경영학박사학위논문

Essays on Performance Measure Drops and Weight Changes

성과지표 탈락과 가중치 변화에 대한 연구

2014년 8월

서울대학교 대학원

경영학과 경영학전공

김 범 준

Essays on Performance Measure Drops and Weight Changes

지도교수 안 태 식

이 논문을 경영학박사학위 논문으로 제출함
2014년 4월

서울대학교 대학원

경영학과 경영학 전공

김 범 준

김범준의 경영학박사 학위논문을 인준함
2014년 6월

위 원 장 신 재 용 (인)

부위원장 황 인 이 (인)

위 원 안 태 식 (인)

위 원 윤 대 희 (인)

위 원 최 연 식 (인)

ABSATRACT

Essays on Performance Measure Drops and Weight Changes

Bum-Joon Kim

Business School

Seoul National University

My dissertation explores performance evaluation system. Specifically, I am interested in performance measure selection and performance measure weight. Thus, I develop two essays on performance measure drops and weight changes.

The purpose of my first essay is to investigate the determinants of performance measure drop which has not been fully explored yet. I examine the economic and behavioral factors that influence a performance measure's drop using a performance evaluation data of state-owned enterprises (SOEs) in Korea. I find that (1) less discriminable performance measures are more likely to be dropped; (2) the lower the objective performance score is, the more likely it is to be dropped; (3) performance measures with non-participative target setting methods are likely to be dropped. This study contributes to the performance evaluation literature by identifying the

determinants of performance measure life span. While most prior studies focus on performance evaluation process, this study focused on the evaluation system design process and its interactions with performance evaluation results to improve the understanding of the dynamics of performance evaluation system. My study sheds light on how the incentives of designer/raters and ratees influence performance evaluation results and hence performance evaluation system design, i.e. performance measure drop.

In my second essay, I empirically investigate how ratees respond under a multi-task agency environment to the change of performance measure weights conducted by the designer of the performance evaluation system

The results are as follows. First, when the designer of the performance evaluation system changes a performance measure weight, ratees change their effort on that performance measure accordingly. This result supports the multi-task agency theory that states that ratees would exert more effort on the performance measure with the increased weight because ratees could benefit more from exerting same level of effort (Brickley, Smith and Zimmerman 2009). Second, when a performance measure weight increases, ratees exert more effort when that measure is a common measure than when it is a unique measure. These results are consistent with prior studies that find that ratees would input more effort on the performance measure that can be subject to relative performance evaluation (Frederickson 1992; Matsumura and Shin 2006) and that the presence of relative performance information increases performance (Tafkov 2013).

Third, ratees reallocate their effort more strongly in measures where they have high task-specific ability than in those of low task-specific ability when performance measure weights are changed.

My study contributes to the performance evaluation literature. First, I empirically show the relation between incentive change and effort allocation more clearly than prior literature. Second, I also provide empirical evidence that the feedback of RPI is a useful mechanism to motivate the ratee to exert more effort. Finally, I present empirical findings that task-specific ability influences the relation between incentive and effort, which has not been empirically explored thoroughly (Bonner and Sprinkle 2002).

Keywords: Discriminability, Performance Measure Drop, Performance Measure Weight, Commonality, Task-Specific Ability, Ratee's incentive

Student Number: 2010 - 30147

TABLE OF CONTENTS

Essay 2. The Change of Performance Measure Weights and Ratee's Incentive

LIST OF TABLES

Essay 1. Determinants of Performance Measure Drop

Essay 2. The Change of Performance Measure Weights and Ratee's Incentive

LIST OF FIGURES

Essay 1. Determinants of Performance Measure Drop

Essay 2. The Change of Performance Measure Weights and Ratee's Incentive

Essay 1

Determinants of Performance Measure Drop

I. Introduction

Performance measure selection is one of the most important elements in performance evaluation and compensation. Prior literature shows that economic and behavioral factors influence the performance evaluation system design to evolve over time by selecting more appropriate performance measures (Meyer and Gupta 1994; Ittner, Larcker and Meyer 2003; Malina and Selto 2004; Hassab-Elnaby, Said and Wier 2005; Ittner and Larcker 2009). This evolution of performance evaluation system is originated not only from changes in the performance evaluation environment but also from the interaction between raters and ratees. In other words, considering both performance environment change and interaction between raters and ratees, designer/raters will make decisions about whether the performance measures should be retained or discarded and whether the relative weight of the performance measure should be adjusted to motivate ratees effectively. However, ratees would also try to exercise influence on the designer/raters to include favorable performance measures and drop unfavorable performance measures for their own interests in the performance evaluation system design process. Thus, the rapid and ongoing evolution in performance evaluation system provides good research opportunities to investigate the influence of economic and behavioral factors on performance measure selection process (Ittner and Larcker 2009).

Many prior studies examined economic and behavioral determinants that affect performance measure choices (Ittner, Larcker and Rajan 1997; Ittner and Larcker 2002;

Malina and Selto 2004). Ittner, Larcker and Rajan(1997) investigates the factors that influence the relative weight between financial and non-financial performance measures in CEO annual bonus contracts. They find that firms have relatively higher weights of non-financial measures when the firms implement an innovation-oriented strategy, exercise a strategic quality initiative, are under regulation, or when the noise of financial measures is high. Ittner and Larcker(2002) examines the determinants of performance measure choice in worker level incentive plans and show that not only economic factors such as upgrading the workforce and linking bonuses to the firm's ability, but also behavioral factors, such as labor union representation and management participation, affect performance measure selection. Nevertheless, these studies could not capture the evolutionary nature of performance measures because the results of prior studies were based on cross-sectional analyses from survey results conducted over a single period.

Using a multi-period setting regarding performance measure changes, Hassab-Elnaby et al. (2005) empirically examines which factors affect the decision to retain, disregard and change the relative weights of non-financial measures. They find that the non-financial performance measures in CEO compensation contracts would be retained when there is an appropriate matching between the firm's characteristics and non-financial measures and when financial performance was subsequently enhanced. Using U.S large manufacturing firm data, Malina and Selto(2004) provides 8 key attributes as a standard to judge the appropriateness of performance measures and shows that these

attributes are important factors to consider in performance measure choice and change. Specially, Meyer and Gupta(1994) argues that when the comparability and variability of performance measure decrease, the performance measure would "run down" due to lack of capability to discriminate between good and bad performance.

However, these prior studies do not fully consider other possible factors such as behavioral factors based on the rater's and ratee's incentives as determinants of performance measure selection. Prior literature mainly focuses on the match between firm level characteristics and performance measure choice (Ittner et al 1997; Ittner and Larcker 2002; Hassab-Elnaby et al 2005) but overlooks the information quality which is used for control and decision-making (Ittner and Larcker 2001). Moreover, these studies do not consider the effect of interaction between rater and ratee on performance measure selection. In addition, prior studies focus on the determinants of performance measure choice, not the opposite, performance measure drop. I investigate the determinants of performance measure drop an unexplored area in performance evaluation literature.

Although a few studies (Meyer and Gupta 1994; Malina and Selto 2004) consider performance measurement characteristics such as comparability and variability, their methods are mostly limited to qualitative methods using interviews and cases. This motivates my study to use empirical data to investigate why performance evaluation system changes, i.e. performance measure drop over multiple periods. my study attempts to fill this void by examining the determinants of performance measure drop

empirically, using data collected from the performance evaluation reports of 15 state owned enterprises (SEO) in South Korea. This data set is composed of almost 30 performance measures per firm for 26 years (1985-2011). More specifically, my data consists of actual performance scores, relative weights, measurement characteristics, such as objective or subjective, common or unique, and target setting methods. my study examines the factors that make designers/raters drop performance measures in the performance evaluation system design process. First, I investigate whether low performance measure discriminability measured by the cross-sectional standard deviation within a measure affects designer/raters to drop the performance measure. Second, by examining the effect of prior performance on objective performance measure drop, I examine whether the ratees conduct influence activities to drop the unfavorable performance measures. Finally, I use the ratee's fairness perception to investigate whether the ratee's participation in target setting affects performance measure drop decisions. I find that (1) the less discriminable performance measures are, the more likely it is to be dropped in the performance evaluation system design process; (2) the lower the objective performance score is, the more likely it is to be dropped; and (3) performance measures using participative target setting method are less likely to be dropped.

My study contributes to the performance evaluation literature in three ways. First, while prior literature largely focuses on the determinants of performance measure choice

cross-sectionally, I address the determinants of performance measure drop in a time-series manner. Second, I also improve the understanding of the dynamic performance evaluation system by disaggregating the performance evaluation system design process from the performance evaluation life cycle (Ittner and Larcker 2009; Du, Tang and Young 2012). Third, I shed light on the incentives of raters and ratees that influence the performance results, and in turn, affect the performance measure drop. Thus, the findings of my study will provide useful insights to the dynamics of the performance evaluation system evolution.

The remainder of the paper is organized as follows. I explain the research setting on performance evaluation systems of SOEs in Korea in Section II. In Section III, I develop my hypotheses based on theory. Then I describe empirical model in Section IV, and the empirical results of my study are the focus of Section V. Finally, Section VI concludes.

II. Research Setting

2.1. Overview of the SOE Performance Evaluation System

The Korean government reports annual performance evaluation results of State-Owned Enterprises (hereafter, SOEs or Ratees) where the government takes equity ownership of more than 50%. As of December 31, 2011, there were 15 SOEs engaged in mining, utilities, security, printing, tourism, and other various fields. <Table 1> shows the details on these SOEs.

[INSERT TABLE 1 ABOUT HERE]

The Korean government legislate the Law for Management of SOE in 1984 which let SOEs conduct business autonomously and applied the market mechanism in the operation of these firms. This new legislation sets up a performance-evaluation system for SOEs and an incentive based compensation scheme. To evaluate SOEs more appropriately, the government (hereafter, designer) appointed evaluation committee members (hereafter, raters) consisting of independent professors, CPAs, professional consultants and researchers[1].

1 Because designer delegates evaluation authorities to professional raters, their participation is limited to providing the evaluation guideline and solving important issues in the performance evaluation process. In the performance evaluation system design process, however, the designer is involved actively to coordinate raters and to reflect government policy.

In my setting, I divide the performance evaluation cycle into two stages; *performance evaluation system design process* and *performance evaluation process*. First, both designer and raters participate together in choosing performance measures *in the performance evaluation system design process* before beginning performance evaluation. Once the performance measures are selected, raters review the reports prepared by ratees and evaluate ratees *in the performance evaluation process* based on the performance evaluation guideline which the designer prepared. When raters finish performance evaluation, they provide feedback on performance results to the designer so that feedback information can be incorporated into subsequent *performance evaluation system design processes*.

Moreover, after performance evaluation is completed, the government reports evaluation results to the President of Republic of Korea and the National Assembly and discloses performance evaluation results like the SOE's total score and its ranking to the public. In order to motivate SOEs, the government recommends that the President reward SOEs which have shown excellent performance or discharge the CEO of SOEs that have had bad performance and should determine how much incentive bonus will be provided to the SOEs based on the evaluation results.

From 1985, the Korean government designed and developed the performance measurement structure to reflect changing business and policy environments. Table 2 presents the overall structure of the performance evaluation system of a SOE in 2011.

The first column of this table shows 3 categories: overall management, main business, and business management, which is composed of human resources, financial/budget management and other operations management. The third column in table 2 indicates that performance measures can be distinguished as objective or subjective measures. First, objective performance measures are evaluated following 4 different target setting methods: (1) actual to target analysis, (2) target-range assignment analysis, (3) trend analysis and (4) beta distribution analysis (Ahn, Hwang and Kim 2010). In my study, I classify target setting methods into two groups which are shown in fourth column of table 2. One group consists of performance measures that have applied a participative target setting method, including the actual to target analysis (Case I) and target-range assignment analysis (Case II). These two methods allow ratees to communicate with raters through voice and explanation (Case I) or through explanation only (Case II) (Ahn, Choi and Hwang 2013). The other group consists of performance measures that apply formula based target methods including trend analysis and beta distribution analysis. These methods do not permit ratees to input any information within the target setting process and determine the target based on a predetermined formula. Second, subjective performance measures have a single evaluation method: ranking on a basis of 9 grade scales. Raters could give ratings entirely based on their subjective evaluation.

[INSERT TABLE 2 ABOUT HERE]

2.2. Descriptive Statistics of Performance Measure Drop

Table 3 shows the number of performance measures that were dropped by year. On average, 23.31% of all performance measures were dropped every year. The smallest number of dropped performance measures is 5 (2.1%) in 1999. In contrast, the largest number of dropped performance measures is 126 (59.72%) in 2007. These dropped performance measures consist of objective performance measures (28.4%) and subjective performance measures (71.6%). As shown in Figure 1, the number of dropped performance measures fluctuated over time. I conjecture that the periodical fluctuation might be partially due to political power shifts, corresponding government policy changes over SOEs, and the following structural changes of performance evaluation system. In addition, it might also be partially due to environmental shocks, like the Asian financial crisis in 1997, because designers and raters are likely to redesign their measurement systems to reflect their current environment and strategies (Kennerley and Neely 2003).

[INSERT TABLE 3 AND FIGURE 1 ABOUT HERE]

Figure 2 presents the trend of the ratio of the number of performance measure drops among the total number of performance measure dropped according to performance measure tenure. At the first year of the performance measure tenure, 37.5% of the

number of total performance measures are dropped. However, beginning with the second year of the performance measure tenure, there is no drastic change in the ratio of the number of performance measure drop on the total number of performance measure drop over time.

[INSERT FIGURE 2 ABOUT HERE]

2.3. The trend of the Number of Participative Target Setting Measure

Figure 3 shows the trend of the number of performance measures with participative target setting methods over time. At the beginning of the performance evaluation system, the targets of all objective performance measures are determined by formula based methods. However, the number of performance measures based on participative methods has grown rapidly from 2000 to 2008. Considering this increasing trend of participative target setting methods, I conjecture that the usefulness of participative target setting methods has been proved in performance evaluation.

[INSERT FIGURE 3 ABOUT HERE]

III. Theory and Hypothesis Development

3.1. Performance Measure Selection

In optimal incentive contracting, the performance measures in the incentive plan should be chosen to motivate agents to work in the manner desired by the owners of the firm (Ittner and Larcker 2002). According to agency theory, the performance measures are selected based on the informativness (or incremental information content) of each measure with regard to the agent's action choice (Holmstrom 1979). Banker and Datar(1989) shows that the relative weight placed on individual performance measures should be a function of sensitivity and precision. A lower weight would be placed on a performance measure with higher distortion in an incentive contract (Feltham and Xie 1994; Baker 2002).

Based on this theoretical background, prior literature investigates the factors that influence the performance measure selection (Ittner, Larcker and Rajan 1997; Ittner and Lacker 2002; Malina and Selto 2004). Ittner et al.(1997) investigates the factors that influence the relative weights placed on financial and non-financial measures in CEO bonus contracts. They find that an innovation-oriented strategy, the adoption of a strategic quality program, a higher level of regulation, and more noise in financial measures would increase the use of non-financial measures. Likewise, Ittner and Larcker(2002) examine the determinants of performance measure choice in work level incentive plans. They show that informativeness, as economic theories have emphasized,

affects performance measure choice. However, other factors which behavioral studies would indicate, such as attracting key talents, pay-for-performance linkage, labor union representation as well as management participation in plan design also influence performance measurement choice. As a comprehensive study regarding performance measure selection in management control theories, Malina and Selto(2004) provides 8 key desirable attributes which performance measures should have (1) being diverse and complementary (Milgrom and Roberts 1995), (2) objective and accurate (Libby, Salterio and Webb 2004), (3) informative (Antle and Demski 1988), (4) more beneficial than costly (Merchant 2012), (5) causally related (Kaplan and Norton 1996), (6) strategic communication devices (Kaplan and Norton 2001), (7) providing incentives for improvement (Rappaport 1999), and (8) being supportive of improved decisions (Luft and Shields 2002). They show that retained performance measures are more likely to have positive forms of these attributes. On the other hand, dropped performance measures are more likely to have negative forms of the attributes.

The number of studies which examine factors influencing the dynamics of performance evaluation design, i.e., drop of performance measures, is relatively scarce (Hassab-Elnaby et al 2005; Meyer and Gupta 1994). Hassab-Elnaby et al. (2005) investigates why firms retain the use of non-financial measures in CEO compensation contracts. They find that US firms which abandon non-financial measures in CEO compensation have bad short-term performance, deficiency of fit between the measures

and the organizational characteristics, CEO change, and changes in regulation. Meyer and Gupta (1994) shows that when the performance measures are less capable of discriminating between good and bad performance regarding the ratee's learning, the performance measures would be discarded. Prior studies consider not only firm-level factors such as strategy, innovation, regulation and union representation but also measure-level factor such as informativeness, accuracy, and objectiveness as factors influencing the performance measure selection. However, they still do not fully explore the rater's and ratee's behaviors and incentives which could affect performance measure selection. In my study focus on rater-ratee interaction and incentives in in selecting or dropping performance measures.

3.2. Rater's Incentives in Performance Evaluation

In my setting, raters are professionals delegated to evaluate the SOE from the government and they pursue to build their reputation as competent raters. They are accountable for their performance evaluation results not only to designers but also to the ratees. Of course their performance evaluation results will be publicized to the public showing indirect accountability. Raters have *accountability,* which refers to the perceived potential to be evaluated by a third-party and being responsible for one's behavior and decision (Frink and Ferrris 1998; Levy and Williams 2004). This *accountability* could motivate raters to input more effort and time to evaluate ratees

accurately. To note is that raters in this would have no incentive to underreport the ratee's performance to reduce the compensation to ratees since they are not final residual claimants(Prendergast 1999). Raters, however, also have incentives to minimize time and effort in the *performance evaluation process* for their own interests. They also have motivation to avoid confrontation with ratees so as not to damage personal relationships (Harris 1994). Specifically, when raters work in a physical location that is remote from that of the ratees, the cost of gathering information which is required to monitor the ratee's performance is high and therefore it is difficult for raters to collect complete performance information (Higgins and Bargh 1987). Such incomplete performance information causes ratings to be heavily centered around the mean on the rating continuum in subjective performance evaluations, a tendency that is referred to as the compression bias (Moers 2005; Bol 2011). In addition, accuracy rating and harsh communication with ratees can damage the personal relationships and raters might be met with criticism (Naiper and Latham 1986). Thus, when raters have incomplete performance information and, at the same time, ratees believe that their rating level is lower than their own perceived level, raters are more likely to evaluate more leniently above the actual performance level to avoid uncomfortable discussions with ratees, a tendency that is referred to as the leniency bias in subjective evaluation(Friedrich 1993; Moers 2005). Moreover, they are more likely to accept the request of changing performance measures from unfavorable measures to favorable ones in objective

evaluation when they believe that external factors cause low performance (Murphy and Cleveland 1995).

3.3. Ratee's Incentives in Performance Evaluation

Compared to the rater's incentives, the ratee's incentives are very simple: maximizing their performance ratings (performance score) while reducing efforts. One way to reduce efforts is by allocating ratees' efforts on the measures with high discriminability (Ahn et al. 2010). Also ratees could also exert influence activities in subjective performance evaluations, attempting to affect a rater's decision for their own benefit (Du et al. 2012). Ratees may engage in positive learning and perverse learning about the performance measure (Meyer and Gupta 1994).

3.4. The Discriminability of Performance Measure

Even when a performance measure is not controlled by the agent, as long as it gives incremental information on the agent's performance through relative performance evaluation, it could be useful in evaluating the performance of the agent (Antle and Demski 1988). Standardized financial measures which capture a diverse set of actions and allow horizontal comparison across diverse divisions may be relatively more informative than unique nonfinancial or behavioral measures (Aray, Glover, Mittendorf and Ye 2005). From a behavioral study, Lipe and Salterio (2000) show that only

common performance measures influence the superior's evaluation because they are easier to use in comparing multiple units due to the judgmental difficulties of the balanced scorecard. Therefore, I conjecture that common measures are considered more important than unique measures.

Meyer and Gupta (1994) shows that performance measures 'run down' over time as managerial actions make the variance[2] among units in the measures shrink, reducing the measure's ability to discriminate good performance from bad performance. This study argues that there are two kinds of learning behaviors which ratees could engage in over time. One is "positive learning" where real improvement takes place and the other is "perverse learning" where there is the appearance of improvement but not actual improvement(Meyer and Gupta 1994). The effects of positive learning and perverse learning would be inseparable because positive and perverse learning might be captured by the same performance measure (Meyer and Gupta 1994). On the other hand, raters have an incentive to compress the discrepancy of ratings among ratees to reduce the cost of confrontation in subjective performance evaluations when they have incomplete performance information (Moers 2005; Bol 2011). Thus, I conjecture that these factors could influence performance measure discriminability to decrease over time.

Meanwhile, Ahn et al. (2010) finds that the degree of performance measure discriminability is positively associated with performance improvement. In order to

[2] Meyer and Gupta (1994) use the terminology, 'comparability and variability' of performance measure instead of discriminability, However, I conjecture that there is no difference between the two concepts in my study.

motivate ratees through relative performance evaluation in common measures, performance measures should provide enough discriminability between good and bad performance (Ahn et al 2010).

Taken together, designer and raters have not only *accountability* by performing appropriate evaluation but also incentive to motivate ratees to input more effort. Therefore when they review the usefulness and appropriateness of performance measures, they should consider the performance measure informativeness that discriminability of the performance evaluation results will provide. Thus, if a performance measure turns out to be less discriminable, it is likely to be dropped in the performance evaluation system design process. Thus, I hypothesize:

Hypothesis 1. The less discriminable a performance measure is, the more likely the performance measure will be dropped.

3.5. Influence Activities in the Performance Evaluation System Design Process

As discussed in previous section, Performance evaluation systems have two processes: the *Performance evaluation system design process* and the *evaluation process.* While designer/raters are mainly responsible for evaluation system designs, ratees could also exert influence by voicing the problems of the evaluation system with the hopes of

having a favorable evaluation result in the next period. During the performance evaluation process, while the rater has full control of the process, ratees can influence raters by providing favorable information to raters during this period (Du et al 2012).

Subjective performance measures involve discretion by definition while objective performance measures limit the flexibility in the evaluation process (Bol 2008). In terms of ratee influence, influence activities are limited for objective evaluation because of its objectivity while various forms of influence activities are possible for subjective evaluation due to its subjective nature (Bol 2008; Du et al 2012). Therefore the ratee's influence is very limited during evaluation processes for objective measures while it is possible to exert influence for the entire evaluation process. For ratees to obtain favorable scores in objective measures, they must influence designer/raters before the *evaluation process* starts during the *evaluation system design process* where designer/raters retain or drop performance measures. In short, the ratee's influence is very limited during the evaluation of objective measures and influence is only possible during evaluation system design process. Ratees could always influence raters during subjective evaluation processes, which thereby reduce the incentives for ratees to exercise influence during the *performance evaluation system design process* for those measures. In other words, if there is influence exerted by ratees during the *performance evaluation system design process*, it will be only for objective measures.

If ratees could influence designer/raters in their objective measure retention or drop

decisions, what kind of objective measures will be candidates for drop; high scored measures or low scored measures? If ratees are obtaining higher scores on average for a certain objective measure relative to other objective measures, they have no incentive to influence designer/raters to drop that measure. If ratees are receiving lower scores on average for a certain measure, reducing absolute total scores, they will have strong incentives to drop that measure. Moreover, raters that are in charge of final performance selection should have accountability for both measure selection and final performance measured in the final score. Lower final scores in objective measures may indicate the failure of measure selection and target setting as well as the failure of ratee motivation (Murphy and Cleveland 1995). There is no reason for designer/raters as well as ratees to retain low-scoring objective performance measures (Mitchell and O'Reilly 1983; Murphy and Cleveland 1995). Therefore I hypothesize the following.

Hypothesis 2. The lower the performance score of an objective performance measure is, the more likely the performance measure will be dropped

3.6. Participative Target Setting in Performance Measures

Prior studies show that when ratees are involved in participative budgeting, ratees feel an enhanced perception of procedural fairness, which motivates ratees to exert more effort and to improve performance (Levy and Williams 2004; Lau and Tan 2005;

Wentzel 2002; Roberts and Reed 1996). For example, Roberts and Reed (1996) find that participation impacts appraisal acceptance which in turn affects appraisal satisfaction, employee motivation and productivity. Wentzel (2002) shows that increased participation in the budgeting process enhances the perception of fairness which fosters the manager's commitment to budgetary goals and subsequently increases performance ratings. Also, Lau and Tan (2005) find that participation in the budgeting process is positively related with the perception of procedural fairness. According to the organizational justice theory, a fair target setting process consists of two components: *voice* and *explanation*. First, voice is defined as the subordinate's ability to deliver his opinion about the target to the superior in the decision making process (Leventhal 1980). Second, explanation refers to the justification that is provided by the superior when the subordinate's opinion does not influence the outcome in the decision process (Bies 1987). Moreover, when voice and explanation are combined, there is a more significant performance improvement (Libby 1999). This result shows that the explanation delivered by the superior in the target setting process is very important in improving the ratee's perception of fairness. In my study, I define target setting methods not only by using both voice and explanation (Type I) but also by using explanation only (Type II) as participative target setting methods. Moreover, participative target setting methods provide opportunities for ratees to set up relatively easy targets to achieve in subsequent periods compared to the formula based target setting method (Ahn, Choi and Kwon

2011). Hence, ratees have the incentive to retain the performance measures with participative target setting methods due to the perception of fairness and the possibility of influencing in the process of setting targets.

On the contrary, with regard to whether the perception of fairness is a determinant of performance measure, Lau and Martin-Sardesai (2012) investigates how fairness influences the choice of performance measures. They find that organizational concern for workplace fairness plays a key role as a determinant of performance measures choice and that it subsequently enhances employee's performance. Thus, raters also have the incentive to retain the performance measures where ratees could participate in target setting process.

Taken together, both raters and ratees have incentive to retain the performance measures that have a participative target setting process. Therefore, designer/raters have the incentive to retain the performance measures which have a participative target setting method in the performance measure selection process. Thus, I hypothesize

Hypothesis 3. Objective performance measures which apply a participative target setting method are less likely to be dropped

IV. Empirical Design

4.1. Testing Hypothesis 1

I develop a regression model for the determinants of performance measure drop, and estimate equation (1) and (2) below by using logistic regression models for survival analysis. This analysis could provide the results from the discrete-time proportional odds model proposed by Cox (Allison 2010). I create an indicator variable (DROP), which would take on the value of one if designer/raters decide to drop a performance measure from performance measures which will be used in the subsequent period and zero if designer/raters keep the performance measure in the performance measures used in the subsequent period. The DROP measure would capture whether designer/raters decide to reduce a performance measure's weight to zero. To investigate H1 regarding whether discriminability among performance measures affects the designer's and rater's decision about performance measure drop from performance measures used in the subsequent period, I regress *DROP* as the dependent variable on *DISCRIMINABILITY* [3]as the independent variable measured as the standard deviation of the performance score in measure j and on year t to capture how much the performance deviates from the mean of performance measures in equation (1) (Ahn et al 2010). I expect the sign of coefficient

3 Ahn et al (2010) define discriminability as the prior 3 year average of the cross-sectional standard deviation of a performance measure. However, because designer/raters review the appropriateness of performance measures every year and there are many samples where the performance measure is dropped within 3 years. In my study, I redefine discriminability as the standard deviation of the performance score of a measure at only the last year.

β_1 to be negative, because low discriminability among ratees provides less informative information for raters.

$$Drop_{j,i,t+1} = \alpha_0 + \beta_1 * Discriminability_{j,t} + \beta_2 * Attain_{j,i,t} + \beta_3 * Weight_{j,i,t} + \beta_4 * M\_Tenure_{j,i,t} + \beta_5 * Ind\_Sub_{j,i,t} + \beta_6 * ln\_Asset_{i,t} + \beta_7 * Roa_{i,t} + \beta_8 * Lev_{i,t} + Year\ dummies + \varepsilon_t \quad (1)$$

where *i*, *j* and *t* indicate the firm, the individual performance measure and the time period respectively.

I include a set of control variables that could influence performance measures drop. Specifically, I control for various measurement characteristics; prior performance, tenure, weight and the subjectivity of performance measures, to capture the effect of prior performance level, performance measure age, relative importance, and subjectivity. I measure *ATTAIN* as the percentage of performance scores that the ratee achieved in the previous year, *M_tenure* as the number of periods since the performance measure was added to the performance measure set, and *WEIGHT* as points allocated to individual performance measures which add up to 100 points for each year. *Ind_Sub* is indicator variable, which takes the value of one if that measure is subjective; otherwise zero. To control for the firm's specific business conditions which would influence performance measure drop, I include *SIZE*, *ROA* and *LEV*. First, I measure *SIZE* as the natural

logarithm of the firm's total assets. Second I measure *ROA* as operating income divided by the book value of total assets to capture firm's profitability. Finally, I measure *LEV* as the book value of liabilities divided by the book value of assets. I incorporated dummy variables that represent years to capture year fixed effect on performance measure drop.

4.2. Testing Hypothesis 2 and 3

Regarding with H2 and H3, I estimate the following regression model (2) which is basically based on regression (1). This regression would be estimated using the subsample with only objective performance measures.

$$Drop_{j,i,t+1} = \alpha_0 + \beta_1 * Discriminability_{j,t} + \beta_2 * Low\_Score_{j,i,t} + \beta_3 * Participative_{,i,t} + \beta_4 * Weight_{j,i,t} + \beta_5 * M\_Tenure_{j,i,t} + \beta_6 * Low\_Score_{j,i,t} * Discriminability_{j,i,t} + \beta_7 * ln\_Asset_{i,t} + \beta_8 * Roa_{i,t} + \beta_9 * Lev_{i,t} + Year\ dummies + \varepsilon_t \quad (2)$$

where *i, j* and *t* indicate the specific firm, the individual performance measure and the time period respectively.

To test H2, I sort objective performance measures into quartiles based on performance score (*ATTAIN*) and then create an indicator variable *Low_Score* which takes on the value of one when it is from the lowest performance score quartile to capture the

objective performance measure group that has bad performance. I am interested in *Low_Score* to capture the effects of the performance level on objective performance measures.

According to H2, ratees would exert influence activities on designer/raters to drop objective performance measures with a low performance level in the *performance evaluation system design process*. Thus, I predict the sign of coefficient β_2 to be positive because the lower the performance score level is, the more likely it will be dropped. Moreover, I incorporate interaction term *Low_Score*Discriminability* to examine whether both bad performance and less discriminable performance measures affect the designer's and rater's decision about performance measure drop. I also expect the sign of coefficient β_6 to be negative because when the performance score is low, ratees have the incentive to exert influence activities to drop unfavorable objective performance measures. Also when performance measure discriminability is low, designer/raters also have the incentive to drop such performance measures.

According to H3, the ratee's participation in target setting process would increase the ratee's perception of fairness which makes raters less likely to drop the performance measure. To capture the effects of the type of target setting process, I incorporate the indicator variable *Participative* which takes the value of one if there is a participative target setting process and zero if it is a formula based target setting process. I expect the sign of coefficient β_3 to be negative because performance measures with participative

target setting process are less likely to be dropped.

All regressions in this section are estimated with Huber-White robust standard errors clustered by firm. These standard errors would be robust to both serial correlation and heteroskedasticity (Rogers 1993).

V. Empirical Results

5.1. Descriptive Statistics

Table 4 presents descriptive statistics of the performance measure drop (*DROP*) variable and its determinants. The total number of observations consist of 6,139 measure-firm-years. The average of Performance Measure Drop (*DROP*) is 0.233, which means that 23.3% of total observations are dropped measures. The average (median) discriminability, which is measured as the cross-sectional deviation among measures at a year is 0.098(0.080). The average(median) performance measure tenure is 4.923 (4) years, suggesting that performance measures are retained for almost 5 years after being included in a performance measurement set. Moreover, descriptive statistics of *WEIGHT* provides that an individual measure has 3.315 point weights on average, and is given from a 0.5 point to a 20 point weight according to its relative importance. The average (median) *ATTAIN* is 0.808(0.870) suggesting that individual performance measures get an approximately 80% rating on average.

The number of objective performance measures is 1,745 measure-firm-years. *Participative* which indicates the participation of ratees in the target setting process, accounts for 38.6% of all objective performance measures, suggesting that 61.4% of all total objective performance measures apply formula based target setting methods.

The average (median) SIZE of firms is 21.493(21.485) as natural log of total assets. The mean value of ROA and LEV are 1.2% and 62.7%, respectively, which implies that

firms are financially stable but less profitable.

[INSERT TABLE 4 ABOUT HERE]

Table 5 presents the Pearson correlation matrix. In general, control variables are not highly correlated with each other. Specifically, *DROP* is not significantly correlated with *discriminability* at a conventional level, suggesting that these results do not seem to support H1. In contrast to d*iscriminability*, *Attain* has a negative correlation with *DROP*, implying that low performance measures are likely to be dropped and this supports H2. However, this effect of *ATTAIN* on *DROP* should be examined more in detail because my sample consists of both subjective and objective measures. Because I do not incorporate control variables in this analysis, the results are only preliminary and an analysis using a multiple regression method should be applied.

[INSERT TABLE 5 ABOUT HERE]

5.2. Determinants of Performance Measure Drop

Table 6 presents the estimation results of equation (1). The dependent variable for column (1) and (2) is the indicator variable indicating the performance measure drop (*DROP*) for my full sample. The pseudo R-squared is about 27.2 percent in column (1)

and 27.4 percent in column (2). As shown in column (1), consistent with H1, I find that coefficient of discriminability is negative (-3.580) and statistically significant at the 1% level, suggesting that performance measures are more likely to be dropped when the performance measure's discriminability decreases. Moreover, in the regression from column (2), the coefficient of *discriminability* is still negative and statistically significant at 1% level. These results suggest that compressed ratings across firms make performance measures less informative in evaluating the ratee's performance due to the lack of capacity for relative performance evaluation[4]. Consequently, I interpret my findings as being consistent with H1: the less discriminable a performance measures is, the more likely the performance measure will be dropped.

[INSERT TABLE 6 ABOUT HERE]

Table 7 provides the estimation results of model (2) that tests H2 and H3 using the subsample of objective performance measures. With regard to H2, among objective performance measures I find that a low performance score measure is more likely to be dropped. The coefficients of *Low_Score* are 0.913 in column (1), 0.844 in column (2), 1.545 in column (3) and they are all statistically significant at the 1% confidence level.

[4] To control the effect of performance measure drop at the first year of measure tenure, I perform an additional test after excluding the sample of performance measure dropped at the first year of measure tenure. The results show that the coefficient of *discriminability* is still statistically significant at conventional level.

Moreover, I incorporate an interaction term, *Low_Score * Discriminability* to capture the interactive effects between low performance score and informativeness based on relative performance evaluation. The coefficient of this interaction term are -4.161 in column (3) and is statistically significant at the 10% confidence level. I find that a performance measure that has both low performance score and less discriminability at the same time is more likely to be dropped, suggesting that when ratees have bad performance in an objective performance measure and raters feel that the performance measure is less informative, there is a synergy effect towards dropping the performance measure which is unfavorable for both raters and ratees. Consistent with my prediction in H2, these results show indirect evidence that ratees increase influence activities on designer/raters to drop unfavorable objective performance measures.

Regarding H3, I find evidence supporting the prediction that a participative target setting process could enhance the ratee's perception of fairness and therefore raters would like to retain the performance measures which have participative target setting methods. The coefficients of *Participative* is -0.529 in column (1), -0.528 in column (2) and are statistically significant at the 1% confidence level. These results support H3 which conjectures that the performance measures having participative target setting are less likely to be dropped.

[INSERT TABLE 7 ABOUT HERE]

5.3. Additional Tests

Table 8 provides additional evidence with regard to H2. For this robustness test, I sort the objective performance measure into quartiles based on *discriminability* and then choose the lowest discriminable performance measures quartile as the subsample. Subsequently, I again categorize this sub-sample into quartiles based on the performance score (*Attain*) to capture the difference between good performing measures and bad performing measure at the same discriminability level. The coefficient of *Low_Perform* is 2.018 and statistically significant at the 1% confidence level, suggesting a bad performing measure is more likely to be dropped at the sample discriminability level. However, the coefficient of *High_Perform* is negative and statistically insignificant, implying that a good performing measure is irrelevant with performance measure drop at the lowest discriminability level. Consequently, I conjecture that bad performing measures at low discriminability level will be more likely to be dropped because not only do raters have the incentive to drop that less informative performance measure but also ratees would place pressure on designer/raters to drop the unfavorable performance measure.

[INSERT TABLE 8 ABOUT HERE]

In addition, I conduct an additional univariate analysis to support H2 by comparing

the performance score of newly added performance measures with that of dropped or that of non-added performance measures. Panel A of Table 9 shows that the performance score (0.876) of the newly added objective performance measures is higher than that (0.807) of dropped performance measures and that this difference is statistically significant at the 1% confidence level. Moreover, Panel B of Table 9 finds that the performance score (0.876) of the newly added objective performance measures is higher than that (0.855) of the non-added performance measures and is marginally significant at the 10% confidence level. Thus, I conjecture that Table 9 indirectly provides additional evidence that the ratee's influence activities exist in *the objective performance evaluation system design process*.

[INSERT TABLE 9 ABOUT HERE]

I perform an additional regression test by examining whether macro-economic factors like GDP and growth of GDP affect the performance measure drop because macro-economic changes such as the Asian financial crisis of 1997 or the global financial crisis of 2008 can make the designer change the performance measures dramatically. I incorporate macro-economic variables like *GDP* and *GDP_growth* as additional control variables. As shown in Panel A of Table 10, I find that macro-economic variables are positive (or negative) and statistically significant at 1% confidence level, suggesting that

national wide economic events like the Asian financial Crisis of 1997 are one of the main drivers in the selection of performance measures. Panel A also shows that the coefficients of *Discriminability* are still negative and statistically significant at 1% confidence level across Column (1) to Column (3), even after controlling the effect of macroeconomic factors. These results provide additional evidence that supports H1.

Panel B of table 10 provides additional evidence that supports H2 and H3 by showing that the coefficient of *Low_Score (participative)* is positive (negative) and statistically significant at the 1% confidence level across all regressions after controlling the effect of macro-economic factors, even though the coefficients of *GDP* or *GDP_growth* are not consistent across all regressions. These results provide additional evidence that supports H2 and H3.

[INSERT TABLE 10 ABOUT HERE]

Although the results are not tabulated, I perform the additional tests as robustness checks. First, to control the regime change of Korea, I incorporate regime change variables which capture each regime of the presidents from 1985 to 2011 as control variables instead of year dummy variables. Although all regime change dummy variables are statistically significant, the coefficients of independent variables of interest are consistent with those of the main analysis. Second, I find that the main results do not

change when the same regressions are conducted after excluding firm characteristics like size, profitability and leverage. Finally, I incorporate an alternative discriminability variable which is calculated as the average value of discriminability among measures during the prior two years, because the designer could consider the informativeness over longer periods before deciding performance measure drop.

IV. Conclusion

When designing a performance evaluation system, designer/raters should select the appropriate performance measures to motivate ratees effectively by reflecting the incentives of raters and ratees as well as the changes of the environment. Performance evaluation systems have evolved to make the performance evaluation system more effective (Ittner and Larcker 2009). In the process of evolution, economic and behavioral factors would affect the designer's and rater's decision about performance measures drop. My research topic regarding the effect of economic factors like discriminability and behavioral factors like the ratee's perception of fairness on the designer's and rater's decision to drop the performance measure is a valuable and intriguing research question.

The purpose of my study is to investigate the determinants of performance measure drop. Using the performance evaluation reports of 15 SOEs in Korea for 26 years, I find that (1) less discriminable performance measures are more likely to be dropped *in the performance evaluation system design process* than more discriminable performance measures; (2) the lower the objective performance score is, the more likely it is to be dropped; and (3) performance measures with participative target setting methods are less likely to be dropped than those with non-participative methods.

The results of my study provide practical implications. First, when designers/raters review the performance evaluation system periodically to determine the retention or

removal of performance measures, my findings provide useful insights into how performance measurement systems are designed and how incentives of designers/raters and ratees interact in the designing process. Second, by separating the *performance evaluation system design process* from the performance evaluation process itself, I am able to identify the different timings of influence activities conducted by ratees. The results show that objective measures with low scores are more likely to be dropped at the design stage. This is because the nature of objective measures makes it almost impossible for ratees to exercise influence during the evaluation process. This ratee's incentive, however, may be well aligned with the rater's incentive as designer/raters are partly responsible for low scores because of erroneous target setting or the failure of appropriate ratee motivation.

My study sheds light on the performance evaluation literature by exploring the determinants of performance measure drop using time series data. The evolution of the performance evaluation system was not fully explored because of the cross-sectional nature of prior studies. By examining time series data, it has been possible to separate *the performance evaluation system design process* from the performance evaluation process in my study. Additionally I focus on the measure-level characteristics rather than those of the firm-level. Finally, my study also investigates how the interaction between raters and ratees influence the performance measurement system design process, shedding light on the interplay between measurement characteristics, designer/raters,

and ratees.

Notwithstanding, my study is subject to many limitations. First, one must give caution in generalizing my findings because my results is based on only 15 public SOEs. Second, the quality of the variables used in my study might not be optimal due to limited data availability. I do not fully consider the firm level or government level control variables like new strategy implementation, business uncertainty, regulation change, and the change of government regime. Third, I measure the ratee's influence activities indirectly using the performance score level of the dropped measure because ratee's true influence activities are unobservable.

In spite of limitations discussed above, my study still provides meaningful insights on the dynamic nature of performance evaluation systems.

Instead of dropped performance measures, new performance measures will be an intriguing research perspective. Although I focus on the determinants of the performance measure drop, in future research, the researcher should also be interested in examining the characteristics of newly added performance measure from the perspective of designer/ratees and ratees incentive.

REFERENCES

Ahn, T. S, Y. S. Choi and D. H. Kwon. 2011. An Examination of Target-Setting: How a Rater revises a Ratee's Target. *The Korean Accounting Review* 36(1):183-223.[Printed in Korean]

Ahn, T. S, I Hwang, and M. I. Kim. 2010. The impact of performance measure discriminability on Ratee incentives. *The Accounting Review* 85(2): 289-417.

Ahn, T. S, Y. S. Choi and I Hwang. 2013. Impact on Ratchet Effect of Superior Subjectivity and Subordinate Influence in Target Setting. *Working Paper*. Seoul National University.

Antle, R. and J. S. Demski. 1988. The controllability principle in responsibility a ccounting. *The Accounting Review* 63(4): 700-718.

Allison, P. D. 2010. *Survival Analysis Using SAS: A Practical Guide, Second Edition*. SAS Institute Inc. Cary, NC.

Arya, A., J Glover, B. Mittendorf, and L. Ye. 2005. On the Use of Customized versus Standardized Performance Measures. *Journal of Management Accounting Research* 17: 7-21.

Baker, G. 2002. Distortion and Risk in Optimal Incentive Contracts. *The Journal of Human Resources* 37(4):728-751.

Banker, R. D. and S. M. Datar. 1989. Sensitivity, Precision, and Linear Aggregati on of Signals for Performance Evaluation. *Journal of Accounting Research* 27 (1):21-39.

Bies, R. J. 1987. Beyond voice: The influence of the decision-maker justification and sincerity in procedural fairness judgments. *Representative Research in Socia l Psychology* 17:3-14.

Bol, J. C. 2008. Subjectivity in Compensation Contracting. *Journal of Accounting Literature* 27:1-24

Bol, J. C. 2011. The Determinants and Performance Effects of Manager's Perform ance Evaluation Biases. *The Accounting Review* 86(5):1549-1575.

Du, F., G. Tang and S. M. Young. 2012. Influence Activities and Favoritism in Subjectivity Performance Evaluation: Evidence from Chinese State-Owned Enter prises. *The Accounting Review* 87(5):1555-1588

Feltham, G. and J. Xie. 1994. Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations. *The Accounting Review* 69(3):429-453.

Friedrich, J. 1993. Primary Error Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena. *Psychological Review* 100(2): 298-319

Frink, D. D. and G. R. Ferris. 1998. Accountability, Impression Management, and Goal Setting in the Performance Evaluation Process. *Human Relations* 51(10): 1259-1283

HassabElnaby, H. R, A. A. Said. and B. Wier. 2005. The Retention of Nonfinanc ial Performance Measures in Compensation Contracts. *Journal of Management Accounting Research* 17:23-42.

Harris, M. M. 1994. Rater Motivation in the Performance Appraisal Context: A Theoretical Framework. *Journal of Management* 20(4):737-756

Higgins, E. T. and J. A. Bargh. 1987. Social Cognition and Social Perception. *Annual reviews of Psychology*. 38:369-425

Holmstrom, B. 1979. Moral Hazard and Observability. *The Bell Journal of Econo mics*. 10(1): 74-91

Ittner, C. D., D. F. Larcker and M. V. Rajan. 1997. The Choice of Performance Measures in Annual Bonus Contracts. *The Accounting Review* 72(2): 231-255.

Ittner, C. D. and D. F. Larcker. 2001. Assessing Empirical Research in manageria l accounting: a value-based management perspective. *Journal of Accounting and Economics* 32: 349-410.

Ittner, C. D. and D. F. Larcker. 2002. Determinants of Performance Measure Choices in Worker Incentive Plans. *Journal of Labor Economics* 20(2): s58-S90.

Ittner, C. D., D. F. Larcker and M. W. Meyer. 2003. Subjectivity and the Weighting of Performance Measures: Evidence from a Balanced Scorecard. *The Accounting Review* 78(3): 725-758.

Ittner, C. D. and D. F. Larcker. 2009. Extending the Boundaries: Non-financial Performance Measures. *Handbook of Management Accounting Research*: 1235-1251.

Kaplan, R. and D. Norton. 1996. *The Balanced Scorecard: Translating Strategy into Action*. Harvard Business School Press, Boston, MA.

Kaplan, R. and D. Norton. 2001. *The Strategy-Focused Organization*. Harvard Business School Press, Boston, MA.

Kennerley, M. and A. Neely. 2003. Measuring performance in a changing business environment. *International Journal of Operations and Production Management* 23(2):213-229.

Lau, C. M. and A. V. Martin-Sardesai. 2012. The role of organizational concern for workplace fairness in the choice of a performance measurement system. *The British Accounting Review* 44. 157-172.

Lau, C. M, and S. L. C. Tan. 2005. The Importance of Procedural Fairness in Budgeting. *Advances in Accounting* 21. 333-356.

Leventhal, G. S. 1980. What should be done with equity theory? : New approaches to the study of fairness in social relationships. In K. J. Gergen, M.S. Greeberg and R. H. Willis, Social Exchange: *Advances in Theory and Research* (pp. 27-54). New York: Plenum.

Levy, R, E. and J, R, Williams. 2004. The Social Context of Performance Appraisal: A Review and Framework for the Future. *Journal of Management* 30(6): 881-905.

Libby, T. 1999. The influence of voice and explanation on performance in a participative budgeting setting. *Accounting, Organizations and Society* 24: 125-137.

Libby, T., S. E. Salterio and A. Webb. 2004. The Balanced Scorecard: The effects of Assurance and Process Accountability on Managerial Judgment. *The Accounting Review* 79(4):1075-1094.

Lipe, M. G. and S. E. Salterio. 2000. The Balanced Scorecard: Judgmental Effects of Common and Unique Performance Measures. *The Accounting Review* 75 (3): 283-298.

Luft, J. L. and M. D. Shield. 2002. Mapping Management Accounting: Graphics and Guidelines for Theory-Consistent Empirical Research. *Working Paper*, Michigan State University. Available at SSRN: http://ssrn.com/abstract=305959.

Malina, M. A. and F. H. Selto. 2004. Choice and change of measures in performance measurement models. *Management Accounting Research* 15:441-469.

Merchant, K. 2012. *Modern Management Control Systems*: Text and Cases. Prentice-Hall, Upper Saddle River, NJ.

Meyer, M. W. and V. Gupta. 1994. The Performance Paradox. *Research in Organizational Behavior* 16:309-369.

Milgrom, P. and J. Roberts. 1995. Complementarities and fit: strategy, structure, and organizational change in manufacturing. *Journal of Accounting Economics* 19(2/3): 179-208.

Mitchell, T.R. and O'Reilly, C.A. 1983. Managing poor performance and productivity in organizations. In K. Rowland & G. Ferris (Eds.), Research in personnel and human resources management (Vol. 1). Greenwich, CT: JAI Press.

Moers, F. 2005. Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations and Society* 30: 67-80.

Murphy, K. R. and J. N. Cleveland. 1995. Understanding Performance Appraisal – Social, Organizational, and Goal-Based Perspectives. SAGA Publication, Thousand

Oaks, CA.

Naiper, N. K. and G. P. Latham. 1986. Outcome expectancies of people who conduct performance appraisals. *Personnel Psychology* 39. 827-837

Prendergast, C. 1999. The Provision of Incentives in Firms. *Journal of Economic Literature* 37:7-63.

Rappaprot, A. 1999. New Thinking on how to link executive pay to performance. *Harvard Business Review* (March-April): 91-101.

Roberts, G. E. and T. Reed. 1996. Performance Appraisal Participation, Goal Sett ing and Feedback The influence of Supervisory Style. *Review of Public Person nel Administration*. 16(4): 29-60.

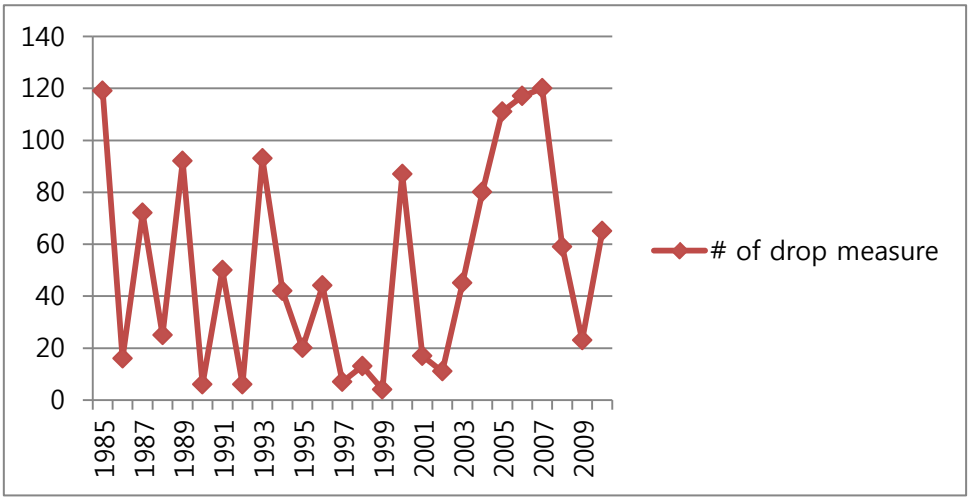
Rogers, W. 1993. Regression standard errors in clustered samples. *Stata Technical Bulletin* 13. 19-23.

Wentzel, K. The Influence of Fairness Perceptions and Goal Commitment on Ma nager's Performance in a Budget Setting. *Behavioral Research in Accounting* 14: 247-271.

FIGURE 1

The Change of the Number of Performance Measure Drop for 1985-2011

Panel A: The Number of Performance Measure Drop by Year



Panel B: The Relative Weight of the Number of Performance Measure Drop by Year

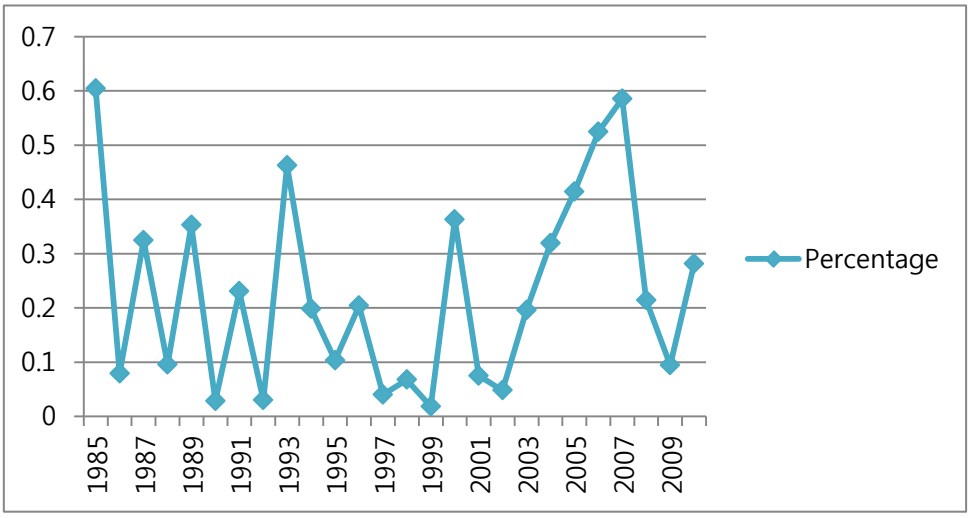


FIGURE 2

The Trend of the Ratio of the Number of Performance Measure Drop among the Total Number of Performance Measure Drop According to Performance Measure Tenure

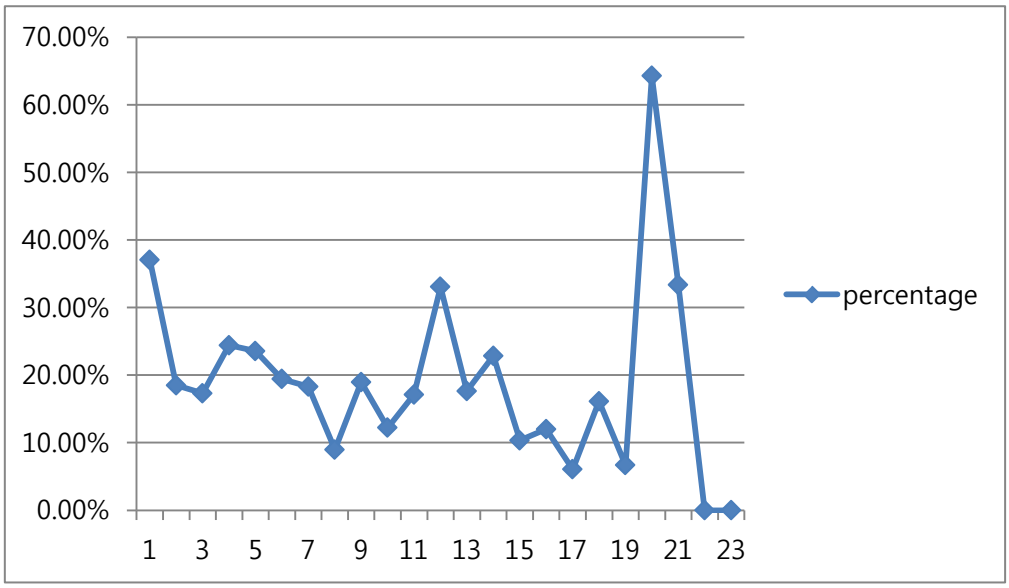


FIGURE 3

The Trend of the Number of Participative Target Measures for 1985-2011

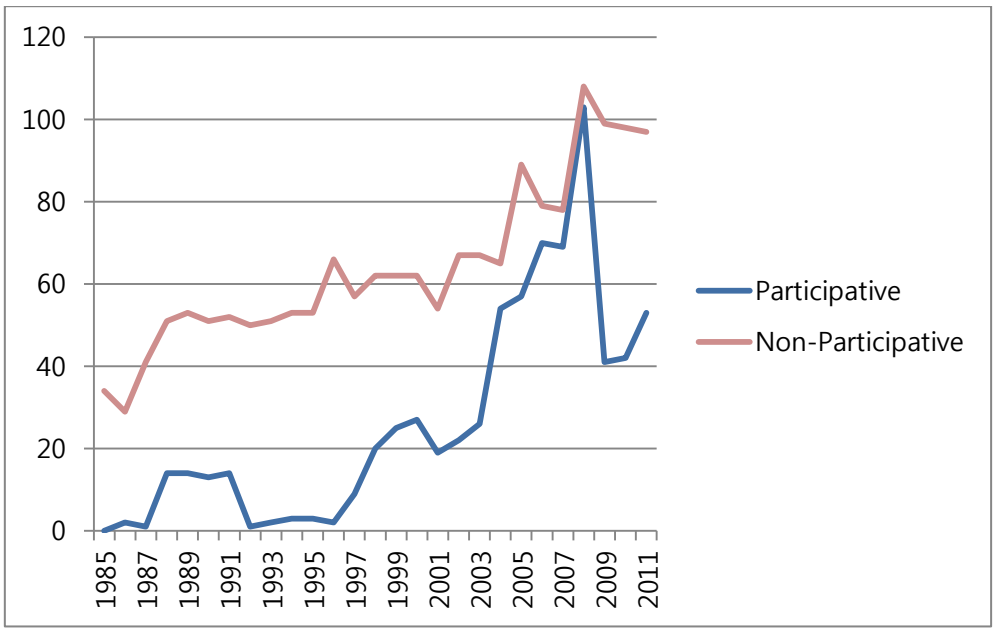


TABLE 1

Details on State Owned Enterprises in Korea

No.	State-Owned Companies	Government Ownership	Total Asset	Net Income	Capital	Roa
1	Korea Electric Power Corporation	21.17%	136,467,850	- 3,292,997	3,209,820	-2.41%
2	Korea Minting & Security Printing Corporation	100%	403,336	- 455	6,623	-0.11%
3	Korea Coal Corporation	100.00%	698,807	- 98,219	106,000	-14.06%
4	Korea Resource Corporation	99.80%	2,941,834	32,731	998,729	1.11%
5	Korea National Oil Corporation	100%	31,567,402	- 152,752	8,965,196	-0.48%
6	Korea Trade Investment Promotion Agency	100%	140,882	39,877	32,000	28.31%
7	Korea Highway Corporation	82.39%	49,279,362	103,174	23,856,515	0.21%
8	Korea National Housing Corporation	89.60%	77,744,381	116,959	10,548,684	0.15%
9	Korea Water Resources Corporation	91.20%	23,429,915	293,267	6,694,987	1.25%
10	Korea Land Corporation	73.34%	50,486,451	65,912	5,000,000	0.13%
11	Korea Rural Community & Agricultural Corporation	100%	7,113,120	92,306	1,305,029	1.30%

12	Agricultural and Fishery Marketing Corporation	100%	1,168,803	11,195	86,200	0.96%
13	Korea National Tourism Organization	55.27%	1,305,930	32,787	47,272	2.51%
14	Korea Railroad Corporation	100%	22,179,329	450,664	9,581,417	2.03%
15	Korea Land and Housing Corporation	86.60%	158,473,034	790,513	22,417,592	0.50%

1. Unit: Million Won, Percent and All data are prepared based on K-IFRS
2. Data are as of December 31, 2011 except for Korea National Housing Corporation and Korea Land Corporation
3. The information of Korea National Housing Corporation and Korea Land Corporation were as of September 30, 2009 because two companies were merged at the end of September, 2009

TABLE 2

An Example of Performance Ratings of an SOE

Category	Individual Measure	Measurement	Target Setting Methods	Weight	Rating	Score	Score rate
1. Leadership and	(1) Leadership	Subjective	RANK	5	B+	4	80
Responsibility	(2) Responsibility Management	Subjective	RANK	3	A0	2.7	90
	(3) Evaluation from Public	Objective	ASSIGN	5	A+	4.993	99.86
	(4) Social Contribution						
	(a) Social Contribution	Subjective	RANK	2	B+	1.6	80
	(b) Conform to Government Policy	Objective	ASSIGN	5	A0	4.264	85.28
2. Management	(1) Operation Efficiency						
Efficiency	(a) Labor Efficiency	Objective	ASSIGN	4	B0	2.629	65.735
	(b) Capital Productivity	Objective	ASSIGN	4	B0	2.681	67.028
	(2) Organization and Human Resource Management	Subjective	RANK	4	B+	3.2	80
	(3) Financial Budget Management and Performance						
	(a) Financial Budget Management	Subjective	RANK	4	B0	2.8	70
	(b) Financial Budget Performance						
	- Total Asset Turnover Rate	Objective	ASSIGN	2	A+	1.851	92.544

	- Interest Coverage Rate	Objective	ASSIGN	2	B0	1.201	60.046
	- Debt Ratio	Objective	ASSIGN	2	D+	0.84	41.996
	(c) Qualitative Management Expense	Objective	ASSIGN	2	B+	1.469	73.434
	(4) Compensation and Performance Management						
	(a) Compensation and Performance Management	Subjective	RANK	4	E0	0.8	20
	(b) Increasing rate of Total Salary Expense	Objective	ASSIGN	4	A+	4	100
	(5) Labor Relation Management	Subjective	RANK	3	A0	2.7	90
3. Core Business	(1) Electricity and Power Demand Management	Subjective	RANK	4	B0	2.8	70
	- Maintaining High Load Factor	Objective	ASSIGN	5	A+	5	100
	- Global Benchmark for High Load Factor	Objective	BENCHMARKING	3	A+	3	100
	(2) Power Transmission and Transformation	Subjective	RANK	4	B+	3.2	80
	- Dysfunction Management for Transmission	Objective	ASSIGN	6	A+	6	100
	- Efficiency Index for Transmission	Objective	ASSIGN	5	A0	4.382	87.648

	Facility						
	(3) Electric Power Distribution	Subjective	RANK	4	B+	3.2	80
	- Right Voltage Management	Objective	ASSIGN	4	A0	3.979	99.467
	- Global Benchmark for Loss rate of transmission	Objective	BENCHMARKING	3	A+	3	100
	- Development of Distribution Management	Objective	ASSIGN	4	A+	4	100
	(4) New Business Development	Subjective	RANK	3	A0	2.7	90
4. Objective Measures in Total				40		29.7	
5. Subjective Measures in Total				60		53.289	
6. Total				100		82.989	

1. This table provides the performance evaluation results from 2011 for a SOE.
2. The evaluation and target setting methods of objective performance measures include trend analysis (TREND), beta distribution analysis (BETA), assigning target process (ASSIGN), and negotiating target process(NEGOTIATE), the subjective performance measures are evaluated by the single evaluation method of ranking analysis.
3, Score rate is a final evaluation result calculated based on evaluation rule by rater.
4. Weight is the relative importance of an individual performance measure ex ante determined by Raters at the beginning of year t.
5. Rating is a rater's evaluation results of individual performance measure under a certain assumption such as distribution of performance

TABLE 3

The Number of Performance Measure's Drop by Year

Year	# of Observation	# of dropped measures	Percentage	# of drop objective measure	% of Objective Measure	# of drop subjective measure	% of Subjective Measure
1985	200	121	60.50%	8	6.61%	113	93.39%
1986	205	18	8.78%	10	55.56%	8	44.44%
1987	225	72	32.00%	19	26.39%	53	73.61%
1988	267	25	9.36%	3	12.00%	22	88.00%
1989	267	93	34.83%	1	1.08%	92	98.92%
1990	218	6	2.75%	0	0.00%	6	100.00%
1991	222	50	22.52%	13	26.00%	37	74.00%
1992	206	6	2.91%	4	66.67%	2	33.33%
1993	207	93	44.93%	3	3.23%	90	96.77%
1994	218	42	19.27%	41	97.62%	1	2.38%
1995	201	20	9.95%	19	95.00%	1	5.00%
1996	222	43	19.37%	33	76.74%	10	23.26%
1997	187	7	3.74%	5	71.43%	2	28.57%
1998	206	14	6.80%	1	7.14%	13	92.86%
1999	238	5	2.10%	1	20.00%	4	80.00%
2000	258	104	40.31%	53	50.96%	51	49.04%
2001	227	17	7.49%	16	94.12%	1	5.88%
2002	230	11	4.78%	4	36.36%	7	63.64%
2003	230	45	19.57%	3	6.67%	42	93.33%
2004	251	80	31.87%	39	48.75%	41	51.25%
2005	274	114	41.61%	19	16.67%	95	83.33%
2006	227	118	51.98%	30	25.42%	88	74.58%
2007	211	126	59.72%	20	15.87%	106	84.13%
2008	236	59	25.00%	23	38.98%	36	61.02%
2009	244	23	9.43%	10	43.48%	13	56.52%

2010	231	65	28.14%	13	20.00%	52	80.00%
Total	5908	1377	23.31%	391	28.40%	986	71.60%

This table reports the number of performance measure drop at the end of each year from 1985 to 2010.

Column (3) indicates the total number of performance measure drop on a year basis.

Column (5) indicates the number of the objective performance measure drop on a year basis and Column (7) indicates the number of the subjective performance measure drop on a year basis

TABLE 4

Descriptive Statistics of the Sample [a]

Variables	N	Mean	Standard Deviation	Min	Q1	Median	Q3	Max
Drop	5,908	0.233	0.423	-	-	-	-	1.000
Discriminability	6,139	0.098	0.071	-	0.046	0.080	0.125	0.602
M_Tenure	6,139	4.923	4.270	1.000	2.000	4.000	7.000	23.000
Weight	6,139	3.315	2.220	0.500	2.000	3.000	4.000	20.000
Attain	6,139	0.808	0.161	-	0.750	0.870	0.935	1.000
Ind_Sub	6,139	0.716	0.451	-	-	1.000	1.000	1.000
Participative	1,745	0.386	0.487	-	-	-	1.000	1.000
Ln_Asset	5,686	21.493	1.974	17.099	20.068	21.486	23.207	25.789
Roa	5,419	0.012	0.064	- 0.469	0.0015	0.008	0.028	0.315
Lev	5,419	0.627	0.317	0.058	0.4228	0.590	0.781	2.168

The sample includes 6,139 unique measure-years and 15 unique State Owned Enterprise in Korea from 1985 to 2011. Data for performance measure's score and characteristics are manually collected from the performance evaluation reports of SOE. Financial performance data and firm characteristics are obtained from Alio web site. The number of observations varies depending on data availability.

a. Variable Definitions

Drop=	1 if measure j of firm i disappears at year t+1, otherwise 0
Discriminability=	standard deviation of attain at measure j and year t
M_Tenure=	period after measure j comes up in performance measure at first
Weight =	points allocated to individual performance measures, adding up to 100 points each year
Attain =	percentage of performance score which individual performance measure achieves
Ind_Sub=	1 if measure is subjective, 0 if measure is objective
Participative=	1 if target method allows ratee's participation, and otherwise 0
Ln_Asset =	log value of total asset of firm i at year t
Roa =	the ratio of operating income divided by total asset of firm i at year t
Lev =	the ratio of total debt divided by total asset of firm i at year t

TABLE 5

Pearson Correlation Matrix [a]

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
(1) Drop	1	-0.009	-0.078***	-0.070***	-0.052***	0.061***	0.024	-0.001	0.027*	-0.006
(2) Discriminability		1	0.160***	0.136***	-0.342***	-0.366***	0.023	0.162***	-0.033**	0.008
(3) M_Tenure			1	0.165***	-0.175***	0.071***	-0.348***	0.054***	-0.032**	0.012
(4) Weight				1	0.047***	-0.355***	-0.128***	0.015	-0.021	0.011
(5) Attain					1	-0.153***	0.050**	-0.004	0.118***	-0.094***
(6) Ind_Sub						1	.	-0.085***	0.026*	-0.007
(7) Participative							1	0.101***	-0.018	0.011
(8) Ln_Asset								1	-0.061***	-0.142***
(9) Roa									1	-0.510***
(10) Lev										1

a. This table presents Pearson correlations in the upper diagonal. The number of observations varies depending on data availability

*, **, *** indicate significance at less than the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed t-tests on mean differences

TABLE 6
Logistic Regressions Estimating the Determinants of Performance Measure Drop

		Dependent variable = Drop	
Independent Variables	Predicted Sign	(1) Coefficient	(2) Coefficient
Intercept		1.167***	0.648**
Discriminability	-	-3.580***	-2.744***
Attain		-0.835***	-0.624**
Weight		-0.104***	-0.087***
M_Tenure		-0.063***	-0.068***
Ind_Sub			0.306***
Ln_Asset		-0.019*	-0.018
Roa		0.550	0.484
Lev		0.069	0.075
Year Fixed Effect		Yes	Yes
Number of Observation		5188	5188
Pseudo R^2		0.272	0.274

This table reports the logistic regression results of the effect of *discriminability* on the performance measure drop by testing hypothesis 1 using equation (1)

$$Drop_{j,i,t+1} = \alpha_0 + \beta_1 * Discriminability_{j,t} + \beta_2 * Attain_{j,i,t} + \beta_3 * Weight_{j,i,t} + \beta_4 * M\_Tenure_{j,i,t} + \beta_5 * Ind\_Sub_{j,i,t} + \beta_6 * ln\_Asset_{i,t} + \beta_7 * Roa_{i,t} + \beta_8 * Lev_{i,t} + Year\ dummies + \varepsilon_t$$

where i, j and t indicate the firm, the individual performance measure and the time period respectively.

Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to the paper for a detailed explanation of these tests.

TABLE 7
Logistic Regression Estimating the Determinants of Objective Performance Measure Drop

		Dependent variable = Drop		
Independent Variables	Predicted Sign	(1) Coefficient	(2) Coefficient	(3) Coefficient
Intercept		-1.493***	-1.074*	-1.178**
Discriminability	-	-0.169	-0.396	0.662
Low_Score	+	0.913***	0.844***	1.545***
Participative	-		-0.529***	-0.528***
*Low_Score*Discrimiability*	-			-4.161*
Weight		-0.074*	-0.065	-0.058
M_Tenure		0.029	-0.001	0.001
Ln_Asset		-0.027	-0.031	-0.034
Roa		1.469**	1.017	0.796
Lev		0.164	0.154	0.128
Year Fixed Effect		Yes	Yes	Yes
Number of Observation		1506	1506	1506
Pseudo R^2		0.334	0.339	0.344

This table reports the logistic regression results of the effect of *attain* on the objective performance measure drop by testing hypothesis 2 and 3 using equation (2). We define Low_Score as the lowest quartile based on the performance score. Thus, Low_Score takes the value of one if performance score is within the lowest quartile and zero if not.

$$Drop_{j,i,t+1} = \alpha_0 + \beta_1 * Discriminability_{j,t} + \beta_2 * Low\_Score_{j,i,t} + \beta_3 * Participative_{,i,t} + \beta_4 * Weight_{j,i,t} + \beta_5 * M\_Tenure_{j,i,t} + \beta_6 * Low\_Score_{j,i,t} * Discriminability_{j,i,t} + \beta_7 * ln\_Asset_{i,t} + \beta_8 * Roa_{i,t} + \beta_9 * Lev_{i,t} + Year\ dummies + \varepsilon_t$$

where i, j and t indicate the firm, the individual performance measure and the time period respectively

Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to the paper for a detailed explanation of these tests.

TABLE 8

Additional Logistic Regression Estimating the Determinants of Objective Performance Measure Drop

	Dependent variable = Drop	
Independent Variables	Predicted Sign	Coefficient
Intercept		-9.735***
High_Perform	?	-0.621
Low_Perform	+	2.018***
M_Tenure		-0.09
Weight		-0.093
Ln_Asset		-0.141*
Roa		3.795
Lev		0.797*
Year Fixed Effect		yes
Number of Observation		375
Pseudo R^2		0.618

This table reports the logistic regression results of the effects of the performance score (*High_Perform or Low_Perform*) within the lowest quartile of discriminability on the objective performance measure drop by testing hypothesis 2 using following equation. In this analysis, we define the *High_Perform (Low_Perform)*as the top (lowest) quartile based on the performance score within the lowest quartile of discriminability. Thus, *High_Perform* (*Low_Perform*) takes the value of one if performance score is within the top (lowest) quartile and zero if not among the lowest quartile based on discriminability level of a performance measure.

$$Drop_{j,i,t+1} = \alpha_0 + \beta_1 * High\_Perform_{j,I,t} + \beta_2 * Low\_Perform_{j,i,t} + \beta_3 * M\text{-}Tenure_{j,i,t} + \beta_4 * Weight_{j,i,t} + \beta_5 * ln\_Asset_{i,t} + \beta_6 * Roa_{i,t} + \beta_7 * Lev_{i,t} + Year\ dummies + \varepsilon_t$$

where i, j and t indicate the firm, the individual performance measure and the time period respectively.

Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to the paper for a detailed explanation of these tests..

TABLE 9
Performance Score Comparison Between Added And Dropped (Non-Added) Measure

Panel A

Variables	Added	Dropped	Difference
Attain	0.876	0.807	-0.070***

Panel B

Variables	Added	Non-Added	Difference
Attain	0.876	0.855	-0.022*

Panel A (Panel B) of this table shows the univariate test results of comparing performance scores (attain) between newly added performance measures and dropped (Non-Added) performance measures. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to the paper for a detailed explanation of these tests.

TABLE 10

Logistic Regression Estimating Discriminability on the Performance Measure Drop After Controlling the Effect of Macroeconomic Factors

Panel A : Full Sample

		Dependent variable = Drop		
Independent Variables	Predicted Sign	(1) Coefficient	(2) Coefficient	(3) Coefficient
Intercept		-15.536***	-1.658***	10.495***
Discriminability	-	-2.744***	-2.407***	-2.407***
Attain		-0.624**	-0.673**	-0.673**
Weight		-0.087***	-0.083***	-0.083***
M_Tenure		-0.068***	-0.068***	-0.068***
Ind_Sub		0.306***	0.306***	0.306***
Ln_Asset		-0.018	-0.02**	-0.020**
Roa		0.484	0.550	0.550
Lev		0.075	0.036	0.039
GDP		0.001***		-0.001***
GDP_growth			23.209***	43.719***
Year Fixed Effect		Yes	Yes	Yes
Number of Observation		5188	5058	5058
Pseudo R2		0.274	0.263	0.263

Panel B : Sub-Sample using Objective Measure

		Dependent variable = Drop		
Independent Variables	Predicted Sign	(1) Coefficient	(2) Coefficient	(3) Coefficient
Intercept		-4.521**	-1.322**	19.198***
Discriminability	-	0.662	0.556	0.557
Low_Score	+	1.545***	1.444***	1.444***
Participative	-	-0.528***	-0.531***	-0.531***

*Low_Score*Discriminability*	-	-4.161*	-3.792*	-3.792*
Weight		-0.058	-0.068	-0.068
M_Tenure		0.001	0.002	0.002
Ln_Asset		-0.034	-0.046**	-0.046**
Roa		0.796	0.718	0.718
Lev		0.128	0.085	0.085
GDP		0.000		-0.002***
GDP_growth			4.909*	39.540***
Year Fixed Effect		Yes	Yes	Yes
Number of Observation		1506	1486	1486
Pseudo R2		0.344	0.345	0.345

Panel A of the table reports the additional regression results of the effect of *discriminability* on the performance measure drop after controlling the macro-economic factor regarding H1. We incorporate GDP and GDP_growth as control variables to capture the macro-economic factors. we get Korean GDP data from Economic statistics system of the Bank of Korea

$$Drop_{j,i,t+1} = \alpha_0 + \beta_1 * Discriminability_{j,t} + \beta_2 * Attain_{j,i,t} + \beta_3 * Weight_{j,i,t} + \beta_4 * M\text{-}Tenure_{j,i,t} + \beta_5 * Ind\_Sub_{j,i,t} + \beta_6 * ln\_Asset_{i,t} + \beta_7 * Roa_{i,t} + \beta_8 * Lev_{i,t} + \beta_9 * GDP_t + \beta_8 * GDP\_growth_t + Year\ dummies + \varepsilon_t$$

where i, j and t indicate the firm, the individual performance measure and the time period respectively

Panel B of the table reports the additional regression results of the effect of *attain and participative* on the performance measure drop after controlling the macro-economic factor regarding H1 and H2.

$$Drop_{j,i,t+1} = \alpha_0 + \beta_1 * Discriminability_{j,t} + \beta_2 * Low\_Score_{j,i,t} + \beta_3 * Participative_{,i,t} + \beta_4 * Weight_{j,i,t} + \beta_5 * M\_Tenure_{j,i,t} + \beta_6 * Low\_Score_{j,i,t} * Discriminability_{j,i,t} + \beta_7 * ln\_Asset_{i,t} + \beta_8 * Roa_{i,t} + \beta_9 * Lev_{i,t} + + \beta_{10} * GDP_t + \beta_{11} * GDP\_growth_t + Year\ dummies + \varepsilon_t$$

where i, j and t indicate the firm, the individual performance measure and the time period respectively

Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. the symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to the paper for a detailed explanation of these tests.

Essay 2

The Change of Performance Measure Weights and Ratee's Incentive

I. Introduction

It is common for many organizations to adopt a financial incentive scheme based on performance and to implement a performance evaluation system in order to motivate agents to exert more effort. This performance evaluation system is based on agency theory in that it is rational for the principal to need to use appropriate performance measures for incentive contracts when it is difficult for that principal to monitor the agent's behavior. Prior studies show that good performance measures should be informative (Holmstrom 1979), be goal congruent (Feltham and Xie 1994) and have precision and sensitivity (Banker and Datar 1989). Other studies indicate that monetary incentives foster an effort-performance relationship by motivating the agent to exert more effort (Bonner and Sprinkle 2002). Therefore, if designers or raters use appropriate performance measures for incentive contracts, ratees will increase their own effort levels to achieve better performance. Moreover, because ratees perform multiple tasks at the same time, rather than just a single task, to maximize their benefits they should allocate their efforts among multiple tasks differently based on the characteristics of the performance measures (Holmstrom and Milgrom 1991). When ratees perform multiple tasks under incentive contracts, they would allocate their effort among individual tasks (performance measures) at the extent to which the marginal benefits to the effort of individual tasks (performance measures) are equal. In addition, it is possible that commonality or task-specific ability of a performance measure could influence ratees in

their effort allocation among multiple tasks (performance measures) because these factors could affect the effort-benefit sensitivity of the ratees. Therefore, my study investigates whether ratees reallocate their effort among multiple tasks when the designer of performance evaluation system (hereafter 'designer') changes performance measure weights under a multi-task agency environment. I also examine how ratees differently respond to the change of a performance measure weight based on commonality and task-specific ability of the performance measure.

First, I examine whether ratees respond by changing their effort level in accordance with performance measure weight adjustment. Second, I investigate how ratees differently respond to the adjustment of performance measure weights on the basis of commonality of the performance measure. Third, I also examine how ratees differently respond to the adjustment of performance measure weights depending on task-specific ability.

The results are as follows. I find that when the designer changes the performance measure weight, ratees adjust their effort on the performance measure in accordance with the changes. In other words, ratees exert more effort on the performance measure with the increased weight. These results support the multi-task agency theory that states that ratees will exert more effort on the performance measure with the increased weight because ratees can benefit more from exerting the same level of effort (Brickley, Smith and Zimmerman 2009). In addition, I find that when a performance measure weight

increases, ratees respond by exerting more effort in common measures than in unique measures. These results are consistent with the prior studies that find that ratees input more effort on performance measures that can be subject to relative performance evaluation (Frederickson 1992; Matsumura and Shin 2006) and that the presence of relative performance information increases performance (Tafkov 2013). I also find that when a performance measure weight increases, ratees exert more effort on the performance measures where they have a high level of task-specific ability than those of low task-specific ability.

My study sheds light on the performance evaluation literature by providing clearer empirical evidence that ratees respond in accordance to the change of performance measure weights, which is in line with Brickley and Zimmerman (2001). Compared with prior studies, my study shows a more clear causal relationship between the specific change of performance measure weight and the change of effort level which ratees exert by accurately capturing the weight change of performance measures weight exactly and by controlling other confounding compounding variables.

My study also contributes to the performance evaluation practice in several ways. First, the change of performance measure weight is found to be a useful mechanism to motivate ratees because the designer could intentionally increase the performance measure weights to motivate ratees to exert more effort on a certain task that the designer places focus upon. Second, my findings also provide empirical evidence that

the feedback of relative performance information influences ratee's effort allocation. Moreover, my study implies that the increase of a performance measure weight is not enough to motivate ratees to exert more effort on a unique measure. Thus, the designer needs to consider other mechanisms to motivate the ratee to exert more effort on a unique measure. Finally, when the ratee has high task-specific ability of a certain performance measure, that ratee is likely to exert more effort on that performance measure rather than on the performance measure with low task-specific ability in accordance with the increase of the performance measure weight. Therefore, the designer must consider how to improve task-specific abilities before adjusting the weights of a performance measure with low task-specific ability.

The remainder of the paper is organized as follows. In Section II, I explain the research setting of SOE performance evaluation system in Korea. Then, I review related literature and develop my hypotheses in Section III. In Section IV, I describe my research design and empirical model, and then present descriptive statistics. The empirical results of my study are the focus of Section V. To check the robustness of my model, I perform additional tests in Section VI. Finally, Section VII concludes.

II. Research Setting

2.1. Overview of the SOE Performance Evaluation System

The Korean Government passed the Law of Management of SOE in 1984 to evaluate the performance of the State-Owned Enterprises (hereafter, SOEs or ratees) where the government owns over 50% of shareholder's equity. The purposes of evaluating SOE performance are to enhance the ownership spirit of executive officers and staff within SOEs, to introduce a business mindset based on autonomy and responsibility, and to implement an appropriate performance based compensation system (Korean Society of Public Enterprise 2003). As of December 31, 2011, there were 15 SOEs engaged in utilities, mining, security, tourism, printing, and other various fields. Table 1 shows the details of these SOEs.

[INSERT TABLE 1 ABOUT HERE]

The performance evaluation of SOEs is performed annually on both SOEs and the CEO of SOEs. To evaluate SOEs independently, the government (hereafter, designer) nominates evaluation committee members (hereafter, raters) composed of academic professors, CPAs, professional consultants, and researchers. They report the performance evaluation results of SOEs to the National Assembly and the president of Korea annually (Ahn, Choi and Kwon 2011).

The structure of the performance evaluation of SOEs is based on the concept of the Balanced Scorecard (Kaplan and Norton 1996). The designer implements the performance measures which are developed from Financial, Customer, Internal Process, and Learning and Growth perspectives to execute the strategy of SOEs.

Table 2 shows that the structure of the performance evaluation system of SOEs consists of three main categories: Leadership and responsibility, management efficiency and core business which has sub-components of human resource, financial/budget management and other operation management. The third column in table 2 indicates that performance measures are composed of objective and subjective measures. Objective performance measures are evaluated on the basis of 4 different target setting methods: (1) actual to target analysis (2) target-range assignment analysis (3) trend analysis and (4) bet distribution analysis (Ahn, Hwang and Kim 2010). On the other hand, subjective performance measures have a single evaluation method: a 9 grade scale (A+ ~ E0), in which raters are able to grade entirely based on their own judgment. Weight indicates the relative importance of an individual performance measure, with all measures accumulating to 100 points. Finally, point is the number of points earned by ratees, and is calculated as the product of the Weight and the Performance Score.

In my research setting, I divide the performance evaluation cycle into two stages; the *performance evaluation system design process* and the *performance evaluation process*. First, both the designer and raters participate together in selecting performance measures

or adjusting performance measure weights before beginning performance evaluation in the *performance evaluation system design process*. Once the performance measures are chosen, raters review the reports prepared by ratees and evaluate ratees in the *performance evaluation process* based on the performance evaluation guideline which designer prepared. After completing performance evaluation, ratees provide feedback on performance results to the designer so that this feedback information can be incorporated into subsequent *performance evaluation system design processes*.

Ratees could receive a diverse compensation package on the basis of the performance evaluation results. After performance evaluation is completed, the designer discloses performance evaluation results, such as the SOE's total score and its ranking, to the public. In order to motivate SOEs, the government recommends that the President reward SOEs which have shown excellent performance, discharge the CEO of SOEs that have had bad performance, and determine how much incentive bonus will be paid to the SOEs based on the evaluation results. Specifically, based on final grade of performance evaluation, the CEOs of SOEs could obtain an incentive bonus of 0 to 200% of their monthly base salary and the employees of SOEs could receive an incentive bonus of 200 to 500% of their monthly base salary. Therefore, ratees have the incentive to exert heavy effort on the performance measure to maximize the performance score. As an example of a compensation scheme, Table 3 shows the incentive bonus structure for SOEs in 2008 (Ahn, Choi and Hwang 2010).

[INSERT TABLE 2 AND 3 ABOUT HERE]

2.2. Descriptive Statistics of Changes in the Performance Measure Weights

The Structure of Performance Evaluation of SOEs has evolved for over 25 years to reflect the changes in the business environment, the business objectives of SOEs, and government policies. Thus, the composition and relative weights of performance measures have also been changed over time (Park, 2003).

Table 4 shows descriptive statistics of changes in the performance measure weights by year. This table indicates how many performance measure weights were adjusted and how much weight of the performance measures were adjusted on a yearly basis. The number of measure weights that were increased is 2,045 (19%) in total. In contrast, the number of weights that were decreased is 2,575(24%) among 10,948 measures in the total sample. Analyzed based on performance measure characteristics, the number of common (unique) measures with increased weights is 1,129 (916). In contrast, the number of common (unique) measures with decreased weights is 1,535 (1,040).

[INSERT TABLE 4 ABOUT HERE]

2.3. The Number of Performance Measure Weight Changes Over Time

Figure 1 shows the pattern of the performance measure weight changes over time. Panel A of Figure 1 shows the number of performance measures weight changes over time by dividing the total sample into 2 groups*: the performance measures with increased weights* and *the performance measures with decreased weights* over time. Panel B and Panel C, respectively, show that the historical pattern of the common and unique measures with weight changes over time.

[INSERT FIGURE 1 ABOUT HERE]

III. Theory and Hypothesis Development

3.1. Effort Allocation under Multi-Task Agency Environments

When a principal hires an agent to perform a specific task for the benefit of the principal, the agency problem can occur where the agent maximizes one's own benefit rather than that of the principal. The reason why this happens is because it is difficulty for the principal to monitor the agent's behavior. Agency theory predicts that if the principal has a performance measure which is informative in capturing the agent's effort level, incentive contracts between the principal and agent could mitigates this agency problem (Holmstrom 1979; Thiele 2007; Brickley, Smith and Zimmerman 2009).

From a different perspective, performance contingent monetary incentive contracts can motivate agents to exert more effort and improve task performance (Bonner and Sprinkle 2002; Banker et al 1996; Banker et al 2001). Banker et al (1996) confirms that stores that adopt performance based incentives increase sales volume and that these effects persist over time when the incentive system is implemented. In a subsequent study, Banker et al (2001) finds that after the implementation of a pay-for-performance plan, performance improvement can result both from the selection effect where more productive employees join the firm and from the effort effect where employees exert more effort to be productive.

It is common that an agent is responsible to perform multiple tasks which contribute differently to the firm's objectives (Thiele 2007). Under a multi-task agency

environment, the agent can decide on how to allocate one's efforts across all relevant work as well as one's effort intensity (Holmstrom and Milgrom 1991; Feltham and Xie 1994; Datar et al 2001; Thiele 2007; Brickley, Smith and Zimmerman 2009). Multi-task agency theory predicts that when an agent performs multiple tasks on behalf of the principal under incentive contracts, the agent would choose an effort level of individual tasks to the extent of which the marginal benefits to the effort of individual tasks are equal (Dewatripont, Jewitt and Tirole 2000). If marginal benefits of individual tasks are not equal, the agent is likely to exert more effort on the task which has a relatively higher marginal value and to exert less effort on the task which has a relatively lower marginal value. Thus, based on this theoretical prediction, to motivate the agent to exert more effort on a specific task, the principal could increase the sensitivity of benefit-to-effort (the relative weight) on the specific task or decrease the sensitivity of benefit-to-effort (relative weight) of other tasks (Brickley, Smith and Zimmerman 2009).

However, it is not easy to motivate the agent to exert balanced effort among multiple tasks appropriately due to the characteristics of performance measures. Holmstrom and Milgorm(1991) indicate that there is an important interaction effect between the performance incentive for one task and that for another task. Specifically, when the principal offers high-powered incentives on the task which is easy to be measured but offers low-powered incentive on another task which is difficult to be measured, the agent would allocate more effort on the task which is expected to earn a higher bonus

but discard or allocate less effort on the task which is expected to earn a lower bonus.

The efficiency of the agent's effort allocation could be improved by configuring multiple performance measures in an optimal combination (Feltham and Xie 1994; Datar et al 2001). Specifically, Feltham and Xie (1994) argue that multiple performance measures, which are intended to capture a multi-dimensional representation of the agent's actions, should be goal congruent to influence the direction and intensity of the agent's effort in the right way. If performance measures are incongruent, they do not capture the individual contribution of each task which the agent performs and thus incentive contracts force agents to implement inefficient effort allocations among relevant tasks. Similarly, Datar et al (2001) contend that even when the sensitivity between an agent's effort and performance measure increases, it is not necessary to increase the relative weight of the performance measure because, in some cases, the combination of an increased compensation weight and increased performance measure sensitivity will motivate the agent to allocate too much effort toward relatively unprofitable actions.

Consequently, it is an important empirical question to investigate whether agents reallocate effort among multiple tasks under a multi-task agency environment when performance measure weights are changed. In addition, it is a valuable research question to examine how agents differently respond to the performance measure weight change based on measure characteristics.

3.2. The Change of Performance Measure Weights

Research which examines the relation between the change of performance measure weights and ratee's effort allocation is very important, but prior empirical studies in this area are scarce (Brickley and Zimmerman 2001). According to Brickley and Zimmerman (2001), professors at universities are responsible for two tasks: teaching and research. Since it is more difficult to quantify teaching quality than research output, promotion decisions are traditionally made on the basis of research performance. Schools, however, could adjust their incentive schemes to improve the quality of teaching. They show that the incentive scheme adopted by the William E. Simon Graduate School of Business Administration at the University of Rochester after 1991 led to a significant improvement in teaching quality. In contrast, research output declined steadily after the implementation of this incentive scheme. These kinds of results are consistent with the prediction of multi-task agency theory which states that agents respond by reallocating their own effort level across various tasks in accordance with the change of performance measure weights. In line with Brickley and Zimmerman (2001), Ahn, Choi, and Seo (2009) examine whether the relative weight of multiple performance measures affects a CEO's strategic effort allocation. Using the performance evaluation data of Korean SOEs, they find that the CEO has the incentives to exert more effort on performance measures which capture both firm and CEO performance rather than those which capture firm performance only. These results

provide implication that if CEOs exert more effort to obtain their own benefits rather than the firm's benefits, the agency problem would be exacerbated between CEOs and SOEs.

Nevertheless, some limitations still exist in prior studies. Both studies do not show the relations between changes in weight and changes in the ratee's effort level explicitly. Specifically, Ahn, Choi, and Seo(2009) does not capture the difference of relative weights between performance measures for both the CEO and the firm and those for the firm only. Moreover, this study does not examine changes, but just focuses on the level analysis. Similarly, Brickley and Zimmerman (2001) do not capture the change of relative weight between teaching and research explicitly because there is implicit weight of teaching and research performance measures. Therefore, it is difficult to draw an inference on the conclusions about the positive relationship between weight changes and the agent's effort allocation from the results of both studies.

To address this interesting and important question, I investigate whether agents reallocate their own efforts across multiple tasks when the designer adjusts the performance measure weights explicitly, using rich and unique performance evaluation results of Korean SOEs. Based on theory and prior studies, I conjecture that when a performance measure weight increases, ratees would respond by exerting more effort on the performance measure. On the contrary, when a performance measure weight decreases, ratees would respond by discarding or exerting less effort on the performance

measure. Thus, I hypothesize:

Hypothesis 1. When a performance measure weight is changed, ratees will change their own efforts on the performance measure in the same direction.

3.3. The Commonality of Performance Measures

The characteristics of performance measures are important elements in the design of the performance evaluation system because they influence the ratee's motivation. My study focuses on the commonality among performance measure characteristics. Performance measures consist of two types of performance measures based on whether they are applied for multiple SOEs or they are used only for individual SOEs.

First, from an economic point of view, common measures are used for Relative Performance Evaluation (thereafter "RPE") which entails information about the performance of a group of peers when evaluating the performance of specific individuals, teams, or organizational units. On the other hand, unique measures are used for performance evaluation of individual ratees.

Prior studies indicate that RPE is effective in motivating ratees to achieve greater performance because RPE reduces common uncertainty in performance evaluation. Specifically, Holmstrom (1992) contends that when the principal cannot observe the effort of multiple agents under common uncertainty in the operating environment, RPE has the benefit of reducing risk in optimal incentive contracting. Fredrickson (1992) supports both economic and behavioral theory regarding RPE by showing that the subject's effort levels increases significantly as the degree of common uncertainty increases with the RPE contacts, but not with the non-RPE contract. In addition, the subject's effort levels were higher under the RPE contract than under the non-RPE

contract regardless of the level of common uncertainty. Matsumura and Shin (2006) also finds that financial performance improves after the implementation of an incentive plan that includes relative performance measures. Finally, Casas-Arce and Martinez-jerez (2009) examine the performance impact of an introduction of contest based incentive schemes using data from a contest among the retailers of a commodities manufacturer. This study shows that the implementation of a contest among retailers within the firm is associated with sales performance growth. Taken together, common measures can provide less noisy information about the effort level that ratees exert because common measures can reduce common uncertainty through RPE. Thus, I conjecture that ratees have incentives to exert more effort on RPE measures than on non-RPE measures due to the reduction of common uncertainty among ratees.

Second, from a behavioral point of view[5], even though ratees do not obtain additional compensation, they have incentives to work harder when Relative Performance Information (thereafter "RPI") is provided than when RPI is not. Social Comparison Theory predicts that ratees have incentives to evaluate their own abilities by comparing themselves with other people continuously (Festinger 1954). Because ratees feel that their sense of self-identity deteriorates when their performances are worse than those of

[5] Lipe and Salterio(2000) argue that only common measures affect the rater's evaluation when evaluating performance among ratees, which is called as "common measure bias". This bias is based on the rater's incentive because common measures are easier to use in comparing ratees (Solvic and Macphilamy 1974). In my research setting, because the weights of performance measures are determined by the designer *in the performance evaluation system design process*, I would like to focus on the ratee's incentive regarding common measures in performance evaluation.

other people, they have incentives to perform better than other people (Tesser and Campbell 1980) or to retain a positive self-image (Beach and Tesser 1995; Tesser 1988). Having higher performance compared to other people provides ratees with positive feelings like pride. On the contrary, having lower performance than other people provides ratees with negative feelings such as shame (Lazarus 1991; Smith 2000). Therefore, I expect that ratees will exert more effort on common measures which enable social comparison than unique measures. As a result, I expect the performance level of common measures to improve.

Moreover, when ratees receive feedback through RPI, they have learning opportunities to evaluate whether their hypothesis about how to improve performance is wrong. Thus, they have motivation to exert more effort on the common measures due to the learning effect of feedback (Kesseler and Ashton 1981; Kluger and DeNisi 1996). Furthermore, providing RPI could improve performance not only under individual performance based contracts (Hannan et al 2008) but also under flat-wage contracts (Kerr et al 2007). Social comparison by disclosing the RPI to the public could enhance the motivation effect on the ratee's effort (Wilson and Benner 1971; Smith and Insko 1987). As a recent laboratory study, Tafkov (2013) finds that the feedback of RPI which is disclosed into public improves performance more under individual incentive contracts than under flat wage contracts.

In my research setting, the RPI captured by common measures is annually disclosed

to the public through performance evaluation reports on SOEs. Moreover, ratees are compensated based on the aggregation of individual performance scores under the tournament incentive scheme[6]. Therefore, I expect that ratees have incentives to respond by exerting more effort on common measure rather than on unique measures when a performance measure weight increases. Taken together, I hypothesize:

Hypothesis 2. When a performance measure weight is changed, ratees would change their own effort level more sensitively to common measures than to unique measures.

6 Even though common measures in my study have individual target and achievement score, raters consider relative performance scores at the aggregate level when they evaluate a ratee's performance score for compensation. Thus, I conjecture that ratees are seems to be evaluated based on relative performance evaluation under a tournament compensation scheme.

3.4. Task Specific Ability

First, from the viewpoint of economic theory, Bonner (1999) contents that it is important to understand exactly what skill the individual brings to improve task performance. Specifically, Awasthi and Pratt (1990) shows that subjects working under incentive contracts exert higher effort duration than subjects under fixed pay, irrespective of ability. However, subjects under incentive contracts did not perform better than those working under fixed pay without a high degree of skill. That is, while monetary incentive contracts motivate ratees to exert more effort, these incentives would increase performance of only ratees with high task-relevant skill. Similarly, Nicholls (1984) argues that ratees that have high task-specific ability could achieve more with equal effort or use less effort than others for to achieve equal performance. Meanwhile, Even though ratees try harder in the presence of incentives, in other words, exhibit higher effort intensity or higher effort duration, if ratees lack the ability required for a specific task, their performance would be invariant to increases in effort (Arkes 1991; Bonner et al 2000; Smith and Walker 1993). In addition, Bonner et al. (2001) also argue that task related skill could influence the incentive-effort as well as effort–performance relation because ratees allocate efforts among tasks based on the effort-performance relation. In other words, ratees, who are responsible for the same task but have different task-specific abilities, have incentives to exert different levels of effort to each task under a performance evaluation system. Thus, the performance measures in which ratees

have high task-specific ability could provide ratees with an effective effort-benefit relationship while performing the task. Therefore, I expect that ratees would respond by exerting more effort when the weight of performance measures in which they have high task-specific ability increases.

In addition, from the viewpoint of behavioral theory, Locke and Latham (1990) argue that when ratees feel that the goal is difficult to achieve due to low task-specific ability, they tend to abandon the goal achievement and in turn the performance becomes worse[7]. Moreover, Kanfer and Ackerman (1989) contend that when the task is not permitted to be abandoned but the performance of that task is bad, the performance of that task becomes worse over time because ratees do not focus on task-related strategies due to exerting cognitive effort on non-task related activities such as being anxious about performance.

From a different viewpoint, under behavioral theory, when ratees have discretion about how to allocate their effort on various tasks, RPI could distort effort allocation of ratees among tasks because ratees may attempt to solve the threat of self-image in one domain by confirming their competitiveness of another performance measure (Steele 1988). Specifically, people prefer to differentiate themselves from other people (Snyder and Fromkin 1980; Lynn and Snyder 2002), and this type of a positive difference could enhance the self-image and social status of ratees (Tesser 1988; Lynn and Snyder 2002;

[7] Leone and rock (2002) argue that when ratees do not achieve the target this year, the target will not be revised downward in the following year due to ratcheting. Thus, I conjecture that ratees would be frustrated because target in the following year is still difficult to achieve when task-specific ability is low.

Simsek and Yalincetin 2010). Brickman and Bulman (1977) shows that ratees prefer the feedback that they perform one task well while performing another task badly to the feedback that they perform both tasks on average. Therefore, the needs for social distinction and self-affirmation can motivate ratees to perform at least one task better at the expense of bad performance for other tasks. Taken together, I expect that when a performance measure weight is increased, relative performance information from common measures could motivate ratees to exert more effort on performance measures in which they have high task-specific ability rather than on those of low task specific ability. Thus, I hypothesize:

Hypothesis 3. When a performance measure weight is changed, ratees would change their effort level more sensitively on the performance measures in which they have high task-specific ability rather than those of low task-specific ability.

IV. Empirical Design

4.1. Testing Hypothesis 1

To test H1, I develop a regression model to investigate whether ratees change effort level in response to the change of performance measure weight and estimate equation (1) by using OLS regression. I incorporate performance measure change (*d_attain*) as the dependent variable to capture the ratee's effort level in the regression equations. *d_attain* is measured as the difference between the percentage of performance score in year t and in year t-1[8].

As the independent variable of interest, the change of performance measure weight (*d_weight*) is calculated as the difference between the weight of an individual performance measure in year t and in year t-1. I expect the sign of coefficient β_1 to be positive, because ratees will exert more effort on the performance measures when the weights of those performance measures increase.

$$d\_attain_{j,i,t} = \alpha_0 + \beta_1 * d\_weight_{j,i,t} + \beta_2 * weight_{j,i,t} + \beta_3 * room_{j,i,t} + \beta_4 * m\_tenure_{j,i,t} + \beta_5 * common_{j,i,t} + \beta_6 * ln\_asset_{i,t} + \beta_7 * roa_{i,t} + \beta_8 * lev_{i,t} + year\ dummies + \varepsilon_t \quad (1)$$

where *i*, *j* and *t* indicate the firm, the individual performance measure, and the time

[8] Following Ahn et al. (2010), I use *d_attain* variable as a proxy of ratee's effort level by assuming that the change of ratee's effort level could be realized into the change of performance score perfectly.

period respectively.

I include various control variables that could affect the ratee's effort (*d_weight*). Specifically, I control for various measurement characteristics; weight, room, tenure, and commonality. *Weight* would influence ratee's incentive to allocate efforts because it captures the relative importance of individual performance measure across SOEs each year (Ahn et al 2010; Ahn, Lee and Park 2011). I include *room*, measured as the log value of 101 percent subtracted by the Performance Score at year t-1, to account for the degree of possible further improvement, because ratees could care more about improving low ratings than improving good ratings (Ahn et al 2010). Because a longer tenure of a performance measure could provide ratees with learning opportunities, I incorporate performance measure tenure (*m_tenure)* which is calculated as the number of years after the performance measures are first included (Meyer and Gupta 1994; Kelly 2010). In addition, because ratees have more informational advantages on unique measures than on common measures, they can achieve a higher performance score on unique measures. I create an indicator variable (*common)*, which would take on the value of one if performance measures are common and zero if performance measures are unique (Ahn, Choi and Hwang 2013) to control the commonality of the performance measures.

To control the firm's specific business condition which could influence the ratee's

effort level, I include *size*, *roa,* and *lev*. First, I measure *size* as the natural logarithm of the firm's total asset. Second, I measure *roa* as operating income divided by the book value of total assets to capture firm's profitability. Third, I measure *lev* as the book value of liabilities divided by the book value of assets. I include dummy variables that represent years to capture differences across sample years.

4.2. Testing Hypothesis 2

Regarding H2, to examine whether ratees respond more sensitively to the weight change of common measures than to that of unique measures, I estimate equation (1) separately for the subsample of common measures and unique measures.

I expect the sign of coefficient β_1 in both subsamples of common measures and unique measures using equation (1) to be positive, because ratees will exert more effort on the performance measures when the weights of those performance measures increase regardless of commonality. However, according to the prediction of H2, I expect that β_1 of equation (1) in the common measures subsample will be greater than that β_1 of equation (1) in the unique measures subsample because ratees would exert more effort on common measures than on unique measures when performance measure weights increase. To capture this difference between common measures and unique measures, I incorporate the interaction terms of both independent variables of interest and all control

variables based on equation (1)[9].

$$d\_attain_{j,i,t} = \alpha_0 + \beta_1 * common_{j,i,t} + \beta_2 * d\_weight_{j,i,t} + \beta_3 * d\_weight_{j,i,t} * common_{j,i,t} + \beta_4 * weight_{j,i,t} + \beta_5 * weight_{j,i,t} * common_{j,i,t} + \beta_6 * room_{j,i,t} + \beta_7 * room_{j,i,t} * common_{j,i,t} + \beta_8 * m\_tenure_{j,i,t} + \beta_9 * m\_tenure_{j,i,t} * common_{j,i,t} + \beta_{10} * ln\_asset_{i,t} + \beta_{11} * ln\_asset_{i,t} * common_{j,i,t} + \beta_{12} * roa_{i,t} + \beta_{13} * roa_{i,t} * common_{j,i,t} + \beta_{14} * lev_{i,t} + \beta_{15} * lev_{i,t} * common_{j,i,t} + year\ dummies + the\ interaction\ term \quad year\ dummies * common_{j,i,t} + \varepsilon_t \quad (2)$$

where *i*, *j* and *t* indicate the firm, the individual performance measure, and the time period respectively.

I expect the sign of the coefficient β_3 of equation (2) to be positive, because ratees will exert more effort on the common measures rather than on the unique measures when the weights of those performance measures increase.

4.3. Testing Hypothesis 3

Regarding H3, I estimate the following equations (3) and (4), which is based on equation (1), to investigate whether ratees allocate more effort on the performance

[9] This regression equation is developed to incorporate the interaction term of all control variables as well as the independent variable of interest with indicator variable "common" because I conjecture that both firm level control variables and measure level control variables could affect the dependent variable *d_attain* differently based on the commonality of performance measures.

measure in which they have high task-specific ability than those of low task-specific ability in accordance with the weight changes of performance measures.

In equation (3), I include, as another independent variable of interest, the task-specific ability measure (*ability*) to capture the capability of ratees with regard to the tasks that the performance measure represents. I also incorporate the interaction term *d_weight*ability* to capture the effect of ability on ratee's effort allocation in accordance with the change of performance measure weight.

There are many possible ways to measure task-specific ability. Agency theory predicts that past performance can provide useful information on measuring an agent's ability (Banker et al 2013; Banker and Hwang 2008). Prior studies use past performance such as prior industry-adjusted return on assets or stock returns as a proxy to capture manager ability (Fee and Hadlock 2004; Rajgopal et al 2006; Milbourn 2003). Moreover, Nicholls (1984) contends that ability could be judged to be high or low with reference to the individual's own past performance or knowledge.

As such, I create the variable *ability* using past performance of the performance measure to capture task-specific ability. Specifically, I sort my sample into quartiles based on the score of performance measures obtained by the ratee in the previous year and then assign the highest (lowest) performance score quartile to high (low) task-specific ability, having the variable *ability* take on the value of 3(0) for these observations. In other words, as a rank variable, *ability* in equation (3) takes on the value

of 3 in high task-specific ability and zero in low task-specific ability. I expect the sign of the interaction term coefficient β_3 in equation (3) to be positive, because the sensitivity of the ratee's effort on the change of measure weight would be greater in high task-specific ability measures than in low task-specific measures.

I perform additional analysis by substituting *ability* as a rank variable with ability as a dummy variable to provide empirical evidence that there is a difference in the ratee's response between high task-specific ability measures and low task-specific ability measures in equation (4). I create an indicator variable (*high_ability*), which would take on the value of one if the performance score of the measure is in the highest performance score quartile and zero if not. Likewise, I create another indicator variable (*low_ability*), which would take on the value of one if the performance score of the measure is in the lowest performance score quartile and zero if not. In line with H3, I expect the sign of β_4 of equation (4) to be positive because ratees would exert more effort on the performance measures in which they have high task-specific ability when the performance measure weights increase. However, I could not predict of the sign of β_5 of equation (4) because it is difficult for ratees to exert effort on the low task-specific ability measure due to the lack of capability to improve performance. According to Ahn et al (2010), I include additional control variable *discriminability* to capture the effect of cross-sectional variation at common measure i in year t-1 in equation (3) and (4),

$$d\_attain_{j,i,t} = \alpha_0 + \beta_1 * d\_weight_{j,i,t} + \beta_2 * ability_{j,i,t} + \beta_3 * d\_weight_{j,i,t} * ability_{j,i,t} + \beta_4 * discriminability_{j,t} + \beta_5 * weight_{j,i,t} + \beta_6 * m\_tenure_{j,i,t} + \beta_7 * ln\_asset_{i,t} + \beta_8 * roa_{i,t} + \beta_9 * lev_{i,t} + year\ dummies + \varepsilon_t \quad (3)$$

$$d\_attain_{j,i,t} = \alpha_0 + \beta_1 * d\_weight_{j,i,t} + \beta_2 * high\_ability_{j,i,t} + \beta_3 * low\_ability_{j,i,t} + \beta_4 * d\_weight_{j,i,t} * high\_ability_{j,i,t} + \beta_5 * d\_weight_{j,i,t} * low\_ability_{j,i,t} + \beta_6 * discriminability_{j,t} + \beta_7 * weight_{j,i,t} + \beta_8 * m\_tenure_{j,i,t} + \beta_9 * ln\_asset_{i,t} + \beta_{10} * roa_{i,t} + \beta_{11} * lev_{i,t} + year\ dummies + \varepsilon_t \quad (4)$$

where *i*, *j* and *t* indicate the firm, the individual performance measure, and the time period respectively.

All regressions in this section are estimated with Huber-White robust standard errors clustered by firm. These standard errors would be robust to both serial correlation and heteroskedasticity (Rogers 1993).

V. Empirical Results

5.1. Descriptive Statistics

Table 5 presents descriptive statistics of the variables used in my regression models. The total number of observations is 11,321 measure-firm-years. The average (median) of *d_attain* is -0.006 (0) and the average (median) of performance measure weight change is -0.035 (0). The average (median) of *discriminability*, which is measured as the cross-sectional deviation among measures at a year is 0.095 (0.078). The average (median) of *ability*, which takes on the value of 3 when high task-specific ability and 0 when low task-specific ability, is 1.555 (2). The average (median) performance measure tenure is 4.58 (3) years, suggesting that performance measure are retained for about 5 years after being included in a performance measure set. Moreover, the average (median) *weight*, which is a weight that ranges from 0.5 points to 27 points according to relative importance, is 3.17 (3). The average (median) room, which shows the degree of possible further improvement, is 2.223(2.603).

The average (median) *size* of firms as the natural log of total assets is 21.480(21.428). The mean value of *roa* and *lev* are 0.013(0.008) and 0.621(0.590), respectively, which implies that the firms are financially stable but not highly profitable.

[INSERT TABLE 5 ABOUT HERE]

Table 6 presents the Pearson correlation matrix. Other than the high correlation between room and ability of -0.728 which is because both of them are calculated on the basis of past performance, most variables are not highly correlated with each other. Specifically, *d_weight* is positively correlated with *d_attain*, which seems to support H1. *Discriminability* is positively correlated with *d_attain*, which is consistent with the results of Ahn et al (2010). Because I do not include control variables in this analysis, the results are only preliminary and an analysis using a multiple regression method should be applied.

[INSERT TABLE 6 ABOUT HERE]

5.2. Performance Measure Weight Change and Ratee's Effort Allocation

Table 7 presents the estimation results of equation (1) to test H1 using the full sample including both common and unique measures. The coefficient of *d_weight* is 0.002 in equation (1) and is statistically significant at the 5% confidence level, which supports the prediction of H1 that states that ratees reallocate their own effort in accordance with the performance measure weight change. The adjusted R-squared of equation (1) is 19.8 percent. Among the control variables that capture the measure's characteristics, the sign

of *weight* and *room* is positive and statistically significant, suggesting that ratees exert more effort on the performance measures which have high relative importance and have potential for future improvement. These results are consistent with Ahn et al (2010). Among firm level control variables, the coefficient of *roa*, which captures the profitability of a firm, is positive and statistically significant. Consequently, I interpret my findings as being consistent with H1: when a performance measure weight is changed, ratees respond by changing their effort level on the performance measure in accordance with the change.

[INSERT TABLE 7 ABOUT HERE]

Table 8 provides the estimation results of equation (1) to test H2 by dividing the full sample into two sub-samples: common measures and unique measures. Using common measures, Column (1) of Table 8 shows that the coefficient of *d_weight* is 0.004 and is statistically significant at 1% confidence level. Meanwhile, using unique measures, Column (2) of Table 8 presents that the coefficient of *d_weight* is -0.003 but is statistically insignificant. From the results of equation (2), column (3) of Table 8 shows that the coefficient of the interaction term, $d\_weight_{j,i,t} * common_{j,i,t}$ in equation(2) (the difference between the two coefficients of the regressions using the sub-samples) is 0.007 and is statistically significant at the 5% confidence level.

These results imply that ratees reallocate their efforts in accordance with the changes when common performance measure weights are changed, but they do not reallocate their own efforts when unique performance measure weights are changed. More importantly, my findings imply that ratees would exert more effort on common measures on the unique measures when performance measure weights are changed. Therefore, I interpret these findings as evidence that supports H2: when performance measure weights are changed, ratees would change their efforts level more sensitively to common measures than to unique measures. The adjusted R-squares of equation (1) in Column (1) and equation (1) in Column (2) of Table 8 are 23% and 18.68%, respectively. The coefficients of control variables are similar with those of table 7.

[INSERT TABLE 8 ABOUT HERE]

Table 9 provides the estimation results of equation (2) and (3) to test H3. Panel A of Table 9 presents that the coefficient of interaction term, *d_weight*ability* which captures the interactive effects between change of performance measure weight and task-specific ability of ratees is 0.004 and is statistically significant at 1% confidence level, which is consistent with H3. The coefficients of the control variables are similar with those of equation (1) and the adjusted R-square of equation (2) is 23.8%. I find evidence that

supports the prediction that ratees allocate more effort on the performance measures in which they have high task-specific ability than those of low task-specific ability in accordance with the weight change of performance measure.

Panel B of Table 9 shows the results of equation (3) regarding H3. I incorporate two dummy variables *high_ability* and *low_ability* which are measured as whether the score is in the highest performance score quartile of each ratee in previous year and as whether the score is in the lowest performance score quartile of each ratee in equation (3). As the independent variable of interest, the coefficient of the interaction term *d_weight*high_ability* which captures the effects of ratees's effort allocation when performance measure weights are changed is 0.009 and is statistically significant at 1% confidence level. On the contrary, the coefficient of interaction term *d_weight** *low_ability* is negative but is statistically insignificant at conventional levels.

I interpret these results as evidence that supports the conjecture that ratees allocate more effort on the performance measure in which they have high task-specific ability than those of low task-specific ability in accordance with the weight change of performance measure. The adjusted R-square of equation (3) is 22.6% and the coefficients of control variables are similar with those of the previous equations above.

Consequently, I find that the results of both Panel A and Panel B of Table 9 support H3 : when a performance measure weight is changed, ratees change their effort level more sensitively on the performance measures in which they have high task-specific

ability rather than those in which they have low task-specific ability.

[INSERT TABLE 9 ABOUT HERE]

VI. Additional Analysis

6.1. Additional Analysis Regarding H1

Table 10 provides additional evidence regarding H1. Ahn et al (2009) uses performance measure weight level as a proxy to capture the relative importance of performance measures and performance score level as a proxy to estimate how much effort ratees exert on the measure. Using the variables of Ahn et al (2009), I perform an additional analysis by including the measure *weight* to capture the relative weight of a performance measure as the independent variable and *attain* to capture the ratee's effort level on a performance measure as the dependent variable. I expect that the performance measure weight is positively correlated with performance score that ratees achieve. As shown in Table 10, the coefficient of weight is 0.003 and is statistically significant at the 5% confidence level, which is still consistent with H1.

[INSERT TABLE 10 ABOUT HERE]

6.2. Additional Analysis Regarding with H2

I also perform a sub-sample analysis using common and unique measures to test H2. The coefficient of *weight* in Column (1) of Table 11 using common measures is 0.004 and is statistically significant at the 1% confidence level. Meanwhile, the coefficient of *weight* in Column (2) of Table 11 using unique measures is -0.001 and is statistically

insignificant at conventional levels. More importantly, the difference between two coefficients using common and unique measures is 0.005 and is statistically significant at the 1% confidence level. Thus, the results of these additional tests support H2.

[INSERT TABLE 11 ABOUT HERE]

6.3. Additional Analysis Regarding H1 and H2

I perform an additional analysis by dividing my common measure sample into two sub groups: *weight increase* and *weight decrease* samples. This is to investigate whether ratees exert more (less) effort on the measure when performance measure weights increase (decrease). I create the indicator variable *inc_weight* (*dec_weight*) which takes on the value of one when a performance measure weight *increases* (*decreases*) and zero if not.

These results of Table 12 provide weak evidence that supports H1 and H2. As shown in Column (1) of Table 12, the coefficient of *inc_weight* is 0.010 and is statistically significant at a marginal level. However, Column (2) of Table 12 shows that the coefficient of *dec_weight* is -0.008 and is statistically insignificant at conventional levels. I interpret these results as weak evidence that supports H1 and H2 because the sign of coefficients is consistent with my prediction even though they are statistically

insignificant in the weight decrease sample[10].

[INSERT TABLE 12 ABOUT HERE]

6.4. Additional Analysis Regarding H3

To check the robustness with regard to H3, I conduct an additional analysis using different task-specific ability variables. I define *ability2* by sorting the performance score into quartiles based on performance score *of all ratees* in the previous year, *not of each ratee (ability)* and define *ability3* by sorting performance score into quartiles based on average performance score *of each ratee* in the previous 2 years (Woods 2012). Thus, I define *ability2* (*ability3*) as rank variables based on performance score quartiles.

Table 13 provides the results of equation (2) using various task-specific ability variables. As shown in Column (1) of Table 13, the coefficient of interaction term *d_weight*ability2* is 0.005 and is statistically significant at the 1% confidence level. In Panel B, the coefficient of *d_weight*ability3* is 0.004 and is statistically significant at a marginal level. Therefore, the results of these additional tests support H3.

[10] Meyer and Gupta (1994) and Kelly (2010) argue that the score of performance measures tend to increase over time due to learning. Moreover, Brickely and Zimmerman (2001) shows that research output declined steadily after the implementation of this incentive scheme which focuses on teaching quality Thus, I conjecture that the learning effect of ratees could cancel out the effect on the ratee's decrease in effort when performance measure weights decrease.

[INSERT TABLE 13 ABOUT HERE]

IV. Conclusion

When agents conduct various tasks on behalf of principal, they allocate their own effort among various tasks to maximize their own benefit under incentive contracts. This effort allocation of agents is influenced by the characteristics of individual performance measures (Holmstrom and Milgrom 1991).

The purpose of my study is to empirically examine whether ratees change their effort level on performance measures when the designer of the performance evaluation system adjusts the performance measure weights under a multi-task agency environment and what types of roles performance measure characteristics such as commonality and task-specific ability have when ratees adjust their effort level in accordance to the changes in performance measure weights.

The results are as follows. First, when the designer of performance evaluation system changes a performance measure weight, ratees reallocate their own effort in accordance to the change in the performance measure. In other words, ratees exert more effort on the performance measure when a performance measure weight increases. These results are consistent with multi-task agency theory which states that ratees would input more effort on the performance measures with increased weights because the ratee can obtain more benefits from exerting the same level of effort (Brickley, Smith and Zimmerman 2009). Second, I find that ratees reallocate their efforts when a common performance measures weight changes, but they do not reallocate their efforts when the weight of

unique performance measure changes. I also find that ratees exert more effort on common measures than on unique measures when performance measure weights increase. These results are consistent with prior findings that ratees would exert more effort on performance measures which allow relative performance evaluation (Frederickson 1992; Matsumura and Shin 2006) and that relative performance information (RPI) positively affects performance when RPI is provided to ratees and publically disclosed (Tafkov 2013). Third, I also present that ratees change their effort level more sensitively on the measures in which they have high task-specific ability than on those of low task-specific ability when performance measure weights change.

My study sheds light on the performance evaluation academic literature by providing evidence on the ratee's response in accordance with performance measure weight change. Multi-task agency theory predicts that ratees reallocate their efforts in accordance with the change of performance measure weights (Brickley, Smith and Zimmerman 2009). However, the empirical research to investigate this theory is very scarce. My study shows a clearer relationship between the performance measure weight change and the ratee's effort level by accurately capturing the performance measure weight change using time series data for 26 years. Moreover, I extend prior research by examining the effects of commonality and task-specific ability on the ratee's effort reallocation according to the change of performance measure weight. Specifically, based on theoretical and experimental studies (Tafkov 2013), I provide empirical evidence that

ratees exert more effort on the common performance measure because RPI is available as feedback and is publically disclosed. I also show empirically that task-specific ability influences the incentive-effort relation. Although the effect of ability on the incentive-effort relation has been discussed extensively, there are few empirical findings to provide the direct evidence regarding this issue (Bonner and Sprinkle 2002).

My study also contributes to the performance evaluation practice as three ways. First, I provide empirical evidence that the increase of performance measure weight is a useful tool for designers to motivate ratees to exert more efforts on the task which designer intends to motivate. Second, I also deliver not only insight that the disclosure of RPI could motivate ratees to exert more efforts to increase performance, but also enlighten the fact that the increase of performance measure weights is not enough for ratees to exert more effort in unique performance measures. Thus, the designer needs to consider the advantage and disadvantages of RPI as a mechanism to motivate the ratee. Third, the adjusting performance measure weights is more effective in motivating ratees to exert effort on the performance measure in which they have high task-specific ability than on those of low task-specific ability. Therefore the designer needs to consider how to improve the ratee's task-specific ability before increasing the weights of performance measures in which ratees have low task-specific ability.

Notwithstanding, my study is subject to many limitations. First, I am very cautious in generalizing my findings because results are based on only 15 Korean public SOEs.

Second, I define task-specific ability variables based on the past performance scores which the ratee achieves in the previous year or the past 2 years. However, these measures could include the measurement errors. In spite of the limitations discussed above, adjusting the performance measure weight would be a general mechanism to motivate ratees in the performance evaluation system. My study still provides meaningful insights on the design of the performance evaluation system. In future research, researchers should consider environmental factors which could influence the ratee's effort allocation when the performance measure weights changes under a multi-task agency environment.

REFERENCES

Ahn, T. S, S. Y. Choi, and I. Hwang. 2010. Empirical Evidence with Respect to Target Ambiguity Interactive Target Setting, and Measurement Noise. *Working Paper*, Seoul National University.

Ahn, T. S, S. Y. Choi, and I. Hwang. 2013. Impact on Ratchet Effect of Superior Subjectivity and Subordinate Influence in Target Setting. *Working Paper*, Seoul National University.

Ahn, T. S, I. Hwang, and M. I. Kim. 2010. The impact of performance measure discriminability on Ratee incentives. *The Accounting Review* 85(2): 289-417.

Ahn, T. S, J. Y. Lee, and J. H. Park. 2011. The effects of firm size and measurement characteristics on performance score. *Korea Journal of Management Accounting Research* 11(2): 107-137. [Printed in Korean]

Ahn, T. S, Y. S. Choi, and D. H. Kwon. 2011. An Examination of Target-Setting: How a Rater revises a Ratee's Target. *The Korean Accounting Review* 36(1):183-223.[Printed in Korean]

Ahn, T. S, Y. S. Choi, and M. J. Seo. 2009. CEO effort allocation under a multi-task setting. *The Korean Accounting Journal* 18(2): 313-342 [Printed in Korean]

Arkes, H.R. 1991. Costs and Benefits of Judgment Errors: Implications for Debiasing. *Psychological Bulletin* 110(3): 486-498.

Awasthi, V., and J. Pratt. 1990. The effects of financial incentives on effort and decision performance: the role of cognitive characteristics. *The Accounting Review* 65(4):797-811.

Banker, R. D. and I. Hwang. 2008. Importance of Measures of Past Performance: Empirical Evidence on Quality of e-Service Provides. *Contemporary Accounting Research* 25(2):307-337.

Banker, R. D. and S. M. Datar. 1989. Sensitivity, Precision, and Linear Aggregation of

Signals for Performance Evaluation. *Journal of Accounting Research* 27(1):21-39.

Banker, R. D., M. N. Darrough, R. Huang and J. M. Plehn-Dujowich. 2013. The Relationship between CEO compensation and Past Performance. *The Accounting Review 88*(1): 1-30

Banker, R. D., S. Y. Lee, and G. Potter. 1996. A field study of the impact of a performance-based incentive plan. *Journal of Accounting and Economics* 21(2).:195-226.

Banker, R. D., S. Y. Lee, G. Potter and D. Srinivasan. 2001. An empirical analysis of continuing improvements following the implementation of a performance-based compensation plan. *Journal of Accounting and Economics* 30: 315-350.

Beach. S. R. H. and A. Tesser. 1995. Self-Esteem and the Extended Self-Evaluation Maintenance Model. Efficacy, Agency, and Self-Esteem, *The Springer Series in Social Clinical Psychology*: 145-170.

Bonner, S. E. 1999. Judgment and decision-making research in accounting. *Accounting Horizons* 13: 385-398.

Bonner, S.E., R. Hastie, G. B. Sprinkle and S. M. Young. 2000. A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting. *Journal of Management Accounting Research* 12: 19-64.

Bonner, S. E., R. Hasite, R., S. M. Young, J. Hesford, and D. Gigone. 2001. Effects of monetary incentives on the performance of a cognitive task: the moderating role of skill. *Working Paper*. University of Southern California.

Bonner, S. E. and G. B. Sprinkle. 2002. The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27:303-345.

Brickley, J. A. and J. L. Zimmerman. 2001. Changing incentives in a multitask environment: evidence from a top-tier business school. *Journal of Corporate Finance* 7: 367-396.

Brickley. J. A., C. W. Smith and J. L. Zimmerman. 2009. *Managerial Economics and Organizational Architecture* Fifth Edition. McGraw-Hill Irwin, New York, NY.

Brickman, P. and R. J. Bulman. 1977. Pleasure and Pain in social comparison. In J. M. Suls and R. L. Miller (eds.) *Social Comparison Processes*. Washington, D.C.: Hemisphere.

Casas-Arce, P. and F. A. Martinez-Jerez. 2009. Relative Performance Compensation, Contests, and Dynamic Incentives. *Management Science* 55(8): 1306-1320

Datar, S., S. C. Kulp, and R. A. Lambert. 2001. Balancing Performance Measures. *Journal of Accounting Research* 39(1): 75-92.

Dewatripont, M, I. Jewitt, and J. Tirole. 2000. Multitasking agency problems: Focus and Task clustering. *European Economic Review* 44: 869-877.

Fee, C. E. and C. J. Hadlock. 2004. Management turnover across the corporate hierarchy. *Journal of Accounting and Economics* 37(1):3-38.

Feltham, G. and J. Xie. 1994. Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations. *The Accounting Review* 69(3):429-453.

Festinger. L. A. 1954. Theory of Social Comparison Processes. *Human Relations* 7:117-140

Frederickson, J. R. 1992. Relative Performance Information: The Effects of Common Uncertainty and Contract Type on Agent Effort. *The Accounting Review* 67(4):647-669.

Hannan, R. L., R. Krishnan and A. H. Newman. 2008. The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans. *The Accounting Review* 83(4): 893-913

Holmstrom, B. 1979. Moral Hazard and Observability. *The Bell Journal of Economics*. 10(1): 74-91

Holmstrom, B. and P. Milgrom. 1991. Multitask Principal-Agent Analysis: Incentive

Contracts, Asset Ownership and Job Design. *Journal of Law, Economics and Organization* 7:24-52

Kanfer. R. and P. L. Ackeman. 1989. Motivation and cognitive abilities: an integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of applied psychology*. 74(4): 657-690

Kaplan, R. and D. Norton. 1996. *The Balanced Scorecard: Translating Strategy into Action*. Harvard Business School Press, Boston, MA.

Kelly, K. 2010. Accuracy of Relative Weights on Multiple Leading Performance Measures: Effects on Managerial Performance and Knowledge. *Contemporary Accounting Research* 27(2): 577-608.

Kerr, N. L, L. A. Messe, D. H. Seok, E. J. Sambolec, R. B. Lount Jr., and E. S .Park. 2007. Psychological Mechanisms Underlying the Kohler Motivation Gain. *Personality and Social Psychology Bulletin* 33: 828-841

Kessler, L. and R. H. Ashton. 1981. Feedback and Prediction Achievement in Financial Analysis. *Journal of Accounting Research* 19(1): 146-161

Kluger, A. N. and A. DeNisi. 1996. The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin* 119(2): 254-284

Lazarus, R. 1991. *Emotion and Adaptation*. New York, NY: Oxford University Press

Leone, A. J., and S. Rock. 2002. Empirical tests of budget ratcheting and its effect on manager's discretionary accrual choices. *Journal of Accounting and Economics* 33: 43-67.

Lipe, M. G. and S. E. Salterio. 2000. The Balanced Scorecard: Judgmental Effects of Common and Unique Performance Measures. *The Accounting Review* 75(3): 283-298.

Locke, E. A., and G. P. Latham. 1990. *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice-Hall.

Lynn, M., and C. Snyder. 2002. Uniqueness seeking. *In Handbook of Positive Psychology,* edited by C. Snyder, and S. Lopez, 395–410. New York, NY: Oxford University Press.

Matsumura, L. M. and J. Y. Shin. 2006. An Empirical Analysis of an Incentive Plan with Relative Performance Measures: Evidence from a Postal Service. *The Accounting Review* 81(3):533-566

Meyer, M. W. and V. Gupta. 1994. The Performance Paradox. *Research in Organizational Behavior* 16:309-369.

Milbourn, T. T. 2003. CEO reputation and stock-based compensation. *Journal of Financial Economics* 68:232-262.

Nicholls, J. G. 1984. Achievement Motivation: Conceptions of Ability, Subjective Experience, Task Choice, and Performance. *Psychological Review* 91(3): 328-346

Park, S. H. 2003. The Analysis of Performance Evaluation System of SOEs over 20 years, *the Journal of Public Administration* 15(1): 11-47. [Printed in Korean]

Rajgopal, S., T. Shevlin, and V. Zamora. 2006. CEO's outside employment opportunitoes and the lack of relative performance evaluation in compensation contracts. *The Journal of Finance* 61(4): 1818-1843

Simsek, O. F. and B. Yalincetin. 2010. I feel unique, therefore I am: The development and preliminary validation of the personal sense of uniqueness (PSU) scale. *Personality and Individual Differences* 49: 576-581.

Smith, R. 2000. Assimilative and contrastive emotional reactions to upward and downward social comparisons. *In Handbook of Social Comparison: Theory and Research,* edited by J. Suls, and L.Wheeler, 173–200. New York, NY: Kluwer Academic/Plenum Publishers.

Smith, R., and C. Insko. 1987. Social comparison choices during ability evaluation: The effects of comparison publicity, performance feedback, and self-esteem. *Personality and Social Psychology Bulletin* 13: 111–122.

Smith, V. L., and J. M. Walker. 1993. Monetary Rewards and Decision Cost in Experimental Economics. *Economic Inquiry* 31: 254-261

Snyder, C., and H. Fromkin. 1980. Uniqueness: *The Human Pursuit of Difference*. New York, NY: Plenum.

Solvic, P. and D. Macphillamy. 1974. Dimensional commensurability and cue utilization in comparative judgment. *Organizational Behavior and Human Performance* 11: 172-194

Steele, C. 1988. The psychology of self-affirmation: Sustaining the integrity of the self. *In Advances in Experimental Social Psychology*, edited by L. Berkowitz, 261–302. San Diego, CA: Academic Press

Tafkov, I. D. 2013. Private and Public Relative Performance Information under Different Compensation Contracts. *The Accounting Review* 88(1):327-350

Tesser, A. 1988. Toward a self-evaluation maintenance model of social behavior. *In Advances in Experimental Social Psychology*, edited by L. Berkowitz, 181–228. San Diego, CA: Academic Press.

Tesser, A. and J. Campbell. 1980. Self-Definition: The Impact of the Relative Performance and Similarity of Others. *Social Psychology Quarterly* 43(3): 341-347.

Thiele, V. 2007. Performance measurement in multi-task agencies. *Research in Economics* 61: 148-163

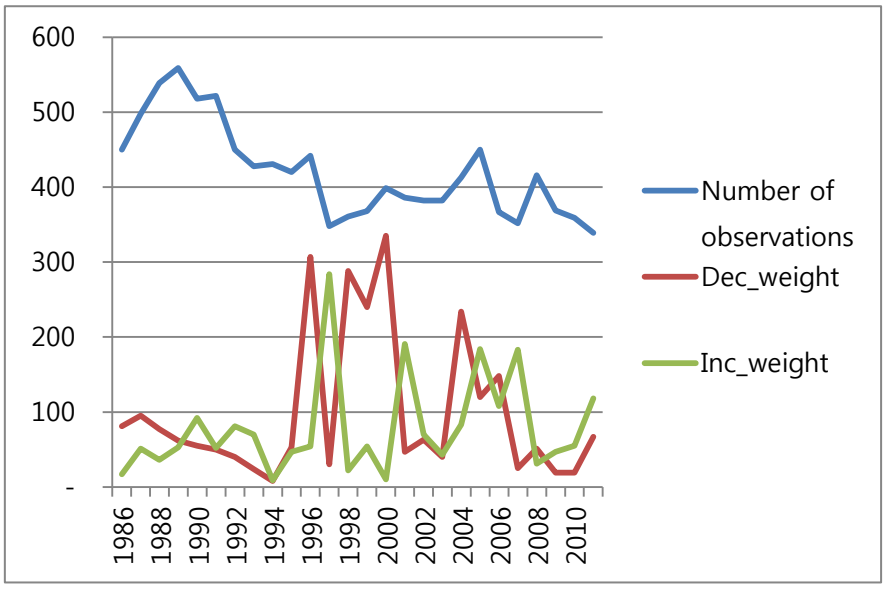
Wilson, S. R. and L. A. Benner. 1971. The Effects of Self-Esteem and Situation Upon Comparison Choices During Ability Evaluation. *Sociometry* 34(3): 381-397

Woods, A. 2012. Subjective adjustments to objective performance measures: The influence of prior performance. *Accounting, Organizations and Society* 37: 403-425

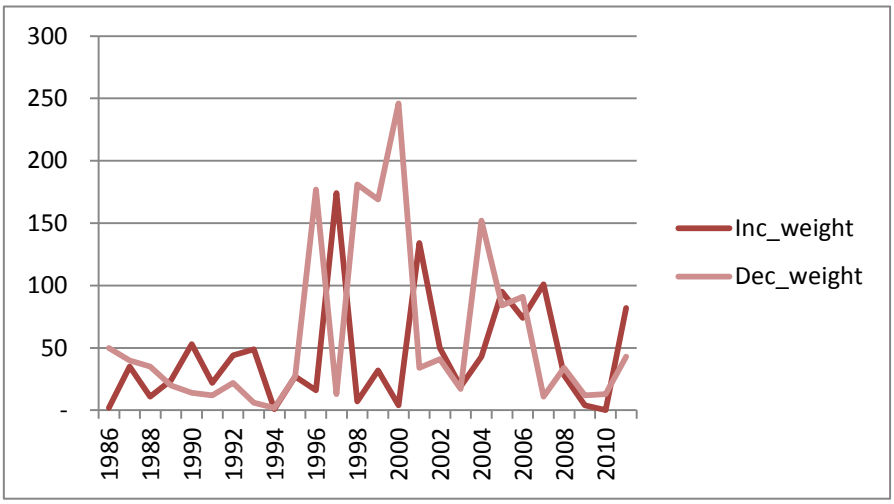
FIGURE 1

The Change of Performance Measure Weights for 1985-2011

Panel A: Full Sample



Panel B: Common Measure Sample



Panel C: Unique Measure Sample

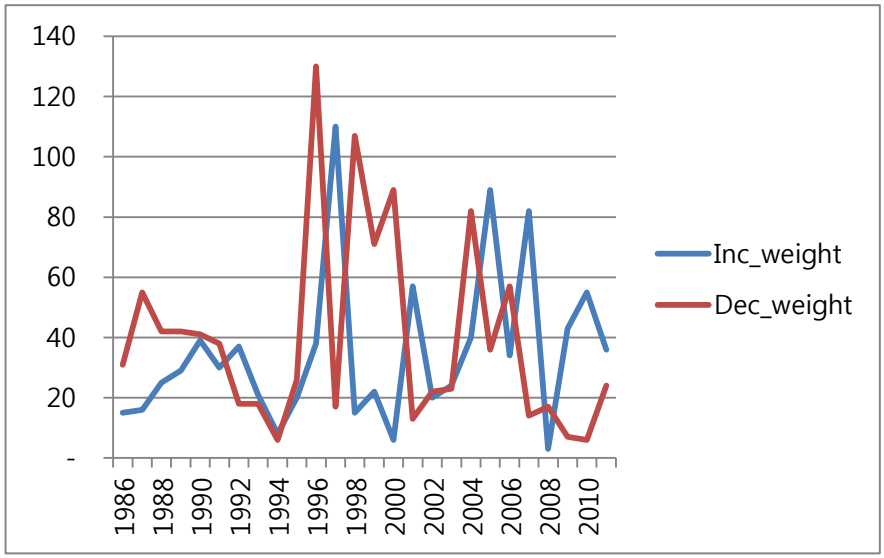


TABLE 1

Details on State Owned Enterprises in Korea

No.	State-Owned Companies	Government Ownership	Total Asset	Net Income	Captial	ROA
1	Korea Electric Power Corporation	21.17%	136,467,850	- 3,292,997	3,209,820	-2.41%
2	Korea Minting & Security Printing Corporation	100%	403,336	- 455	6,623	-0.11%
3	Korea Coal Corporation	100.00%	698,807	- 98,219	106,000	-14.06%
4	Korea Resource Corporation	99.80%	2,941,834	32,731	998,729	1.11%
5	Korea National Oil Corporation	100%	31,567,402	- 152,752	8,965,196	-0.48%
6	Korea Trade Investment Promotion Agency	100%	140,882	39,877	32,000	28.31%
7	Korea Highway Corporation	82.39%	49,279,362	103,174	23,856,515	0.21%
8	Korea National Housing Corporation	89.60%	77,744,381	116,959	10,548,684	0.15%
9	Korea Water Resources Corporation	91.20%	23,429,915	293,267	6,694,987	1.25%
10	Korea Land Corporation	73.34%	50,486,451	65,912	5,000,000	0.13%
11	Korea Rural Community & Agricultural Corporation	100%	7,113,120	92,306	1,305,029	1.30%
12	Agricultural and Fishery Marketing Corporation	100%	1,168,803	11,195	86,200	0.96%

13	Korea National Tourism Organization	55.27%	1,305,930	32,787	47,272	2.51%
14	Korea Railroad Corporation	100%	22,179,329	450,664	9,581,417	2.03%
15	Korea Land and Housing Corporation	86.60%	158,473,034	790,513	22,417,592	0.50%

1. Unit: Million Won, Percent
2. All data are prepared based on K-IFRS
3. Data are as of December 31, 2011 except Korea National Housing Corporation and Korea Land Corporation
4. The information of Korea National Housing Corporation and Korea Land Corporation were as of September 30, 2009 because two companies were merged at the end of September, 2009

TABLE 2

An Example of Performance Ratings

Category	Individual Measure	Measurement	Common/ Unique	Target Setting Methods	Weight	Rating	Score	Score rate
1. Leadership	(1) Leadership	Subjective	Common	RANK	5	B+	4	80
and	(2) Responsibility Management	Subjective	Common	RANK	3	A0	2.7	90
Responsibility	(3) Evaluation from Public	Objective	Common	ASSIGN	5	A+	4.993	99.86
	(4) Social Contribution							
	(a) Social Contribution	Subjective	Common	RANK	2	B+	1.6	80
	(b) Conform to Government Policy	Objective	Common	ASSIGN	5	A0	4.264	85.28
2. Management	(1) Operation Efficiency							
Efficiency	(a) Labor Efficiency	Objective	Common	ASSIGN	4	B0	2.629	65.735
	(b) Capital Productivity	Objective	Common	ASSIGN	4	B0	2.681	67.028
	(2) Organization and Human Resource Management	Subjective	Common	RANK	4	B+	3.2	80
	(3) Financial Budget Management and Performance		Common					
	(a) Financial Budget Management	Subjective	Common	RANK	4	B0	2.8	70
	(b) Financial Budget Performance							
	- Total Asset Turnover Rate	Objective	Common	ASSIGN	2	A+	1.851	92.544

	- Interest Coverage Rate	Objective	Common	ASSIGN	2	B0	1.201	60.046
	- Debt Ratio	Objective	Common	ASSIGN	2	D+	0.84	41.996
	(c) Qualitative Management Expense	Objective	Common	ASSIGN	2	B+	1.469	73.434
	(4) Compensation and Performance Management							
	(a) Compensation and Performance Management	Subjective	Common	RANK	4	E0	0.8	20
	(b) Increasing rate of Total Salary Expense	Objective	Common	ASSIGN	4	A+	4	100
	(5) Labor Relation Management	Subjective	Common	RANK	3	A0	2.7	90
3. Core Business	(1) Electricity and Power Demand Management	Subjective	Unique	RANK	4	B0	2.8	70
	- Maintaining High Load Factor	Objective	Unique	ASSIGN	5	A+	5	100
	- Global Benchmark for High Load Factor	Objective	Unique	BENCHMARKING	3	A+	3	100
	(2) Power Transmission and Transformation	Subjective	Unique	RANK	4	B+	3.2	80
	- Dysfunction Management for Transmission	Objective	Unique	ASSIGN	6	A+	6	100

- Efficiency Index for Transmission Facility	Objective	Unique	ASSIGN	5	A0	4.382	87.648
(3) Electric Power Distribution	Subjective	Unique	RANK	4	B+	3.2	80
- Right Voltage Management	Objective	Unique	ASSIGN	4	A0	3.979	99.467
- Global Benchmark for Loss rate of Transmission	Objective	Unique	BENCHMARKING	3	A+	3	100
- Development of Distribution Management	Objective	Unique	ASSIGN	4	A+	4	100
(4) New Business Development	Subjective	Unique	RANK	3	A0	2.7	90
Objective Measures in Total				40		29.7	
Subjective Measures in Total				60		53.289	
Total				100		82.989	

1. This table provides the performance evaluation results from 2011 for a SOE
2. The evaluation and target setting methods of objective performance measures include trend analysis (TREND), beta distribution analysis (BETA), assigning target process (ASSIGN), and negotiating target process(NEGOTIATE), the subjective performance measures are evaluated by the single evaluation method of ranking analysis
3. Score rate is a final evaluation result calculated based on evaluation rule by rater
4. Weight is the relative importance of an individual performance measure ex ante determined by Raters at the beginning of year t
5. Rating is a rater's evaluation results of individual performance measure under a certain assumption such as distribution of performance.

TABLE 3

Incentive Scheme of Korean SOEs

Beneficiary	Incentive Bonus	How to Calculate Incentive Rate
Employee	Monthly Base Salary X Incentive Rate	200% +{(Total Score[1] - Lowest Score[2])/(Highest Score[3] - Lowest Score[2])}X 300% * Incentive rate range : 200% ~ 500%
CEO/ Board of Director	Monthly Based Salary X Incentive Rate	{(Aggregated Score[4] - Lowest Score[5])/(Highest Score[6] - Lowest Score[5])}X 100% * Incentive rate range : 0% ~ 200%

1. Total Score : The total score which a SOE earns from the performance evaluation at year t.
2. Lowest Score : The lowest Score among 14 SOE's total scores
3. Highest Score : The highest Score among 14 SOE's total scores
4. Aggregated Score : {SEO's total score + (CEO's total Score X2)}/3
5. Lowest Score : The lowest Score among 14 CEO's aggregated scores
6. Highest Score : The Highest Score among 14 CEO's aggregated scores

TABLE 4

Descriptive Statistics of Change of Performance Measure Weight

year	# of obs	Full sample				Common Sample		Unique sample	
		Inc_ weight	%	dec_ weight	%	Inc_ weight	Dec_ weight	Inc_ weight	Dec_ weight
1986	450	17	4%	81	18%	2	50	15	31
1987	498	51	10%	95	19%	35	40	16	55
1988	539	36	7%	77	14%	11	35	25	42
1989	559	53	9%	62	11%	24	20	29	42
1990	518	92	18%	55	11%	53	14	39	41
1991	522	52	10%	50	10%	22	12	30	38
1992	450	81	18%	40	9%	44	22	37	18
1993	428	70	16%	24	6%	49	6	21	18
1994	431	9	2%	8	2%	1	2	8	6
1995	420	47	11%	53	13%	27	27	20	26
1996	442	54	12%	307	69%	16	177	38	130
1997	348	284	82%	30	9%	174	13	110	17
1998	361	22	6%	288	80%	7	181	15	107
1999	368	54	15%	240	65%	32	169	22	71
2000	399	10	3%	335	84%	4	246	6	89
2001	386	191	49%	47	12%	134	34	57	13
2002	382	70	18%	63	16%	50	41	20	22
2003	382	43	11%	40	10%	19	17	24	23
2004	413	83	20%	234	57%	43	152	40	82
2005	450	184	41%	120	27%	95	84	89	36
2006	367	108	29%	148	40%	74	91	34	57
2007	352	183	52%	25	7%	101	11	82	14
2008	416	31	7%	51	12%	28	34	3	17
2009	369	47	13%	19	5%	4	12	43	7
2010	359	55	15%	19	5%	0	13	55	6

2011	339	118	35%	67	20%	82	43	36	24
Total	10948	2045	19%	2578	24%	1131	1546	914	1032

This table shows that the number of weight increased (decreased) performance measure on a yearly basis by full sample, common measure sample and unique measure sample.

TABLE 5

Descriptive Statistics of the Sample [a]

Variables	N	Mean	Standard Deviation	Min	Q1	Median	Q3	Max
d_attain	8,129	- 0.006	0.137	- 1.000	- 0.046	-	0.043	1.000
d_weight	8,129	- 0.035	1.202	- 16.000	- 0.333	-	0..091	7.727
discriminability	4,833	0.095	0.072	-	0.045	0.078	0.122	0.602
ability	8,129	1.555	1.070	-	1.000	2.000	2.000	3.000
m_tenure	11,321	4.584	4.052	1.000	2.000	3.000	6.000	25.000
weight	8,129	3.178	2.068	0.500	2.000	3.000	4.000	27.000
room	8,129	2.223	1.310	-	1.619	2.603	3.258	4.615
common	11,321	0.578	0.494	-	-	1.000	1.000	1.000
ln_asset	10,146	21.468	1.973	17.099	19.975	21.428	23.162	25.789
roa	9,629	0.013	0.061	- 0.469	0.002	0.008	0.032	0.315
lev	9,629	0.621	0.304	0.058	0.423	0.590	0.764	2.168

The sample includes 11,321 unique measure-years and 15 unique State Owned Enterprise in Korea from 1985 to 2011. Data for performance measure's score and characteristics are manually collected from the performance evaluation reports of SOE. Financial performance data and firm characteristics are obtained from Alio web site. The number of observations varies depending on data availability.

a. Variable Definitions

d_attain =	performance score at measure j and year t - performance score at measure j and year t-1
d_weight =	the weight of measure j at year t - the weight of measure j at year t-1
discriminability =	standard deviation of attain at measure j and year t only for common measures
ability=	quartile ranking based on prior performance score at firm i and year t-1.
m_tenure=	period after measure j comes up in performance measure at first
weight=	points allocated to individual performance measures, adding up to 100 points each year
room =	measured by log (101%-performance score at measure j and year t-1)
common=	1 if measure is common among SOEs, 0 if measure is unique for specific SOE
participation =	1 if target method allows ratee's participation, and otherwise 0
ln_asset =	log value of total asset of firm i at year t
roa =	the ratio of operating income divided by total asset of firm i at year t
lev =	the ratio of total debt divided by total asset of firm i at year t

TABLE 6

Pearson Correlation Matrix [a]

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*
(1) d_attain	1	0.026**	0.105***	-0.305***	-0.007	-0.020**	0.282***	-0.007	-0.004	0.053***	-0.025**
(2) d_weight		1	-0.024	-0.028**	-0.004	-0.432***	0.014	-0.029***	0.001	-0.017	0.011
(3) discriminability			1	-0.049***	0.143***	0.116***	0.088***	0.047***	0.139***	-0.024	0.005
(4) ability				1	-0.037***	0.051***	-0.728***	-0.023**	0.003	-0.01	0.005
(5) m_tenure					1	0.180***	0.157***	0.088***	0.075***	-0.062***	0.027***
(6) weight						1	-0.024**	0.103***	-0.008	-0.029**	0.027**
(7) room							1	0.287***	0.018	-0.023*	0.022*
(8) common								1	0.043***	-0.01	0.003
(9) ln_asset									0	-0.051***	-0.130***
(10) roa										1	-0.514***
(11) lev											1

a. This table presents Pearson correlations in the upper diagonal. The number of observations varies depending on data availability
*, **, *** indicate significance at less than the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed t-tests on mean differences

TABLE 7
The Effect of the Change of Performance Measure Weights on Ratee's Effort

	Dependent variable = *d_attain*	
Independent Variables	Predicted Sign	Coefficient
Intercept		-0.115***
d_weight	+	0.002**
weight		0.003**
room		0.033***
m_tenure		-0.000
common		-0.027***
ln_asset		0.001
roa		0.115***
lev		0.002
Year Fixed Effect		Yes
Number of Observation		7087
adjusted R^2		0.198

This table reports the results of the effect of performance measure weight change on ratee's effort allocation by testing hypothesis 1 using Equation (1)

$$d\_attain_{j,i,t} = \alpha_0 + \beta_1 * d\_weight_{j,i,t} + \beta_2 * weight_{j,i,t} + \beta_3 * room_{j,i,t} + \beta_4 * m\_tenure_{j,i,t} + \beta_5 * common_{j,i,t} + \beta_6 * ln\_asset_{i,t} + \beta_7 * roa_{i,t} + \beta_8 * lev_{i,t} + year\ dummies + \varepsilon_t$$

Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to Table 5 for detailed variable definition

TABLE 8

The Effect of Commonality on Ratee's Effort with changes in Performance Measure Weights

		Dependent variable = *d_attain*		
		common	unique	difference
Independent Variables	Predicted Sign	(1) Coefficient	(2) Coefficient	(3) Coefficient
Intercept		-0.242***	-0.026	-0.216***
d_weight	+	0.004***	-0.003	0.007**
weight		0.006***	-0.003	0.009***
room		0.048***	0.022***	0.026***
m_tenure		-0.001**	0.000	-0.001**
ln_asset		0.003**	-0.001	0.005***
roa		0.201***	-0.012	0.214***
lev		0.01	-0.008	0.018***
Year Fixed Effect		Yes	Yes	Yes
Number of Observation		4300	2787	7087
adjusted R^2		0.23	0.187	0.216

This table reports the regression results of the effect of performance measure weight change on ratee's effort allocation across common and unique measure groups by testing hypothesis 2 using Equation (1)

$$d\_attain_{j,i,t} = \alpha_0 + \beta_1 * d\_weight_{j,i,t} + \beta_2 * weight_{j,i,t} + \beta_3 * room_{j,i,t} + \beta_4 * m\_tenure_{j,i,t} + \beta_5 * ln\_asset_{i,t} + \beta_6 * roa_{i,t} + \beta_7 * lev_{i,t} + year\ dummies + \varepsilon_t$$

where *i, j* and *t* indicate the firm, the individual performance measure and the time period respectively.

Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to Table 5 for detailed variable definition.

TABLE 9
The Effect of Task-Specific Ability on Ratee's Effort with changes in Performance Measure Weights

Panel A

	Dependent variable = *d_attain*	
Independent Variables	Predicted Sign	Coefficient
Intercept		0.008
d_weight		-0.005
ability		-0.048***
*d_weight*ability*	+	0.004***
discriminability		0.107***
weight		0.005***
m_tenure		-0.000
ln_asset		0.000
roa		0.148***
lev		0.006
Year Fixed Effect		Yes
Number of Observation		4061
adjusted R^2		0.238

Panel B

	Dependent variable = *d_attain*	
Independent Variables	Predicted Sign	Coefficient
Intercept		-0.068***
d_weight		0.000
high_ability		-0.055***
low_ability		0.088***
*d_weight*high_ability*	+	0.009***

*d_weight*low_ability*	?	-0.004
discriminability		0.078
weight		0.004***
m_tenure		0.000
ln_asset		0.001
roa		0.144***
lev		0.007
Year Fixed Effect		Yes
Number of Observation		4061
adjusted R^2		0.226

Panel A of this table reports the regression results of the effect of performance measure weight change on ratee's effort allocation with the interactions of ability by testing hypothesis 3 using Equation (2)

$$d\_attain_{j,i,t} = \alpha_0 + \beta_1 * d\_weight_{j,i,t} + \beta_2 * ability_{j,i,t} + \beta_3 * d\_weight * ability_{j,i,t} + \beta_4 * discrimiabilty_{j,i,t} + \beta_5 * weight_{j,i,t} + \beta_6 * m\_tenure_{j,i,t} + \beta_7 * ln\_asset_{i,t} + \beta_8 * roa_{i,t} + \beta_9 * lev_{i,t} + year\ dummies + \varepsilon_t$$

where *i, j* and *t* indicate the firm, the individual performance measure and the time period respectively.

$$d\_attain_{j,i,t} = \alpha_0 + \beta_1 * d\_weight_{j,i,t} + \beta_2 * ability_{j,i,t} + \beta_3 * ability_{j,i,t} + \beta_4 * d\_weight * high\_ability_{j,i,t} + \beta_5 * d\_weight * high\_ability_{j,i,t} + \beta_6 * discrimiabilty_{j,i,t} + \beta_7 * weight_{j,i,t} + \beta_8 * m\_tenure_{,j,i,t} + \beta_9 * ln\_asset_{i,t} + \beta_{10} * roa_{i,t} + \beta_{11} * lev_{i,t} + year\ dummies + \varepsilon_t$$

where *i, j* and *t* indicate the firm, the individual performance measure and the time period respectively.

We incorporate indicator variable *high_ability (low_ability)* as measured highest (lowest) performance score quartile of each ratee in previous year in equation (3)

In equation (2) and (3) of this table, standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to Table 5 for detailed variable definition

TABLE 10
The Effect of Performance Measure Weights on Performance Level

		Dependent variable = *attain*
Independent Variables	Predicted Sign	Coefficient
Intercept		0.844***
weight	+	0.003**
room		-0.063***
m_tenure		0
common		-0.008*
ln_asset		0.005**
roa		0.168***
lev		-0.008
Year Fixed Effect		Yes
Number of Observation		7087
adjusted R^2		0.475

This table reports the regression results of the effect of performance measure weight on ratee's effort allocation measuring the performance Score (*attain*). we define attain as percentage of performance score which individual performance measure achieves

$$attain_{j,i,t} = \alpha_0 + \beta_1 * weight_{j,i,t} + \beta_2 * room_{j,i,t} + \beta_3 * m\_tenure_{j,i,t} + \beta_4 * common_{j,i,t} + \beta_5 * ln\_asset_{i,t} + \beta_6 * roa_{i,t} + \beta_7 * lev_{i,t} + year\ dummies + \varepsilon_t$$

where *i, j* and *t* indicate the firm, the individual performance measure and the time period respectively.

Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to Table 5 for detailed variable definition

TABLE 11

The Effect of Performance Measure Commonality on Ratee's Effort in Accordance with Performance Measure Weight Changes

		Dependent variable = *attain*		
		common measure	unique measure	difference
Independent Variables	Predicted Sign	(1) Coefficient	(2) Coefficient	(3) Coefficient
Intercept		0.775***	0.941***	-0.166***
weight	+	0.004***	-0.001	0.005***
room		-0.061***	-0.064***	0.004
m_tenure		0.000	-0.001	0.001
ln_asset		0.006**	0.002	0.005**
roa		0.256***	0.017	0.239***
lev		-0.005	-0.010	0.005
Year Fixed Effect		Yes	Yes	Yes
Number of Observation		4300	2787	7087
adjusted R^2		0.411	0.544	0.48

This table reports the regression results of the effect of performance measure weight on ratee's effort allocation measuring the performance Score (attain) across common and unique measure groups. we define attain as percentage of performance score which individual performance measure achieves

$$attain_{j,i,t} = \alpha_0 + \beta_1 * weight_{j,i,t} + \beta_2 * room_{j,i,t} + \beta_3 * m\_tenure_{j,i,t} + \beta_4 * common_{j,i,t} + \beta_5 * ln\_asset_{i,t} + \beta_6 * roa_{i,t} + \beta_7 * lev_{i,t} + year\ dummies + \varepsilon_t$$

where *i*, *j* and *t* indicate the firm, the individual performance measure and the time period respectively.

Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to Table 5 for detailed variable definition

TABLE 12
The Effect of the Sign of Performance Measure Weight Change on Ratee's Effort Allocation

		Dependent variable = *d_attain*	
Independent Variables	Predicted Sign	Coefficient	Coefficient
Intercept		-0.252	-0.244***
inc_weight	+	0.010*	
dec_weight	-		-0.008
weight		0.005***	0.005***
discriminability		0.209***	0.206***
m_tenure		-0.001**	-0.001**
room		0.047***	0.047***
ln_asset		0.003***	0.003**
roa		0.186***	0.186***
lev		0.006	0.006
Year Fixed Effect		Yes	Yes
Number of Observation		4061	4061
adjusted R^2		0.233	0.232

This table reports the regression results of the Effect of the sign of performance measure weight change on ratee's effort allocation. We define an indicator variables inc_weight (dec_weight) which takes on the value of one when a performance measure weight increases (decreases) and zero .the regression equation is as follows.

$$d\_attain_{j,i,t} = \alpha_0 + \beta_1 * inc\_weight_{j,i,t} + \beta_2 * dec\_weight_{j,i,t} + \beta_3 * weight_{j,i,t} + \beta_4 * room_{j,i,t} + \beta_5 * m\_tenure_{j,i,t} + \beta_6 * room_{j,i,t} + \beta_7 * ln\_asset_{i,t} + \beta_8 * roa_{i,t} + \beta_9 * lev_{i,t} + year\ dummies + \varepsilon_t$$

where *i, j* and *t* indicate the firm, the individual performance measure and the time period respectively.

Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to Table 5 for detailed variable definition

TABLE 13

The Effects of Task Specific Ability on Ratee's Effort with Changes in Performance Measure Weights

		Dependent variable = *d_attain*	
Independent Variables	Predicted Sign	(1) Coefficient	(2) Coefficient
Intercept		-0.072	-0.104**
d_weight		-0.005	-0.006
ability2		-0.050***	
ability3			-0.038***
*d_weight*ability2*	+	0.005***	
*d_weight*ability3*	+		0.004*
discriminability		0.105**	0.120**
weight		0.005***	0.004***
m_tenure		-0.000	-0.000
ln_asset		0.004**	0.004**
roa		0.213***	0.230***
lev		0.003	0.014
Year Fixed Effect		Yes	Yes
Number of Observation		4061	3247
adjusted R^2		0.252	0.22

This table reports the regression results of the effect of performance measure weight change on ratee's effort allocation with the interactions of different *ability* variables by testing hypothesis 3 using Equation (2). We define ability2 as sorting performance score into quartile based on performance score of all ratees in previous year, not of each ratee (ability) and define ability3 as sorting performance score into quartile based on average performance score of each ratee in previous 2 years (Woods 2012). Thus, we define ability2 (ability3) as rank variables based on performance score quartile.

$$d\_attain_{j,i,t} = \alpha_0 + \beta_1 * d\_weight_{j,i,t} + \beta_2 * ability2(3)_{j,i,t} + \beta_3 * d\_weight * ability2(3)_{j,i,t} + \beta_4 * discrimiabilty_{j,i,t} + \beta_5 * weight_{j,i,t} + \beta_6 * m\_tenure_{j,i,t} + \beta_7 * ln\_asset_{i,t} + \beta_8 * roa_{i,t} + \beta_9 * lev_{i,t} + year\ dummies + \varepsilon_t$$

where *i, j* and *t* indicate the firm, the individual performance measure and the time period respectively.

Standard errors are corrected for heteroskedasticity using the Huber-White robust standard errors clustered by firm. The symbols *, **, and *** correspond to the 10-percent, 5-percent, and 1-percent significance levels (two-tailed), respectively. Please refer to Table 5 for detailed variable definition

국문초록

성과지표 탈락과 가중치 변화에 관한 연구

본 학위논문은 성과지표 탈락의 결정요인과 성과지표의 가중치 변화가 피평가자의 인센티브에 미치는 영향을 분석한 2개의 연구논문으로 구성되어 있다. 첫 번째 연구는 성과평가시스템에서 사용되고 있는 성과지표가 탈락되는 요인을 연구하였다. 성과지표의 탈락요인과 관련된 연구는 아직까지 많이 연구되지 않은 분야이다. 본 연구에서는 한국의 공기업 경영평가 자료를 활용하여 성과지표의 탈락을 결정하는 경제적이고 행동과학적인 요인들을 연구하였다. 연구 결과는 다음과 같다. 첫째, 성과지표가 피평가자들을 차별화하지 못하는 경우 차년도 지표선정에서 탈락될 가능성이 높은 것으로 나타났다. 둘째, 계량지표의 경우 성과수준이 낮을수록 차년도 지표선정에서 탈락할 가능성이 높았다. 셋째, 목표설정과정에 피평가자가 참여하지 않는 성과지표가 차년도 지표선정에서 탈락될 가능성이 높은 것으로 나타났다.

본 연구는 다음과 같은 측면에서 공헌점이 있다. 첫째, 성과지표 생존기간의 결정요인을 찾아냄으로써 성과평가연구에 기여하였다. 둘째, 기존연구들이 주로 성과평가 수행단계에 관심을 두고 있는 반면, 본 연구에서는 전년도 성과평가결과가 차년도 성과평가시스템 설계에 미치는 영향을 분석함으로써

성과평가시스템의 역동적인 측면을 이해하는데 기여하였다. 마지막으로, 본 연구는 평가자와 피평가자의 인센티브가 어떻게 성과평가 제도설계(성과지표 탈락)에 영향을 미치는지 분석하였다는 점에서 성과평가연구에 공헌하고 있다.

두 번째 연구는 다중업무 대리인 환경하에서 성과평가시스템의 설계자가 성과지표의 가중치를 변경하는 경우 피평가자들이 어떻게 반응하는지 실증적으로 연구하였다.

본 연구의 결과는 다음과 같다. 첫째, 성과평가제도 설계자들이 성과지표의 가중치를 변경하는 경우 피평가자들은 성과지표의 가중치 변화방향과 동일한 방향으로 자신의 노력을 변화시킨다. 이러한 결과는 성과지표의 가중치가 증가하는 경우 피평가자는 동일한 노력을 기울여도 더 많은 혜택을 누릴 수 있기 때문에 피평가자들이 가중치가 증가된 성과평가지표에 더 많은 노력을 기울일 것이라는 다중업무 대리인 이론을 지지하고 있다(Brickley, Smith and Zimmerman 2009). 둘째, 고유지표의 가중치가 증가하는 경우보다 공통지표의 가중치가 증가하는 경우 피평가자는 해당지표에 더 많은 노력을 기울인다. 이러한 결과는 피평가자들이 상대평가 지표에 더 많은 노력을 기울인다는 선행연구들(Frederickson 1992; Matsumura and Shin 2006) 과 상대평가 지표가 존재하는 경우 성과가 향상된다(Tafkov 2013)는 선행연구결과와 일치한다. 셋째, 성과지표의 가중치가 증가하는 경우, 피평가자들은 해당 지표와 관련된 능력이 높은 경우 더 많은 노력을 기울인다.

본 연구는 다음과 같은 측면에서 성과평가 연구에 공헌하고 있다. 첫째, 본 연구는 선행연구와 비교하여 성과지표의 가중치 변화와 이에 따른 피평가자의 노력의 변화를 명확하게 실증적으로 보이고 있다. 둘째, 본 연구는 상대평가정보의 피드백이 피평가자가 더 많은 노력을 기울이도록 하는 유용한 수단이라는 점을 실증적으로 보였다. 마지막으로, 본 연구는 성과지표와 연계된 업무능력이 성과지표의 가중치 변화와 피평가자의 노력간의 관계에 영향을 미친다는 점을 실증적으로 제시하였다는 점에서 성과평가 연구에 기여하고 있다(Bonner and Sprinkle 2002).

주요어: 차별화 수준, 성과지표 탈락, 성과지표 가중치, 공통지표, 성과지표와 연계된 업무능력, 피평가자의 인센티브
학번: 2010-30147