Public Personnel Management
2014, Vol. 43(4) 387–414
© The Author(s) 2014
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0091026014530675
ppm.sagepub.com
SAGE

Motivation of Public Managers as Raters in Performance Appraisal: Developing a Model of Rater Motivation

Seejeen Park[1]

Abstract

The motivation of public managers to provide accurate personnel performance appraisals in the public sector is an unexplored area of study in public administration. This study provides preliminary steps in developing a rater motivation model (RMM) for the public administration field. The RMM is built using three primary determinants of rater motivation (felt accountability, incentive structures, and public service motivation). In addition, this study examines the variations in raters' perceived rating accuracy depending on the existence of a reward system for accurate appraisers and an organizational forced distribution rating system. The sample consists of public officials working in various levels of the U.S. and Korean governments. This study uses confirmatory factor analysis to construct measures of rater motivation and multiple regression analysis to observe the relationship between the determinants and the rater's perceived performance rating accuracy. The findings suggest that raters are significantly affected by all three determinants of rater motivation as well as by the presence of rewards for accurate appraisal and a forced distribution rating system. Implications for both the public administration literature and practice are suggested.

Keywords

rater motivation, public service motivation, performance appraisal

Introduction

In human resource management, the supervisor, who stands between employees and management, has the crucial task of helping the organization make the best use of its

[1]Myongji University, Seoul, Republic of Korea

Corresponding Author:
Seejeen Park, Public Administration Department, Myongji University, 34 Geobukgol-ro, Seodaemun-gu, Main building 5th floor S1529, Seoul, 120728, Republic of Korea.
Email: sjpnpm@gmail.com

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

core asset: employees (Berman, Bowman, West, & Van Wart, 2006). Performance appraisal is a tool used by supervisors to manage employees in a way that will induce them to work effectively for an organization by providing rewards and recognition (e.g., promotion, pay increases) or sanction (e.g., delaying promotions, withholding bonuses, and pay increases) based on each employee's performance. Therefore, performance appraisal, whether conducted in government or private organizations, is a standard and important task of supervisors.

Although supervisors are responsible for conducting performance appraisals, the motivation of public managers, as supervisors, in the performance appraisal context remains an unexplored area of study within the field of public administration. The majority of the literature concentrates on explaining the success or failure of the appraisal system from the subordinate's perspective rather than from the supervisor's perspective. Such research can be divided into two streams.

The first stream of this research examines the ways in which the pay-for-performance (PFP) system can be implemented and managed. Since the enactment of the Civil Service Reform Act (CSRA) of 1978, contemporary studies concerning performance appraisal have been strongly tied to implementation of the PFP system (see Perry, Engbers, & Jun, 2009, for a comprehensive review). To name a few areas, scholars sought to address problems in the PFP implementation process by examining the validity of the performance appraisals in differentiating between employees with different performance levels (Pearce & Perry, 1983; Perry, Petrakis, & Miller, 1989), and employee perceptions and attitudes toward PFP appraisal system implementation (Perry et al., 2009; Underhill & Oman, 2007). The findings suggest that PFP is ineffective, yet continues to be used in the public sector (Ingraham, 1993; Park & Berry, 2012).

The second stream of research investigates how the effectiveness of the appraisal system can be improved. For instance, popular areas of research include the role of employee participation and acceptance of the appraisal system (Roberts, 1994, 2003), perceived stress and appraisal discomfort (Gbadamosi & Ross, 2012; Smith, Harrington, & Houghton, 2000), agreement between managers and subordinates on appraisal (Longenecker & Nykodym, 1996), and rater training (Duncan, Atkins, Fletcher, & Stillman, 2005).

In both streams of research, scholars often used survey instruments to ask evaluated employees about their perceptions regarding fairness, links between performance and pay, and satisfaction with the appraisal. Despite a number of studies (e.g., Miller & Thornton, 2006; Selden & Sowa, 2011) proposed rater's perspective on the appraisal system, the effectiveness of the appraisal system was mainly assessed based on the subordinate's view. Consequently, individual motivation in appraisal research was operationalized from the perspective of the subordinates. Although raters (supervisors) are responsible for determining the final outcome of the performance appraisal process in the form of performance ratings, the motivation of raters in public organizations has rarely been considered in empirical research.

Rater motivation poses a significant issue because the ways in which raters respond to various motivating or demotivating forces while conducting performance appraisals

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

may significantly affect the accuracy of those appraisals. An inaccurate appraisal may result in insufficient discrimination between employees with different levels of performance and reduce the utility of appraisals in making administrative decisions regarding promotions and financial incentives (Blume, Baldwin, & Rubin, 2009; Decotiis & Petit, 1978).

In the current study, rater motivation is defined as the motivation of performance raters to be more or less accurate in making appraisals. The definition is based on Murphy and Cleveland's (1991, 1995) argument that raters are not always motivated to evaluate subordinates accurately and are likely affected by various motivating forces that incentivize raters to intentionally distort their rating accuracy. The rating accuracy affected by rater motivation in this study is introduced as perceived performance rating accuracy, which represents the level of accuracy that a rater perceives his or her final performance rating distributed to the subordinates reflects.

This study provides preliminary steps in developing a rater motivation model (RMM) for the public administration field. The study focuses on discovering the motivational bases of public managers in the public performance appraisal context and imports relevant theories (e.g., rater motivation) from the business sector for this purpose. It has long been assumed that public and private organizations share similar management functions, and analogous management practices can be used in both settings (Rainey, 2003; Rainey & Bozeman, 2000).

The diffusion of PFP from the private to the public sector by the CSRA of 1978 (Park & Berry, 2012), the total quality management (TQM) approach in performance appraisal (Bowman, 1994), and the George W. Bush Administration's President's Management Agenda of "strategic management of human capital" provide examples of cases in which business human resource management practices were adopted in the public sector.

The current study constructs RMM pertinent to the public sector by adding public service motivation (PSM) to the model. RMM could contribute to extending the theories related to PFP and PSM by addressing the important appraisal function from the supervisor's viewpoint and testing the influence of PSM in the appraisal process.

The RMM

A significant portion of previous rater accuracy research has centered around topics such as rating format modification, the cognitive process of raters, and rater training (Arvey & Murphy, 1998; Roch, McNall, & Caputo, 2011). Such approaches assume that raters possess an inherent motivation to rate accurately, and the accuracy of the appraisal can therefore be improved by enhancing the rating instrument or rater cognition.

More recently, researchers began to address the effects of rater motives on performance appraisal. The basic premise was that inaccurate appraisals may be intentional rather than a result of the raters' inability to rate accurately (Murphy & Cleveland, 1991, 1995; Wang, Wong, & Kwong, 2010). Several examples of major topics in the current rater motivation literature are the rater's goal-directed perspectives for

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

appraisal (Wang et al., 2010), the impact of non-performance information on a rater's rating behavior (Spence & Keeping, 2010), the effect of impression management on performance ratings (Wayne & Liden, 1995), the politics of performance appraisal (Longenecker, Sims, & Gioia, 1987), and the relationship between felt accountability and appraisal (Mero & Motowidlo, 1995).

Nevertheless, most of the literature has examined rater motivation by investigating the relationship between a single aspect of rater motives (e.g., social context or rater goals) and appraisal accuracy. Only a small number of studies in the business literature (e.g., Harris, 1994; Murphy & Cleveland, 1991, 1995) have proposed integrated models of rater motivation. However, these models are primarily conceptual and have not been extensively tested in organizational settings. Consequently, gaps and limitations exist with regard to explaining the RMM in a comprehensive manner. Furthermore, the previous works have been conducted mostly in the private sector rather than in the public sector and have not accounted for the potential effects of PSM on appraisal performance.

The present study seeks to construct a comprehensive RMM that incorporates the extrinsic and intrinsic motivation of the raters (Harris, 1994; Murphy & Cleveland, 1995), the relationship with the rater's supervisor and subordinates, as represented by felt accountability (Mero & Motowidlo, 1995), and PSM. By doing so, the current study presents the various motivational factors in public organization appraisal settings and addresses the limitations of the extant literature that focused on the unidimensional aspect of rater motivation. The raters' motivation for their individual benefit and satisfaction, consideration for social context, and the level of PSM are assumed to affect the appraisal outcome.

As shown in Figure 1, in RMM, raters are assumed to feel accountable to their supervisors and subordinates involved in the appraisal process (felt accountability), perceive incentives and disincentives in making accurate appraisals (incentive structure), and be affected by PSM when making appraisal decisions. This research is mainly interested in how the determinants of rater motivation affect the outcome of raters' perceived performance rating accuracy. In addition, the study examines the variations in perceived rating accuracy depending on the existence of an organizationally established reward system for accurate appraisers and a forced distribution rating system. In the current study, perceived performance rating accuracy represents the raters' perception of the closeness of the rating to the true reflection of the employee's performance and how much the raters are satisfied with the ratings they made. Raters' levels of satisfaction with the performance appraisal experience, their own self-assessment of their appraisal accuracy, and their perception of how the final rating reflected their willingness to rate accurately were all included in forming the overall perceived rating accuracy. Accordingly, factors of rater motivation will either positively or negatively affect the rater's level of perceived rating accuracy.

The perceived performance rating accuracy is distinguished from direct or objective measurement of appraisal accuracy. In general, direct measurement of judgmental rating accuracy is conducted by assessing the degree by which the ratings match the true score that represents an unbiased rating of the rater while in an optimal setting

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

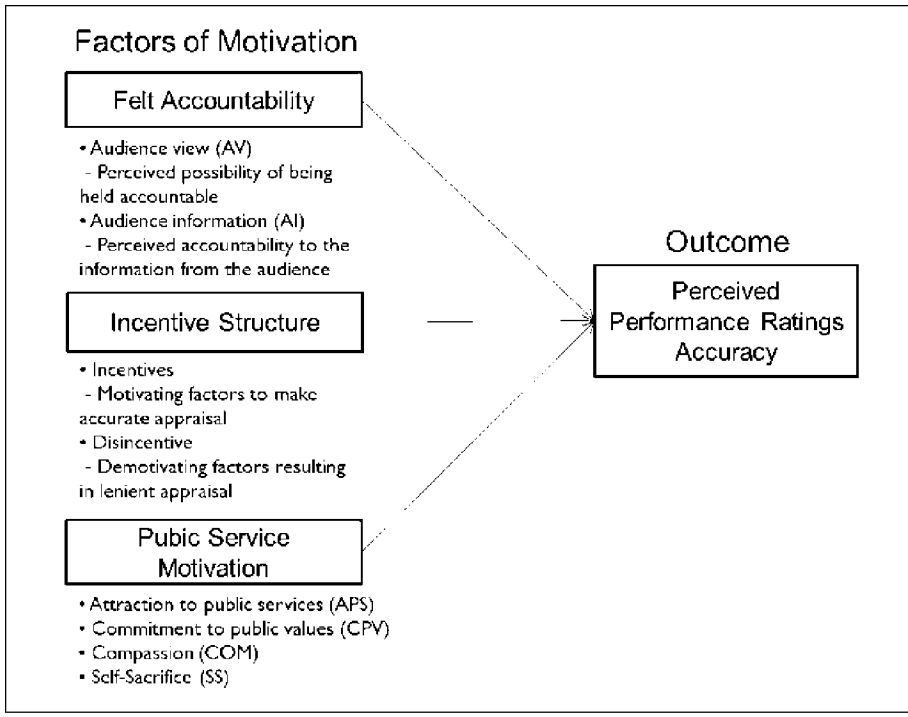

Figure 1. Model of RMM.
Note. RMM = rater motivation model.

(Murphy & Cleveland, 1991). However, such a direct measurement cannot be used to explain the rater's self-evaluation of the rating accuracy because the discrepancy between the utmost possible accurate rating and the rater's actual rating accuracy can only be perceived by the rater.

Hypotheses

Felt Accountability

Felt accountability (or rater accountability) can be described as a "perceived potential to be evaluated by someone and be held responsible for one's decisions or behaviors" (Frink & Ferris, 1998; Levy & Williams, 2004, p. 887). A number of scholars have posited that in organizational settings, when decision makers are aware of their audience's views, they tend to make decisions that are acceptable to that audience (Klimoski & Inks, 1990; Mero, Guidice, & Brownlee, 2007; Tetlock, 1985). These audiences include various stakeholders, such as supervisors, managers, subordinates, customers, and suppliers (Mero et al., 2007). In the performance appraisal context, considering the audiences' views when making appraisal decisions can be considered

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

as a motivating factor for the rater (Mero & Motowidlo, 1995). In the present study, two types of audiences—the subordinate being evaluated and the supervisor of the rater—will be used to reflect rater felt accountability.

This study tests and builds on the work of Mero and Motowidlo (1995), and explores new aspects of rater accountability. The reliability of the empirical results (felt accountability as a motivating factor) developed in laboratory settings used in previous research is tested by replicating the research in actual organizations with data from managers who evaluate subordinates. The following two aspects of felt accountability in the appraisal context are then introduced: (a) audience view and (b) audience information.

Audience view is a representation of the degree of the raters' perceived possibility of being held accountable to the appraisal audiences. Audience information reflects the level of a rater's perceived accountability to information from the appraisal audience (e.g., expected rating from subordinate, supervisor's indication of good performance). It is assumed that a higher expectation to be evaluated and more information from the audience results in a higher level of felt accountability. Consequently, as raters' overall level of felt accountability increases, the perceived performance rating accuracy should improve because the raters are likely to believe that they have made appraisal decisions that are acceptable to the audience and become satisfied with the ratings they delivered.

Hypothesis 1: The level of a rater's felt accountability and perceived performance rating accuracy is positively related.

Incentive Structure

The conceptual models of Murphy and Cleveland (1995) and Harris (1994) are used as a foundation for constructing the incentive structure, which includes both incentives and disincentives. In addition, Landy and Farr's (1980) argument that raters are indifferent to different types of rating formats is tested by arguing that inconvenience in rating format may be a disincentive.

Murphy and Cleveland (1991) posited that raters are motivated to rate accurately based on the existence of rewards for accurate ratings and the probability of receiving those rewards. However, raters expect to experience negative consequences of accurate and low ratings and are motivated to distort ratings to avoid such consequences. Therefore, according to Murphy and Cleveland (1995), rating behavior is an integrated outcome of two contradictory motivations. In the current study's model, incentives and disincentives reflect the two contrary motivations.

Harris (1994) built a "model of rater motivation" (p. 739), in which situational (e.g., organizational human resource management strategy) and personal variables (e.g., self-efficacy) affect motivational factors, such as rewards, negative consequences, and impression management, that generate the final outcome of performance appraisal behaviors. Harris divided the types of rewards in appraisal into two categories: direct rewards and indirect rewards.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Landy and Farr's (1980) proposition that performance raters may prefer various rating formats, but that their preferences have little impact on actual rating behavior shifted the direction of performance appraisal research in the 1980s from rating instruments to the assertion that the appraisal process matters (Ferris, Judge, & Rowland, 1994; Murphy & Cleveland, 1995). Landy and Farr's (1980) argument implies that rating format is neither a motivating nor a demotivating factor for the raters because raters are indifferent to different types of rating formats when making their appraisals. Departing from Landy and Farr (1980), the present model assumes that if raters view their organization's rating format as burdensome or time-consuming, they may become dissatisfied, will not fully use the contents in the format, and will produce less accurate appraisals.

Incentives. Incentives can be indicators of why raters invest time and resources in performance appraisal. Rewards can take the form of both direct (formal and informal reward structure) and indirect rewards (Harris, 1994). A first factor to consider is whether the organization has a formal reward system for accurate appraisers. Although Murphy and Cleveland (1991) indicated that such a reward system rarely exists, it is possible that organizations reward accurate appraisals, which may positively affect appraisal accuracy.

A second factor is the informal incentive to comply with the organization's culture of promoting accurate performance appraisals. Management practices, including appraisal, are considered to be a type of organizational culture, and these practices play the role of a control mechanism that accepts or disapproves of certain patterns of behavior in organizations (Martin & Siehl, 1983). Therefore, raters are motivated to make a more accurate appraisal in an organizational culture that promotes accurate appraisal.

A third factor is the indirect reward that stems from making an accurate appraisal, which can be of two types. The first type of reward is the extrinsic reward of a pay increase or promotion of the rater based on unit performance improvement (Harris, 1994). In Harris' view, the raters may believe that unit performance can be improved by providing accurate appraisals with feedback to the members of the unit (subordinates). The second type of reward is the intrinsic reward of the rater's inherent satisfaction with being involved in performance appraisal (Harris, 1994). Harris (1994) posited that the "increased esteem and recognition from subordinates or supervisors" (p. 740) after the rater provides helpful feedback is a satisfier.

Disincentives. Both extrinsic and intrinsic disincentives can discourage raters from conducting accurate appraisals. Regarding extrinsic disincentives, with the expectation of increasing unit performance, raters will be less accurate if they believe that inflated ratings will increase a subordinate's future performance (Murphy & Cleveland, 1991). Another extrinsic disincentive is the organizational culture of promoting lenient appraisal. In organizations where the norm is to assign high ratings, the raters who act against the norm may confront disapproval from peer managers (Murphy & Cleveland, 1995). Accordingly, raters are motivated to make inflated appraisal under an organizational culture that encourages lenient appraisal.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For intrinsic disincentives, although raters may be satisfied by providing helpful feedback to subordinates based on an accurate appraisal, they can also be dissatisfied by delivering negative feedback (e.g., informing employees that their performance is below average) and may inflate ratings to avoid negative reactions from the ratees (Gbadamosi & Ross, 2012; Murphy & Cleveland, 1995).

Another intrinsic disincentive is the rater's dissatisfaction with the appraisal formats used in their organization. If raters perceive their appraisal format as burdensome or time-consuming, they may make a less accurate appraisal by not fully utilizing the content in the format.

Incentive structure. Both the incentives and disincentives are included in the incentive structure in the current model. As discussed above, most of the disincentives are developed from Murphy and Cleveland's (1991, 1995) theory, which mainly concentrated on why raters would distort ratings. Interestingly, Harris (1994) defined the same factors for the incentives, such as giving feedback to subordinates, as extrinsic and intrinsic incentives. For that reason, the current study combines the work of Harris (1994), and Murphy and Cleveland (1991, 1995), and assesses whether the same determinants of motivation can act as both incentives and disincentives. It is expected that if the same factors can be used as incentives and disincentives, the relationship between the two different types of motivators will be negative. The specific formal incentive structure system used by an organization is tested separately because it can be easily identified in an objective manner and does not involve any individual perception of the raters. All other components (discussed above) are combined to form the incentive structure. The incentive structure represents the level of a rater's incentive to be accurate in their appraisal. The higher the raters' incentive level, the more they will perceive that their appraisal decisions are precise.

Hypothesis 2: Raters will show a higher perceived performance rating accuracy when there are direct formal incentives than when there are no direct incentives.
Hypothesis 3: A rater's level of incentive structure and perceived performance rating accuracy is positively related.

PSM

PSM research, based on the belief that "unique motives are found among public servants that are different from those of their private sector counterparts" and that this affects the ways in which public servants act, has been a major area of public administration research for more than 20 years (Perry, Hondeghem, & Wise, 2010, p. 681; Wright & Grant, 2010). Previous research has studied the impact of PSM on organizational behavior, quality of work, and other topics (see Perry et al., 2010, for a comprehensive review). The current study examines the association between PSM and job performance.

In one of the first works of the PSM literature, Perry and Wise (1990) proposed that "in public organizations, public service motivation is positively related to individual performance" (p. 370). The proposition led to the argument that "employees with high

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PSM are motivated to perform more effectively" (Wright & Grant, 2010, p. 694). Since then, a number of studies have found positive relationships between PSM and individual performance (Naff & Crum, 1999; Vandenabeele, 2009), whereas other studies have revealed mixed results (e.g., Alonso & Lewis, 2001, Frank & Lewis, 2004) or no direct significant influence (Bright, 2007).

The current study assumes that the positive relationship between PSM and work performance can be applied to the appraisal setting because performance appraisal is one type of task of the rater. Work performance in the appraisal context is dependent on rating accuracy, meaning that higher appraisal accuracy equals higher appraisal performance. Therefore, PSM and appraisal performance are assumed to be positively associated as raters with higher PSM are likely to be more motivated to perform their tasks than raters with lower PSM. In sum, this study expects that a higher level of PSM will increase raters' perceived rating accuracy.

Hypothesis 4: The level of a rater's PSM and perceived performance rating accuracy is positively related.

Forced Distribution Rating System

Organizations that implement a forced distribution rating system require raters to discriminate between high and low performers by assigning fixed portions in rating categories (Blume et al., 2009). Several scholars provide a number of relevant implications regarding the impact of a forced distribution system on appraisal. For example, according to Blume and colleagues (2009), individuals (undergraduate students) become more favorable toward the forced distribution system when the system does not punish low performers often. In addition, Schleicher, Bull, and Green (2009) found that individuals (MBA students) perceived that the forced distribution rating system was "more difficult and less fair than a more traditional rating format" that does not assign forced proportions of ratings (Schleicher et al., 2009, p. 899).

The current study conducts a direct comparison of raters depending on the existence of a forced distribution rating system in their organization. The assumption is that unless the organization's expectation for the rating distribution and the rater's opinion of what rating each subordinate deserves match exactly, the perceived accuracy of the rating is compromised. Therefore, this study expects that the perceived accuracy of an appraisal will be higher in an organization with no forced rating distribution system.

Hypothesis 5: The relationship between the existence of a forced distribution rating system and a rater's perceived performance rating accuracy is negative.

Research Design

The empirical analysis of this study involves three stages. First, surveys were administered to public officials in the U.S. (state and local) and Korea (central, provincial, local) governments. Confirmatory factor analysis (CFA) was then conducted with the

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

collected data to develop measures of rater motivation. Finally, a cross-sectional multiple regression analysis (MRA) was executed to observe the relationship between the motivating factors and the final outcome, the perceived performance rating accuracy. The unit of analysis for this research was individuals in public organizations.

Sample

The sample consisted of public officials in public agencies with supervisory appraisal experience. The ideal population of the sample may be all public officials in the United States and Korea. However, conducting random sampling of all officials from both countries was not feasible due to limited resources and difficulty in attaining access to all agencies. Therefore, convenience sampling was used because this study selected a limited number of agencies in the United States and Korea. In an attempt to increase the representativeness of the data, surveys were administered to officials in various levels of government. The survey population in the United States consisted of state and local government managers enrolled in a certified public manager program (CPM) in a large southern U.S. state. The population in Korea consisted of officials in central, provincial and local governments. The sample is described in the survey procedure section.

Survey Procedure

Surveys were distributed to U.S. and South Korea civil servants. The target group for data collection was public officials with performance appraisal experience. Organizations in both countries have appraisal systems in which the final outcome of the process is a performance rating. Before the actual distribution, the survey's content validity was checked by several public administration professors in both the United States and Korea (including translations from English to Korean), the director of a CPM in a large southern state university (United States), and public officials (Korea). Survey anonymity was ensured by omitting personal information that could be used to identify the respondents. To increase the number of respondents, on-site surveys were used rather than online or mailed surveys.

For the United States, surveys were distributed to public officials enrolled in the CPM conducted by a large southern state university. The CPM program is certified by the National Certified Public Manager Consortium and provides training and development programs for public sector managers from state and local governments. The time frame for distribution was December 2012 to February 2013. The overall response rate was 51.3% (182/355). Of the respondents, 114 were appraisers.

In Korea, 600 questionnaires were sent out between December 2012 and January 2013 to four central government agencies (Ministry of Gender, Equity & Family, Ministry of Foreign Affairs and Trade, Ministry of Government Legislation, and Ministry of Public Administration and Security) and three provincial or local government-level offices (SeodaemunGu Office, Yongin City Office, and Goyang City Office). The response rate was approximately 70% (422/600). Of the respondents, 122 respondents were appraisers. Table 1 provides information regarding the descriptive information of the surveys in both countries.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1. Description of Survey Respondents.

Variable	Category		United States (n = 114)	Korea (n = 122)	All (N = 236)
Gender	Men	n (%)	51 (44.7)	92 (75.4)	143 (60.6)
	Women		63 (55.3)	30 (24.6)	93 (39.4)
Education	High school/general equivalency diploma	n (%)	8 (7.0)	5 (4.1)	13 (5.5)
	Some college		27 (23.7)	5 (4.1)	32 (13.6)
	Bachelor's degree or equivalent		40 (35.1)	66 (54.1)	106 (44.9)
	Some graduate school		11 (9.6)	7 (5.7)	18 (7.6)
	Master's degree or higher		28 (24.6)	39 (32.0)	67 (28.4)
Tenure (years)	Average		20.01	12.8	16.53
	SD		9.02	6.83	8.8

Measures

All measures (provided in Appendix A) were constructed from the survey data either by using questions and scales from previous research or by developing new scales. Felt accountability was measured by modifying or adding items to the eight-item felt accountability measure developed by Hochwarter, Kacmar, and Ferris (2003). The modified and developed items are listed in Appendix A, Table A1. For the incentive structure, there were no directly applicable measures in the literature. Therefore, the author developed items for this study (Appendix A, Table A2).

To measure PSM, this study used a mixture of recently developed measures. Recently, modifying Perry's (1996) 24-item measure of PSM, Kim and his colleagues' (2013) measures were tested in 12 countries, including the United States and Korea. The authors suggested modifying their finalized 16-item measure of PSM to apply it to Korea and posited that the measures were appropriate for use in the United States. Although Kim and his colleagues (2013) found the need to modify the 16-item measure for Korea, Kim's (2011) work on Korean firefighters suggested a four-dimension, 12-item measure as a valid measure of PSM.

Consequently, the present study combines these two PSM measures to create a scale of PSM applicable to both the United States and Korea. The four dimensions of PSM were concurrent with Kim and his colleagues' (2013) dimensions as follows: attraction to public services (APS), commitment to public values (CPV), compassion (COM), and self-sacrifice (SS). The 23 PSM items and their descriptive information are presented in Appendix A (Table A3).

With no prior measures developed in the literature, perceived performance rating accuracy was measured by a three-item index that was developed to measure the degree of accuracy that the raters felt they had achieved in their appraisal. The measurement model showed strong Cronbach's Alpha (α = .819), and factor analysis confirmed all of the items loading on a single factor. A composite index of perceived

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2. Availability of Reward and Forced Distribution Rating System (N = 236).

Item	Question	% of availability	
1	My organization has a system of rewarding accurate performance appraisers (e.g., best supervisor award, including rating accuracy as one category of a supervisor's own performance appraisal).	United States	9.6
		Korea	37.7
2	My organization has a fixed-performance rating distribution system, such that there are assigned proportions for each given rating.	United States	73.7
		Korea	88.5
Country	Examples of rewards for accurate appraisers		
United States	Monetary reward, awards, plaques, good evaluation in the rater's appraisal (accuracy of appraisal is included in task performance evaluation) (common in both countries)	—	
Korea		Promotion, salary increase, transfer to preferable or important position within the organization, training opportunity, trip to abroad package, vacation (days) (only in Korea)	

performance rating accuracy was created by adding the three items into a single index, as provided in Appendix A (Table A4).

The survey respondents were asked whether their organizations had a reward system for accurate appraisers or a forced distribution rating system. Both measures were coded into dichotomous variables (0 = yes, 1 = no). The description of the availability of the systems and examples of rewards for accurate appraisers is presented in Table 2. Table 1 describes the control variables, which included gender, education, and years of service in their current organization.

Analysis

Before executing CFA, the data from the U.S. and Korean appraisers were integrated to create the data set used in the current analysis. There were 236 total observations, and there were no significant outliers or miscoding in the data. Data screening for missing values revealed that 1.24% of the data were missing. Using the PRELIS 2.80 expectation maximization (EM) algorithm method, missing values were filled using multiple imputation techniques.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CFA

CFA was conducted using AMOS 21 with the maximum likelihood estimation. CFA is a proper tool because this study has hypotheses about the expected specified relationship between the indicators (factors) and the unobserved (latent) variables. The relationship between the indicators and the latent variables is defined as a reflective relationship. Such a definition follows the management literature in which the measures are "usually treated as reflective, meaning they are observed manifestations of unobserved latent constructs" (Diamantopoulos & Siguaw, 2006, Williams, Vandenberg, & Edwards, 2009, p. 551). Accordingly, in this study, the measures (items) of each determinant of motivation are reflective indicators of each latent variable (felt accountability, incentive structure, PSM) that were used to develop measures of the determinants of rater motivation.

For the evaluation of the model fit, the goodness of fit index (GFI), the comparative fit index (CFI), the normative fit index (NFI), and the root mean square error of approximation (RMSEA) are reported. In general, the criteria for an acceptable fit of each index were as follows: (a) GFI (>.90), (b) CFI (>.90), (c) NFI (>.90), and (d) RMSEA (<.08; Kim, 2009; Kline, 2005; K. Yang, 2009). The current study followed the advice of several scholars who regard the chi-square/*df* ratio as an improper indicator of model fit for CFA and do not report it (Hair, Anderson, Tatham, & Black, 1998; Hatcher, 1998; K. Yang, 2009).

In the present study, three separate CFA models were conducted to test three measures: felt accountability, incentive structure, and PSM. After running CFA on the initial measurement model, the model fit was assessed and the model was re-specified until it attained an acceptable fit. Standardized factor loadings (parameter estimates) and modification indices (MI) were used as criteria to modify the measurement model. After the measures were finalized, the items in each model were summed to create composite indexes (listed in Appendix B, Table B4) for MRA. To conserve space in the current article, detailed information regarding the statistical and technical issues of CFA used in this study's analysis is provided in a previous study by the author (Park, 2013).

Felt accountability. The one-factor, four-item model showed a good fit in general (GFI = .989, NFI = .962, CFI = .976, and RMSEA = .078). All factor loadings of the items were significant at the $p < .001$ level with a range of .46 to .695. The results show (Appendix B, Table B1) that Hochwarter and colleagues' (2003) felt accountability measure can be modified for use in the public sector performance appraisal context and that one of the three developed items was a valid indicator.

Incentive structure. The incentive structure construct was composed of both incentives and disincentives. The two-factor, four-item correlated model provided a good model fit (GFI = .997, NFI = .988, CFI = .996, and RMSEA = .046). Appendix B (Table B2) shows the list of the items used. All of the standardized factor loadings were statistically significant at a level of $p < .05$ or lower, ranging from 0.379 to 1.456 (see Appendix C for a more detailed description). The covariance between incentives and

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

disincentives was −0.251 at a $p < .086$ level, reconfirming the contradictory relationship between the incentives and disincentives of appraisal. The result suggests that the two dimensions developed by extending the conceptual models of Murphy and Cleveland (1991, 1995) and Harris (1994) are valid measures.

PSM. Previous PSM studies have verified a four-correlated-factor model as a proper model to measure PSM (Kim, 2009; Kim et al., 2013; Perry, 1996). Consistent with these findings, the final result of the modification process was the four-factor correlated 15-item model. As shown in Appendix B (Table B3), the four-factor correlated model displayed a good model fit (GFI = .923, NFI = .911, CFI = .957, RMSEA = .060). All of the standardized factor loadings of the 15 items were statistically significant at a level of $p < .001$, ranging from .603 to .812.

The results show that the four-factor correlated PSM measurement model developed by integrating Kim's (2011) and Kim and his colleagues' (2013) work provided good measures for all four dimensions. This, once again, confirms the widely accepted four-factor PSM measurement models in the public administration literature (Kim, 2009; Kim et al., 2013; Perry, 1996).

MRA

MRA was conducted to observe how the determinants of rater motivation affect the perceived rating accuracy of the raters. The composite indexes created from CFA were used as predictors (or determinants of rater motivation) and dependent variables (perceived performance rating accuracy). The independent variables were assumed to be distinctive from each other and to individually affect the dependent variable without interaction. The additive assumption in MRA requires that "each independent variable's effect on the dependent variable does not vary depending on the values of the other independent variables" (Berry, 1993, p. 5). Thus, MRA was consistent with the assumptions of this study. Table 3 provides information about each variable of the MRA.

Prior to the regression analysis, the classical assumptions of ordinary least squares (OLS) regression suggested by Berry (1993) were examined to ensure that OLS would be an appropriate estimator. The analyzed data revealed no serious violation of any assumptions of OLS. Table 4 provides the intercorrelation matrix of the variables.

Discussion

Table 5 summarizes the multiple regression estimates of the models. The adjusted R^2 value (= .274) was strong because the model explains approximately 27.4% of the variance in the perceived performance rating accuracy. The F-statistic was statistically significant at the $p < .001$ level. Furthermore, a regression model that excluded the control variables was executed to examine whether the addition of control variables affected the statistical significance of the independent variables.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3. Description of Variables in Multiple Regression Analysis.

Independent variables	Scale	Dependent variable
Determinants of rater motivation		
Felt accountability (composite index)	5-point Likert-type scale (index range = 4-20)	Perceived performance rating accuracy (composite index; index range = 3-15, 5-point Likert-type scale)
Incentive structure (composite index)	5-point Likert-type scale (index range = 4-20)	
Public service motivation (composite index)	5-point Likert-type scale (index range = 15-75)	
Availability of fixed rating distribution system	Dichotomous variable (0 = yes, 1 = no)	
Availability of system for rewarding accurate performance appraisers	Dichotomous variable (0 = yes, 1 = no)	
Control variables		
Country (United States/ Korea)	Dichotomous variable (0 = Korea, 1 = United States)	
Gender	Dichotomous variable (0 = Men, 1 = Women)	
Education	1 = High school/general equivalency diploma, 2 = some college, 3 = bachelor's degree or equivalent, 4 = some graduate school, 5 = master's degree or higher	
Tenure	Number of years in position	

Table 4. Intercorrelation Table of Continuous Variables (N = 236).

	M	SD	1	2	3	4	5	6
1. Perceived performance rating accuracy	10.839	2.013						
2. Felt accountability	13.733	2.423	.27*					
3. Incentive structure	13.211	2.425	.131*	.196*				
4. Public service motivation	59.449	6.942	.245*	.218*	.117			
5. Tenure	16.53	8.8	.168*	.048	−.221*	−.015		
6. Education	3.398	1.19	.08	−.08	−.291*	.133*	−.089	

*p < .05.

As displayed in Table 5, there were no major differences in the regression coefficients, R^2, and statistical significance between the complete model and the model with only the study's theoretically motivated variables. This demonstrates that the control

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5. Cross-Sectional Regression Estimates of Perceived Performance Rating Accuracy.

	Complete model		Only study variables	
Independent variables	Coefficient	*p* value	Coefficient	*p* value
Felt accountability	.142*	.004	.144*	.004
Incentive structure	.241**	.000	.225**	.000
Public service motivation	.046 *	.008	.049*	.004
RSAA (0 = yes, 1 = no)	−.475	.097	−.480	.098
FDRS (0 = yes, 1 = no)	.626*	.039	.737*	.015
Country (0 = Korea, 1 = United States)	−1.308**	.000	−1.618**	.000
Education	.139	.180	—	
Gender	−.684*	.006		
Tenure	.010	.478		
R^2	.302		.266	
Adjusted R^2	.274		.247	
F statistic	10.850		13.82	
Probability > *F*	.000**		.000**	
N	236		236	

Note. RSAA = reward system for accurate appraisers; FDRS = forced distribution rating system.
*$p < .05$. **$p < .001$.

variables minimally affected the regression estimates. A summary of the support for each hypothesis is presented in Table 6.

Determinants of Motivation

All three determinants of rater motivation (H1, H3, H4) had strong positive relationships with the dependent variable of perceived performance rating accuracy. Consistent with Mero and Motowidlo's (1995) contention that rater accountability can be a motivating force for the rater, the current results show that raters who more strongly perceive felt accountability toward audiences have higher expectations of being held accountable and acknowledge more information about the preferences of the audiences. Accordingly, the raters with high felt accountability (H1) were motivated to meet the expectations and preferences of the audiences. Thus, their perceived rating accuracy is positively affected.

Regarding the incentive structure (H3), the findings illustrate that if raters believe that they would benefit from making an accurate appraisal, the incentive structure level of the individual increases, eventually leading to higher perceived performance rating accuracy. Therefore, the conceptual propositions proposed but not tested by Murphy and Cleveland (1995) and Harris (1994) were supported by empirical evidence. The significant relationship between PSM and perceived performance rating accuracy (H4) demonstrates for the first time that the positive relationship between

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 6. Summary of Support for Hypotheses.

Hypothesis	Statement	Results	*p* value
1	The level of a rater's felt accountability and perceived performance rating accuracy is positively related.	Supported	<.05
2	Raters will show a higher perceived performance rating accuracy when there are direct formal incentives than when there are no direct incentives.	Partially supported	<.1
3	A rater's level of incentive structure and perceived performance rating accuracy is positively related.	Supported	<.001
4	The level of a rater's PSM and perceived performance rating accuracy is positively related.	Supported	<.05
5	The relationship between the existence of a forced distribution rating system and a rater's perceived performance rating accuracy is negative.	Supported	<.05

Note. PSM = public service motivation.

PSM and work performance (Naff & Crum, 1999; Perry & Wise, 1990; Vandenabeele, 2009) is applicable to the performance appraisal context.

Presence of Reward and Forced Distribution Rating System

Raters in organizations with a formal reward system for accurate appraisal (H2) showed higher perceived rating accuracy than raters in organizations without such systems. This result provides empirical support for Murphy and Cleveland's (1995) proposition that a formal reward system can be regarded as a direct incentive to rate more accurately. However, the availability of reward systems for accurate ratings was more predominant in Korea (37.7%) than in the United States (9.6%). Next, as expected, the negative relationship between a forced distribution rating system and perceived rating accuracy (H5) was observed. In the presence of a forced distribution system, raters were not able to truly reflect their perception of accuracy on the ratings because they were obligated to comply with the organization's expectations for rating distribution.

Control Variables

Among the control variables, country (Korea = 0, United States = 1) and gender (men = 0, women = 1) were statistically significant at the $p < .05$ level. For the country control variables, despite the generic approach of motivation of this study, it is possible that the ways in which raters perceive their appraisal accuracy may be affected by different organizational cultures. The United States and South Korea represent two distinct types of organizational culture in this study, reflecting the cultures of Western individualism and Eastern Confucianism, respectively.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

According to Ralston, Holt, Terpstra, and Kai-Cheng (1997), in Western countries, including the United States, Judeo-Christian religious values that emphasize "personal achievement and individual self-worth" (p. 9) have been major influences on individualism. In Eastern countries characterized by collectivism, such as Korea, the primary distinctive organizational culture is Confucianism, and the underlying leadership style in the Korean organization (management system) tends to be authoritative, promote harmony among employees, and show paternalism (Lee & Yoo, 1987; I. Yang, 2006). Accordingly, it is expected that raters under the influence of Confucianism are likely to be more lenient in making their performance ratings to maintain harmony with their subordinates and the organization's paternalistic culture.

Contrary to this study's expectations, Korean officials reported higher levels of perceived rating accuracy than did U.S. officials. In other words, Korean officials were more willing to accurately rate their employees and thus perhaps received rewards for making accurate ratings because the reward system was more available in Korea than it was in the United States (Table 2). In addition, truthfully rating employees may have been considered more important than promoting harmony or paternalistic culture in the workplace.

In regard to the gender control variable, although there is little research comparing male and female managers in the appraisal context, Benedict and Levine (1988) found that "females were more lenient with poor performers" (Bretz, Milkovich, & Read, 1992, p. 325). The current results support Benedict and Levine's (1988) findings in the sense that although the reason is unclear, female managers self-declare a lower degree of perceived performance rating accuracy than their male counterparts.

Conclusion

This study breaks new ground by developing a comprehensive RMM for the public sector and testing it in a preliminary fashion to investigate why raters are motivated to make accurate or lenient appraisals. Whereas the sample used in this study has limitations and the results cannot be broadly generalized, the large convenience sample of public officials demonstrates that the model has merit and should be tested further in more rigorous designs. This study underscores several factors that affect the motives of raters in public organizations. The results indicate that when assessing the accuracy of their performance ratings of subordinates, raters are affected by the appraisal audiences (supervisors and subordinates) and the expected benefits of providing accurate appraisal. More importantly, this study adapts private sector models to public sector conditions and demonstrates the importance of additional factors of rater motives, such as PSM. Raters with higher PSM tended to believe that they performed more accurately in their appraisal, validating the applicability of the positive PSM–work performance relationship (Naff & Crum, 1999; Perry & Wise, 1990) in the appraisal context.

The findings also reveal that raters considered the formal rewards for accurate appraisals as incentives to become more accurate in appraisal and the existence of a forced distribution rating system as a negative influence on appraisal accuracy. In

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

summary, the performance of appraisals may be dependent both on the raters' intention to be more or less accurate and on how the institutional reward and rating distribution systems are designed in organizations.

Considering the value of this research for human resource management in public organizations, this study presents a new perspective on performance appraisal and PFP. To create a more successful appraisal system, the current study argues that contemporary research needs to account for the motivational basis of not only the subordinates but also the administrators (raters) in appraisal. This study also finds that institutional rules make a difference in how raters perform their jobs.

The study offers recommendations for future research. Studies can examine the PFP system from the supervisor's perspective in greater depth. Researchers may study the alignment or conflict between the rater's motives in appraisal and PFP to improve the implementation process of the system. Although the PFP system needs accurate appraisals to properly reward employees, as demonstrated in the current study, raters are not always motivated to make accurate appraisals. Thus, future research may address how PFP systems can be modified to encourage raters to make accurate appraisals.

In addition, researchers can investigate the validity of the appraisal system by cross-examining the perceived accuracy of appraisal of both the supervisors and the subordinates. By doing so, future studies may reveal how performance ratings can have similar or different interpretations for rating accuracy depending on the types of employees.

The findings in this study also provide several implications for managing performance appraisal in public organizations. Because the raters perceive certain reward systems for accurate appraisers as an incentive to improve rating accuracy, public organizations should consider using these rewards to increase valid appraisal performance (Table 2 lists the examples). To decide the eligibility for the rewards for accurate appraisal, one recommendation is to create mechanisms through which the rater's supervisor monitors the appraisals in the PFP system. For example, because raters feel accountable to their supervisors in the appraisal context, the rater's supervisor could evaluate the subordinates of the rater on a preliminary basis and compare the results with those of the rater. With this mechanism, the rater will be more likely to conduct a more precise or defendable appraisal to account for the view of the rater's supervisor, and the raters that show high congruence with preliminary ratings assigned by the rater's supervisor can be considered as accurate appraisers and receive rewards. Finally, organizations can include the rater's appraisal accuracy as one of the performance evaluation criteria of the raters and provide additional rewards (e.g., pay, promotion, vacation hours) based on the results.

Limitations

This study has a number of limitations. It is a preliminary study with a newly developed model. Some measures of rater motivation have been introduced for the first time and must be refined in future research. Issues such as low Cronbach's alpha, factor

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

loadings, and convergent validity may be a result of the convenience sample and the preliminary tests undertaken. These limitations should be overcome in future research by developing more refined measures from this study. For that reason, this study serves the purpose of introducing a pilot model of rater motivation.

Another limitation is the possibility of the common-source bias (CSB) that may result from relying on the same survey respondents for independent and dependent variables. Recent research stresses that CSB is not a major issue if respondents are asked for "observable behavior" or "strategy questions about reacting" (Kwon, Berry, & Jang, 2012; Meier & O'Toole, 2010, p. 17). The majority of survey questions in this study are related to these two categories. Thus, the risk of CSB is mitigated.

Finally, the results of the study have limited generalizability to a broad group of public officials. The sample is composed of officials in selected government organizations in a southern state in the United States and Korea. Therefore, despite the fact that this research does not involve any political issues and pursues a generic approach in motivation research, it cannot be argued that the results are generalizable across all levels of governments, countries and business organizations. Future research may conduct national-level random sampling techniques to address this issue.

Appendix A

Table A1. Items of Felt Accountability.

Item (α = .609)	Question
FA1 Modified Hochwarter, Kacmar, and Ferris's (2003) measure	I am held very accountable for my performance appraisal as a rater.
FA2	I often have to explain about the performance ratings I assigned to my subordinates.
FA3	My supervisor holds me accountable for my performance rating decisions.
FA4	If performance appraisal does not go the way that they should, I will hear about it from my supervisor.
FA8	In the grand scheme of things, my efforts at work are very important.
FA9	Co-workers, subordinates, and bosses closely scrutinize my efforts at work.
FA5 Newly developed items	I am usually aware of what rating my subordinates expects to receive (self-rating).
FA6	I am aware of the subordinates who my supervisor likes or favors.
FA7	I understand my supervisor's definition of good indication of performance.

Note. Measured in Likert-type scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *neither agree nor disagree*, 4 = *agree*, 5 = *strongly agree*). FA = felt accountability.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table A2. Items of Incentive Structure.

Item	Dimension	Question
IS1	Incentive (α = .539)	My organization's current performance appraisal rating format is convenient to use.
IS3		My organization has a culture of promoting accurate performance appraisal.
IS4		I believe giving accurate performance appraisal feedback to my subordinates would lead to a performance increase in my unit.
IS5		I feel good when I give honest and accurate performance appraisal feedback to my subordinates.
IS2	Disincentive (α = .529)	Complying with my organization's performance appraisal rating format takes too much time.
IS6		My organization tends to encourage lenient performance appraisal.
IS7		I believe giving lenient performance appraisal feedback to my subordinates would lead to a performance increase of my unit.
IS8		I feel uncomfortable when I give negative performance appraisal feedback even though they are based on an accurate appraisal I made.

Note. Measured in Likert-type scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *neither agree nor disagree*, 4 = *agree*, 5 = *strongly agree*). IS = Incentive structure.

Table A3. Items of Public Service Motivation.

Item	Dimension	Question
PSM1	APS (α = .865)	I admire people who initiate or are involved in activities to aid my community.
PSM2		It is important to contribute to activities that tackle social problems.
PSM3		Meaningful public service is very important to me.
PSM4		It is important for me to contribute to the common good.
PSM5		I am interested in making public programs that are beneficial for my country of the community I belong to.
PSM6		Sharing my views on public policies with others is attractive to me.
PSM7		Seeing people get benefits from the public program I have been deeply involved in brings me a great deal of satisfaction.
PSM8	CPV (α = .763)	I consider public service my civic duty.
PSM9		I would prefer seeing public officials do what is best for the whole community even if it harmed my interests.
PSM10		I think equal opportunities for citizens are very important.
PSM11		It is important that citizens can rely on the continuous provision of public services.
PSM12		It is fundamental that the interests of future generations are taken into account when developing public policies.
PSM13		To act ethically is essential for public servants.
PSM14	COM (α = .836)	It is difficult for me to contain my feelings when I see people in distress.

(continued)

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table A3. (continued)

Item	Dimension	Question
PSM15		I am often reminded by daily events how dependent we are on one another.
PSM16		I feel sympathetic to the plight of the underprivileged.
PSM17		I empathize with other people who face difficulties.
PSM18		I get very upset when I see other people being treated unfairly.
PSM19		Considering the welfare of others is very important.
PSM20	SS (α = .777)	Making a difference in society means more to me than personal achievements.
PSM21		I am prepared to make enormous sacrifices for the good of society.
PSM22		I believe in putting duty before self.
PSM23		I would agree to a good plan to make a better life for the poor, even if it costs me money.

Note. Measured in Likert-type scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *neither agree nor disagree*, 4 = *agree*, 5 = *strongly agree*). PSM = public service motivation; APS = attraction to public services; CPV = commitment to public values; COM = compassion; SS = self-sacrifice.

Table A4. Items of Perceived Performance Rating Accuracy (N = 236).

Item	Question	*M*	SD	α
PRC1	How would you describe your level of satisfaction with the performance appraisal experience with your subordinates in the most recent evaluation period?	3.470	0.817	.819
PRC2	In general, as a rater, how would you assess your rating accuracy in the last evaluation period?	3.653	0.737	
PRC3	How closely do the performance ratings you gave to your subordinates reflect your willingness to make accurate appraisals?	3.716	0.794	

Note. Measured in Likert-type scale (1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree). The possible range of the composite index was from 3 to 15 with a mean of 10.839 and a standard deviation of 2.013. PRC = perceived performance rating accuracy.

Appendix B

Table B1. Standardized Factor Loading Estimates for the One-Factor Four-Item Model (Felt Accountability).

Factor	Item	*M*	*SD*	Factor loading	*t* value	R^2	α
Felt accountability	FA1	3.758	0.839	0.656**	8.417	.43	.645
	FA3	3.394	0.923	0.695**	8.805	.483	
	FA4	2.983	0.889	0.46**	6.066	.211	
	FA7	3.597	0.827	0.449**	5.909	.201	

Note. FA = felt accountability.
*p < .05. **p < .001.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table B2. Standardized Factor Loading Estimates for the Two-Factor Four-Item Correlated Model (Incentive Structure).

Factor	Item	*M*	*SD*	Factor loading	*t* value	R^2	α (Overall = .515)
Incentive	IS4	3.576	0.894	0.379*	2.142	.144	.71
	IS5	3.919	0.819	1.456*	2.292	2.121	
Disincentive	IS6	3.068	0.956	0.431**	3.943	.186	.419
	IS8	3.216	1.107	0.62**	4.363	.384	

Note. IS = Incentive structure.
*p < .05. **p < .001.

Table B3. Standardized Factor Loading Estimates for the Four-Factor Fifteen-Item Correlated Model (PSM).

Item	Dimension	*M*	*SD*	Factor loading	α
PSM2	APS	3.936	0.659	0.741	.873
PSM3		4.059	0.675	0.788	
PSM4		4.182	0.656	0.782	
PSM5		3.822	0.837	0.812	
PSM7		4.013	0.699	0.7	
PSM10	CPV	4.203	0.653	0.739	.747
PSM11		4.072	0.788	0.794	
PSM12		4.271	0.641	0.603	
PSM15	COM	3.864	0.683	0.629	.826
PSM16		3.962	0.763	0.81	
PSM17		4.114	0.612	0.757	
PSM19		4.021	0.642	0.764	
PSM20	SS	3.708	0.757	0.805	.779
PSM21		3.453	0.800	0.754	
PSM22		3.767	0.709	0.629	
Interfactor covariance	1		2		3
1. APS					
2. CPV	.629				
3. COM	.589		.83		
4. SS	.608		.532		.614

Note. All factors loadings and interfactor covariances are significant at p < .001. PSM = public service motivation; APS = attraction to public services; CPV = commitment to public values; COM = compassion; SS = self-sacrifice.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table B4. Composite Index of Variables (N = 236).

Variables	Items	Index range	Minimum	Maximum	*M*	*SD*
Felt accountability	FA1, FA3, FA4, FA7 (4 items)	4-20	6	19	13.733	2.423
Incentive structure	IS4, IS5, IS6, IS8 (IS6 and IS8 are reversed; 4 items)	4-20	6	19	13.211	2.425
Public service motivation	PSM2, PSM3, PSM4, PSM5, PSM7, PSM 10, PSM11, PSM12, PSM15, PSM16, PSM17, PSM19, PSM20, PSM21, PSM22 (15 items)	15-75	33	75	59.449	6.942
Perceived performance rating accuracy	PRC1, PRC2, PRC3	3-15	6	15	10.839	2.013

Note. Each item is measured in Likert-type Scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *neither agree nor disagree*, 4 = *agree*, 5 = *strongly agree*). Survey questions for each item are available in Appendix A. FA = felt accountability; IS = incentive structure; PSM = public service motivation; PRC = perceived performance rating accuracy.

Appendix C

While researchers generally expect the standardized factor loadings to be below 1, the factor loading value of 1.456 for the item incentive structure (IS5; Appendix B, Table 2) is not necessarily problematic in confirmatory factor analysis (CFA). According to Jöreskog (1999), the misunderstanding that completely standardized factor loading need to be smaller than 1 originates from

> classical exploratory factor analysis where factor loading are correlations if a correlation matrix is analyzed and the factors are standardized and uncorrelated (orthogonal). However, if the factors are correlated (oblique), the factor loadings are regression coefficients and not correlations and as such they can be larger than one in magnitude. (p. 1)

Jöreskog (1999) also stated that factor loadings above 1 may be an evidence of multicollinearity in the data. In this two-factor four-item correlated model, as the two factors together measure the construct of incentive structure, the presence of some multicollinearity does not necessarily mean that the item should be dropped. Therefore, item IS5 is included in the model.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Alonso, P., & Lewis, G. B. (2001). Public service motivation and job performance. *The American Review of Public Administration, 31*, 363-380.

Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology, 49*(1), 141-168.

Benedict, M. E., & Levine, E. L. (1988). Delay and distortion: Tacit influences on performance appraisal effectiveness. *Journal of Applied Psychology, 73*, 507-514.

Berman, E. M., Bowman, J. S., West, J. P, & Van Wart, M. (2006). *Human resource management in public service: Paradoxes, processes, and problems*. Thousand Oaks, CA: Sage.

Berry, W. D. (1993). *Understanding regression assumptions*. Newbury Park, CA: Sage.

Blume, B. D., Baldwin, T. T., & Rubin, R. S. (2009). Reactions to different types of forced distribution performance evaluation systems. *Journal of Business and Psychology, 24*, 77-91.

Bowman, J. S. (1994). At last, an alternative to performance appraisal: Total quality management. *Public Administration Review, 54*, 129-136.

Bretz, R. D., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management, 18*, 321-352.

Bright, L. (2007). Does person-organization fit mediate the relationship between public service motivation and the job performance of public employees? *Review of Public Personnel Administration, 27*, 361-379.

Decotiis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. *The Academy of Management Review, 3*, 635-646.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management, 17*, 263-282.

Duncan, J. R., Atkins, S. G., Fletcher, R. B., & Stillman, J. A. (2005). Frame of reference training for assessment centers: Effects on interrater reliability when rating behaviors and ability traits. *Public Personnel Management, 34*, 17-30.

Ferris, G. R., Judge, T. A., & Rowland, K. M. (1994). Subordinate influence and the performance evaluation process: Test of a model. *Organizational Behavior and Human Decision Processes, 58*, 101-135.

Frank, S. A., & Lewis, G. G. (2004). Government employees working hard or hardly working? *The American Review of Public Administration, 34*, 36-51.

Frink, D. D., & Ferris, G. R. (1998). Accountability, impression management, and goal setting in the performance evaluation process. *Human Relations, 51*, 1259-1283.

Gbadamosi, G., & Ross, C. (2012). Perceived stress and performance appraisal discomfort: The moderating effects of core self-evaluations and gender. *Public Personnel Management, 41*, 637-659.

Hair, J., Anderson, R., Tatham, R., & Black, W. (1998). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice-Hall.

Harris, M. M. (1994). Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management, 20*, 735-756.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hatcher, L. (1998). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.

Hochwarter, W. A., Kacmar, C. J., & Ferris, G. R. (2003, April). *Accountability at work: An examination of antecedents and consequences*. Paper presented at the annual meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.

Ingraham, P. W. (1993). Of pigs in pokes and policy diffusion: Another look at pay-for-performance. *Public Administration Review*, *53*, 348-356.

Jöreskog, K. G. (1999). *How large can a standardized coefficient be?* Retrieved from http://www.ssicentral.com/lisrel/techdocs/HowLargeCanaStandardizedCoefficientbe.pdf

Kim, S. (2009). Testing the structure of public service motivation in Korea: A research note. *Journal of Public Administration Research and Theory*, *19*, 839-851.

Kim, S. (2011). Testing a revised measure of public service motivation: Reflective versus formative specification. *Journal of Public Administration Research and Theory*, *21*, 521-546.

Kim, S., Vandenabeele, W., Wright, B. E., Andersen, L. B., Cerase, F. P., Christensen, R. K., . . .Vivo, P. D. (2013). Investigating the structure and meaning of public service motivation across populations: Developing an international instrument and addressing issues of measurement invariance. *Journal of Public Administration Research and Theory*, *23*, 79-102.

Klimoski, R., & Inks, L. (1990). Accountability forces in performance appraisal. *Organizational Behavior and Human Decision Processes*, *45*, 194-208.

Kline, R. B. (2005). *Principles and practices of structural equation modeling* (2nd ed.). New York, NY: Guilford.

Kwon, M., Berry, F., & Jang, S. (2012). A theoretical framework on the determinants of strategic cities: Empirical results from Florida City governments. *The American Review of Public Administration*, *43*, 438-459.

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, *87*, 72-107.

Lee, S. M., & Yoo, S. (1987). The K-type Management: A driving force of Korean prosperity. *Management International Review*, *27*(4), 68-77.

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, *30*, 881-905.

Longenecker, C. O., & Nykodym, N. (1996). Public sector performance appraisal effectiveness: A case study. *Public Personnel Management*, *25*, 151-164.

Longenecker, C. O., Sims, H. P., & Gioia, D. A. (1987). Behind the mask: The politics of employee appraisal. *The Academy of Management Executive*, *1*, 183-193.

Martin, J., & Siehl, C. (1983). Organizational culture and counterculture: An uneasy symbiosis. *Organizational Dynamics*, *12*(2), 52-64.

Meier, K., & O'Toole, L. (2010, September). *Organizational performance: Measurement theory and an application: Or, common source bias, the Achilles Heel of public management research*. Paper presented at APSA 2010 Annual Meeting, Washington, DC.

Mero, N. P., Guidice, R. M., & Brownlee, A. L. (2007). Accountability in a performance appraisal context: The effect of audience and form of accounting on rater response and behavior. *Journal of Management*, *33*, 223-252.

Mero, N. P., & Motowidlo, S. J. (1995). Effects of rater accountability on the accuracy and the favorability of performance ratings. *Journal of Applied Psychology*, *80*, 517-524.

Miller, C. E., & Thornton, C. L. (2006). How accurate are your performance appraisals? *Public Personnel Management*, *35*, 153-162.

Murphy, K. R., & Cleveland, J. (1991). *Performance appraisal: An organizational perspective*. Boston, MA: Allyn & Bacon.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Murphy, K. R., & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal based perspectives*. Thousand Oaks, CA: Sage.

Naff, K. C., & Crum, J. (1999). Working for America: Does public service motivation make a difference? *Review of Public Personnel Administration, 19*(4), 5-16.

Park, S., & Berry, F. (2012). Successful diffusion of a failed policy: The case of pay-for-performance in the US federal government. *Public Management Review*. Advance online publication.

Park, S. (2013). *Rater motivation in performance appraisal of public organizations: Effects of motivating factors on appraisal accuracy* (Doctoral dissertation). Retrieved from Electronic Theses, Treatises and Dissertations. ETD: 7961.

Pearce, J. L., & Perry, J. L. (1983). Federal merit pay: A longitudinal analysis. *Public Administration Review, 43*, 315-325.

Perry, J. L. (1996). Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory, 6*, 5-22.

Perry, J. L., Engbers, T. A., & Jun, S. Y. (2009). Back to the future? Performance-related pay, empirical research, and the perils of persistence. *Public Administration Review, 69*, 39-51.

Perry, J. L., Hondeghem, A., & Wise, L. R. (2010). Revisiting the motivational bases of public service: Twenty years of research and an agenda for the future. *Public Administration Review, 70*, 681-690.

Perry, J. L., Petrakis, B. A., & Miller, T. K. (1989). Federal merit pay, round II: An analysis of the performance management and recognition system. *Public Administration Review, 49*, 29-37.

Perry, J. L., & Wise, L. R. (1990). The motivational bases of public service. *Public Administration Review, 50*, 367-373.

Rainey, H. G. (2003). *Understanding and managing public organizations* (3rd ed.). San Francisco, CA: Jossey-Bass.

Rainey, H. G., & Bozeman, B. (2000). Comparing public and private organizations: Empirical research and the power of the a priori. *Journal of Public Administration Research and Theory, 10*, 447-470.

Ralston, D. A., Holt, D. H., Terpstra, R. H., & Kai-Cheng, Y. (1997). The impact of national culture and economic ideology on managerial work values: A study of the United States, Russia, Japan, and China. *Journal of International Business Studies, 28*, 177-207.

Roberts, G. E. (1994). Maximizing performance appraisal system acceptance: Perspectives from municipal government personnel administrators. *Public Personnel Management, 23*, 525-549.

Roberts, G. E. (2003). Employee performance appraisal system participation: A technique that works. *Public Personnel Management, 32*, 89-98.

Roch, S. G., McNall, L. A., & Caputo, P. M. (2011). Self-judgments of accuracy as indicators of performance evaluation quality: Should we believe them? *Journal of Business and Psychology, 26*, 41-55.

Schleicher, D. J., Bull, R. A., & Green, S. G. (2009). Rater reactions to forced distribution rating systems. *Journal of Management, 35*, 899-927.

Selden, S., & Sowa, J. E. (2011). Performance management and appraisal in human service organizations: Management and staff perspectives. *Public Personnel Management, 40*, 251-264.

Smith, W. J., Harrington, K. V., & Houghton, J. D. (2000). Predictors of performance appraisal discomfort a preliminary examination. *Public Personnel Management, 29*, 21-32.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Spence, J. R., & Keeping, L. M. (2010). The impact of non-performance information on ratings of job performance: A policy-capturing approach. *Journal of Organizational Behavior, 31*, 587-608.

Tetlock, P. E. (1985). Accountability: The neglected social context of judgment and choice. In L. L. Cummings & B. M. Straw (Eds.), *Research in organizational behavior (Vol. 7*, pp. 297-332). Greenwich, CT: JAI Press.

Underhill, J., & Oman, R. (2007). A critical review of the sweeping federal civil service changes. *Review of Public Personnel Administration, 27*, 401-420.

Vandenabeele, W. (2009). The mediating effect of job satisfaction and organizational commitment on self-reported performance: More robust evidence of the PSM—Performance relationship. *International Review of Administrative Sciences, 75*, 11-34.

Wang, X. M., Wong, K. F. E., & Kwong, J. Y. Y. (2010). The roles of rater goals and ratee performance levels in the distortion of performance ratings. *Journal of Applied Psychology, 95*, 546-561.

Wayne, S. J., & Liden, R. C. (1995). Effects of impression management on performance ratings: A longitudinal study. *Academy of Management Journal, 38*, 232-260.

Williams, L. J., Vandenberg, R. J., & Edwards, J. R. (2009). Structural equation modeling in management research: A guide for improved analysis. *The Academy of Management Annals, 3*(1), 543-604.

Wright, B. E., & Grant, A. M. (2010). Unanswered questions about public service motivation: Designing research to address key issues of emergence and effects. *Public Administration Review, 70*, 691-700.

Yang, I. (2006). Jeong exchange and collective leadership in Korean organizations. *Asia Pacific Journal of Management, 23*, 283-298.

Yang, K. (2009). Examining perceived honest performance reporting by public organizations: Bureaucratic politics and organizational practice. *Journal of Public Administration Research and Theory, 19*, 81-105.

Author Biography

Seejeen Park is an adjunct instructor at Myongji University Public Administration Department. His research interests include human resource management, organization theory, and strategic management.

Downloaded from ppm.sagepub.com by Pro Quest on March 16, 2015

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.