACR. A BPATC study (1989) revealed that in many instances where the rater gave a ‘Excellent’ grade to a ratee, he recommended not to give promotion. Significant variation (30%) in the performance score of the same employee in the consecutive years were also found, giving the impression of a low reliability in the measurement process. Apart from a large scope for the discretionary judgment, such low reliability might be the result of the absence of training and guideline for the raters in preparing the ACR. India has a specific rule, i.e. The All India Services (Performance Appraisal Report) Rules 2007, and Pakistan has a comprehensive guide to performance evaluation. Bangladesh has some guidance for filling up the ACR in the Establishment Manual, but does not have a comprehensive guideline. Interestingly, the ACR part of the Manual was not been updated since 1982 though much changes were made in the appraisal format afterwards.

1.3 Criteria of Rating

Both the personality and the performance factors are rated in the current PA. Indeed, the personality traits are emphasized by a ratio of 13:12 over the performance factors. The inherent assumptions are:

1. As public service is rule oriented, it is much necessary to evaluate the right kinds of personalities for the public sector;
2. As the class I civil servants frequently rotate from job to job of varied nature, it is more important to assess their mental aptitudes to cope up with the changes;
3. Performance in one position, in fact, does not guarantee a better performance in other positions.

But problem arises from the usage of the global traits which are not only difficult to define but also subject to varied interpretation across the raters. This violates the principle of the construct validity. For example, traits like personality, intelligence, sense of proportion cannot be defined to universal satisfaction. However, operational definition for each trait can be provided so that minimum standard can be maintained during the judgment. Apart from the rating discretion, the rater can recommend³ on the aptitude of the ratee, his honesty, training need, eligibility for promotion etc. Performance criteria of the fourth part of the ACR are also undefined leaving their judgment to the rater's discretion. Haque's study (2010) found that the civil servants came up with three dominant definitions of the 'Quantity of the tasks performed': 1) the volume of task that the supervisor assigns to the ratee, 2) the average volume of task in the concerned desk, and 3) the volume of task in response to demand.

Another source of low construct validity is the use of correlated and even synonymous performance factors which is likely to result in unnecessary inflation in rating a particular trait as well as the overall rating. The following table provides some synonymous and highly correlated factors used in the current ACR.

Table 8: Non-valid Factors

Synonymous/Interchangeable Factors	Correlated Factors
1. Devotion to duty and Sense of responsibility	1. Between reliability, sense of responsibility and supervisory capacity
2. Cooperation and Relation with peers	2. Between behavior with people and cooperation
3. Capacity to implement decision and promptness in implementing orders	3. Between intelligence and sense of judgment and proportion
	4. Between initiative and interest in work

Source: prepared by the author

1.4 Method of Rating

Graphic Rating Scale (GRS) is used to measure the personality and the performance traits in the current system. A four point scale is used, where none of the anchor is defined. This allows discretion of the rater. Moreover, the 4-point scale is unsuitable to differentiate relative worth of 35,000 cadre officials having different levels of skills. Such non-standardization of the scale not only reduces the construct validity of the appraisal, but might also encourage its political abuse by the rater sometimes becoming lenient and at other times strict. BPATC study (1989) showed that on average 25% employees were given 'Excellent' grade despite the Manual (section 3.15) categorically said that the 'Excellent' grade should not surpass 5% of the employees supervised by a rater and can only be given to those employees who make unique contribution to the service or render exemplary performance.

A pen-picture is provided in the fifth part of the ACR. However, this part has no bearing over the summary rating and therefore, became a formalistic one other than using for inserting adverse remark. The Manual asks to fill it with qualities of the ratee that the rater thinks have important impact in performing the job but did not included in the 3rd and the 4th part of the ACR.

1.5 Ratee's Right

The current appraisal system has mixed implication on the perceived fairness. It gives importance on adequate hearing. The Manual requires the Rater to provide direction about serious deviance of the ratee and give him opportunity to be corrected. If the ratee fails to correct himself, the rater will deliver him about his judgment. In such cases, the ratee has the right to appeal. In making adverse remark on the ratee, the Manual says that the rater should not accuse the ratee without concrete evidence. Section 4.8 of the Manual mentions that if the rater falsely accuses the ratee and is proved so, a departmental disciplinary action will be taken against the rater for such malafide accusation. Section 4.7 assures the ratee's right to seek proof from the rater on accusation for disloyalty and dishonesty. All these suggest that the current system gives weight on the evidence based rating, especially in case of unsatisfactory rating which is an important element of perceived fairness. According to section 4.6, if an officer gets adverse remarks for two consecutive years from the same rater, he will be withdrawn and placed under a different rater. This also gives the ratee the opportunity of adequate notice.

Stage 4: Feedback and Action

The current appraisal system has no provision of giving feedback to the official except information on the adverse remark if made any. If an employee gets 'Below average' grade in cumulative rating, or gets '1' on the GRS in the third and fourth part of the ACR, he is considered got adverse remark. Therefore, the system is partial and negative as far as feedback is concerned. The confidential nature of the system tends to generate tension among the ratees and also deprive them getting constructive direction for improving performance in the future.

The Manual provides some measures to ensure procedural justice against the adverse remarks. First, the CSO has a duty to carefully crosscheck the ground upon which ad-

verse remark was made. If the CSO disagrees with the rater, he will rectify the performance score and such rectified score would be final. If the CSO agrees with the rater, the concerned Ministry will inform the employee about the adverse remark and seek application for redress. Based on defensive argument provided by the rater and the ratee, the concerned ministry under which the ratee belongs to will make the final decision.

CONCLUSIONS

The foregoing discussion leads us make the following observations about the current PA system of BCS:

1. Though predictive validity of the current system theoretically seems satisfactory, it considerably lacks content and construct validity. Therefore, administrative decisions like promotion, placement, salary increase, disciplinary action etc. based on the current rating will likely to violate distributional and procedural justice.
2. Reliability in the appraisal process is also low due to non-standardization of the performance factors and the rating scale. The absence of training and proper guideline on the appraisal process make the reliability even worse. This provides the scope for the political abuse of the ACR.
3. Perceived fairness in post-rating phase seems sufficiently high. At the same time, pre-rating phase has low perceived fairness due to confidential nature of the appraisal process and the high discretionary power of the rater.
4. Though multiple objectives were set, the current ACR have no relevance to career planning and performance improvement since the ACR does not focus on job performance and does not allow performance related feedback except the adverse remark.

From the foregoing observation we can reasonably conclude that the efficacy of the current PA system of BCS is questionable. As we told earlier, a faulty appraisal system might be counterproductive, the current PA system might create more problems than solutions. Extensive empirical study is needed to measure actual efficacy of the present system and identify ways of reforms.

Mohammad Ashraful Haque is Lecturer in Public Administration at Jahangirnagar University, Bangladesh. He is currently studying public policy management at the Heinz College of Carnegie Mellon University, Pittsburgh, PA, USA. Email: haqash.01@gmail.com.

NOTES

1. Gresham's Law of Money Circulation says that if commodity value of a currency exceeds the face value, it is driven out of market. Similarly, if the performance value of an official exceeds the norm value set by the appraiser, he/she is driven out of the core civil service group. Therefore, appraisal system becomes fallacious-what was chosen to reward the merit eventually ends up in punishing it.
2. Deputy secretary position is a midlevel managerial position but entry level policy position in BCS. Promotion to this level and onward is highly regarded in Bangladesh and has been a source of inter-cadre rivalry.
3. In the 6th part of the ACR, the rater writes recommendations at his discretion.

REFERENCES

- Administrative and Service Reorganization Committee (ASRC), 1973. *ASRC Report*. Government of Bangladesh.
- Anderson, J. r. 2002. "Measuring Human Capital: Performance Appraisal Effectiveness", Paper presented in Midwest Academy of Management Conference, University of Kentucky.
- Armstrong, M. and Baron, A., 1998. *Performance Management: The New Realities*, Institute of Personnel and Development: London.
- BPATC, 1989. *Performance Appraisal System of Class I officers of Bangladesh Civil Service*. Savar: Bangladesh Public Administration Training Center (BPATC).
- Bretz, R.D., Milkovich, G. T., and Read, W., 1992. "The Current State of Performance Appraisal Research and Practice: Concerns, Directions, and Implications". *Journal of Management*, 18(2): 321-352.
- Burgess, R., Philippe, A. Redding, S., and Zilibotti, F. 2004. "The Unequal Effects of Liberalization: Theory and Evidence from India." Working paper, London School of Economics.
- Byrne, Z.S. and Cropanzano, R. 2001. "The History Of Organizational Justice: The Founders Speak". In Cropanzano (Ed.), *Justice In The Workplace: From Theory To Practice*, 3-26, Mahwah, NJ, Lawrence Erlbaum Associates, Inc.
- Cascio, W. F., 1982. *Applied Psychology in Personnel Management*. 2nd Edition. Reston VA: Reston Publishing Company.
- Central Government of India, 2007. *All India Services (Performance Appraisal Report) Rules, 2007*.
- Cleveland, J.N. and Murphy, K.R. 1992. "Analyzing Performance Appraisal as a Goal-Directed Behavior". *Research In Personnel and Human Resources Management*, 10, 121-185.

- Cleveland, J. N., & Murphy, K.R., 1995. *Understanding Performance Appraisal: Social, Organizational, and Individual Differences*. Thousand Oaks, CA: Sage.
- Cronbach, L., 1990 *Essentials of Psychological Testing*. New York: Harper and Row.
- Daley, D. M. 1993. "Performance Appraisal As an Aid in Personnel Decisions," *American Review of Public Administration*, 23, 201-214.
- Demmke, C. 2007. *Study for the 48th meeting of the Directors-General of the Public Service of the Member States of the European Union*. Maastricht: European Institute of Public Administration.
- DeNisi, A., Cafferty, T., and Meglino, B., 1984. "A Cognitive Model of The Performance Appraisal Process: a Model and Research Propositions". *Organizational Behavior and Human Performance* 21(3):358-367.
- Dewberry, C. 2001. "Performance Disparities Between Whites and Ethnic Minorities: Real Differences or Assessment Bias?" *Journal of Occupational and Organizational Psychology*, 74, 659-673.
- Establishment Division of Pakistan, 2004. *A Guide to Performance Evaluation*. Pakistan Public Administration Research Center.
- Folger, R., Konovsky, M. A. and Cropanzano, R., 1992. "A Due Process Metaphor for Performance Appraisal", *Research in Organizational Behavior*, 14: 129-177.
- Haque, Mohammad Ashraful, 2010. "From Legality to Result: Revisiting Performance Appraisal System of Bangladesh Civil Service". Paper presented at the annual Social Sciences Deans Conference, Jahangirnagar University, Savar: Bangladesh.
- Hossain, A., 2004. *Civil Service Management*. Unpublished Doctoral Thesis, University of Dhaka: Bangladesh.
- Ivancevich, J.M., 2001. *Human resource management*. 8th Edition. New York: McGraw-Hill Irwin.
- Lam, S.S.K., and Schaubroeck, 1999. "Total Quality Management and Performance Appraisal: An Experimental Study of Process Versus Results and Group Versus Individual Approaches". *Journal of Organizational Behavior*, 20: 445-457.
- Landy, F. and Farr, J., 1983. *The Measurement of Work Performance*, NY: Academic Press.
- Latham, G. P., and Wexley, K. N., 1981. *Increasing Productivity Through Performance Appraisal*. Reading, MA: Addison-Wesley.
- Lawler, E. E., Mohrman, A. M., and Resnick, S. M., 1984. "Performance Appraisal Revisited". *Organizational Dynamics*, 13 (1): 20-35.
- McDonald, M., 2005. "Validity, Data Sources". *Encyclopedia of Social Measurement*. K. Kempf-Leonard, Elsevier, 3: 939-948.
- McGregor, D. 1987. *The human side of enterprise*. Middlesex: Penguin Books.

- Milkovich, G. T. and Wigdor, A. K., 1991. *Pay for Performance: Evaluating Performance Appraisal and Merit Pay*. National Research Council: USA.
- Ministry of Public Administration, 2010. *Establishment Manual*. Government of Bangladesh.
- Misra, B. B. 1970. *The Administrative History of India: 1834-1947*. London: Oxford University Press.
- Murphy, K., and Cleveland, J. 1991 *Performance Appraisal: An Organizational Perspective*. Boston: Allyn and Bacon.
- Newstrom, J.W. & Davis, K. 1993. *Organisational behaviour: human behaviour at work*. New York: McGraw-Hill.
- Organ, D.W. (1997). "Organizational citizenship behavior: It's construct cleanup time", *Human Performance*, 10: 85-97.
- Pay and Services Commission, 1977. *Pay and Service Commission Report*. Government of Bangladesh.
- Pfeffer, J., 1981. *Power in Organization*. Marshfield, MA: Pittman.
- Prentice, R. A. and Keohler, J., 2003. "A Normality Bias in Legal Decision Making", *Cornell Law Review*, 88.
- Scott, S.G., and Einstein, W.O., 2001. "Strategic Performance Appraisal in Team-based Organizations: One Size Does Not Fit All". *Academy of Management Executive*, 15(2),107-116.
- Siddiquee, N.A. 2006. "Public management reform in Malaysia", *International Journal of Public Sector Management Journal*, 19(4), 339-358.
- Skarlicki, D. P. and Folger, R., 1997. "Retaliation in the Workplace: The Roles of Distributive, Procedural and Interactional Justice". *Journal of Applied Psychology*, 82: 434-443.
- Snell, s. and Bohlander, G. 2007. *Human Resource Management*. 3rd Edition. Delhi: Cengage Learning.
- Somerick, N. M., 1993. "Strategies for Improving Employee Relations By Using Performance Appraisals More Effectively". *Public Relation Quarterly*. 38 (3), 37-39, 1993.
- Spangenberg, H.H., 1990. *Assessing Managerial Competence*. Cape Town: Juta.
- Spencer, L.M. & Spencer, S.M., 1993. *Competence at Work*. New York: Wiley.
- Tziner, A., Joanis, C., & Murphy, K., 2000. "A Comparison of Three Methods of Performance Appraisal with Regard to goal Properties, Goal Perceptions, and Ratee Satisfaction". *Group and Organization Management*, 25:175-190.

About IPMR

IPMR The International Public Management Review (IPMR) is the electronic journal of the International Public Management Network (IPMN). All work published in IPMR is double blind reviewed according to standard academic journal procedures.

The purpose of the International Public Management Review is to publish manuscripts reporting original, creative research in the field of public management. Theoretical, empirical and applied work including case studies of individual nations and governments, and comparative studies are given equal weight for publication consideration.

IPMN The mission of the International Public Management Network is to provide a forum for sharing ideas, concepts and results of research and practice in the field of public management, and to stimulate critical thinking about alternative approaches to problem solving and decision making in the public sector.

IPMN includes over 1300 members representing about one hundred different countries, both practitioners and scholars, working in all aspects of public management. IPMN is a voluntary non-profit network and membership is free.

ISSN 1662-1387