Melco Management Accounting Research Discussion Paper Series

Why is Subjective Evaluation Used for Long-Term Incentives?

March 2020

Toru Ogasawara
Lecturer of Faculty of Commerce, Kumamoto Gakuen University
2-5-1 Oe, Chuo-ku, Kumamoto 862-8680 JAPAN
togasawara@kumagaku.ac.jp

The Melco Foundation

Nagoya, Japan

Why is Subjective Evaluation Used for Long-Term Incentives?

Toru Ogasawara

Faculty of Commerce, Kumamoto Gakuen University, Kumamoto, Japan

Abstract

Subjective evaluation, an assessment of employees to gauge their ability and achievements, can be used to determine long-term incentives such as promotion and salary increases. However, existing literature has not yet fully established the relationship between long-term incentives and subjective evaluation. This study uses personnel data from a Japanese Company to empirically demonstrate the role of subjective evaluation in accumulating historical information to identify with ease the high-performing employees. In particular, the study tested characteristics of subjective evaluation: historical information and the difference between high skill and low skill employees. The study found that the performance and promotion in the previous year positively correlate with subjective evaluation in the current year and the discrimination of subjective evaluation increases over time. Additional analysis shows that higher performance in previous years can enhance performance in the coming years. The findings prove why subjective evaluation is often used to determine long-term incentives.

Key Words: Subjectivity, Performance Evaluation, Long-term incentives, Promotion

1. Introduction

Incentive systems are important for companies to ensure that their employees work efficiently (Jensen and Meckling, 1976; Holmström, 1979; Banker and Datar, 1989; Feltham and Xie, 1994; Lambert 2001). It is important for companies to set the appropriate combination of performance evaluation and incentives (Jensen and Meckling, 1976; Lambert, 2001). Many studies examine the design of incentive systems by considering the relationship between performance (performance measure or evaluation) and short-term incentive plan, such as financial measures (Banker, 1996), non-financial measures (Banker, 2001), subjective assessment (Ittner et al., 2003; Gibbs et al., 2004), relative performance evaluation (Matsumura and Shin, 2006). In addition, companies often use long-term incentive plans,

with multiple payments, such as promotion and salary increase, over time. Previous literature shows that these long-term incentives play some important roles, such as (1) replacing the short-term incentive (Gibbons and Murphy, 1992; Ederhof, 2011 and (2) giving incentive to white-collar employees whose performance is difficult to measure (Baik et al., 2016). To examine the relationship between these long-term incentives and performance evaluation empirically, internal corporate data should be used. Few studies have empirically explored the relationship between performance evaluation and long-term incentives. For instance, Cichello et al. (2009) and Campbell (2008) examine the relationship between long-term incentives and performance evaluation. Cichello et al. (2009) also examine the relationship between promotion and financial indicators. Campbell et al. (2011) studies the relationship between non-financial measures and promotion as long-term incentives. These papers provide useful knowledge. Subjective evaluation, such as employee assessment (e.g., ability and achievements) by superiors, can be used to determine long-term incentives (e.g., promotion, salary increases). However, the relationship between long-term incentives and subjective evaluation has not been fully investigated. Grabner and Moers (2013) empirically investigated the relationship between subjective evaluation and promotion, but did not focus on the role of promotion as a long-term incentive. They used data from a bank in the Netherlands, which has different types of jobs. In their findings, they show that subjective evaluation strongly correlates with promotion to perform different jobs, whereas objective evaluation strongly correlates with promotion for the same job. Thus, subjective evaluation can capture performance of an employee in a new job, which objective measures in the current job cannot measure. However, Grabner and Moers could not explain why subjective evaluation is used for long-term incentives regardless of job change. An answer to this research question can reveal the role of subjective evaluation and indicate the limits of objective performance measures. Therefore, the research contributes not only to subjective evaluation (Bol, 2008), but also to objective evaluation. To answer the research question, this study statistically analyzes personnel data from a Japanese Company. It also presents empirical evidence on the role of subjective evaluation in long-term incentives.

Another difference between long-term and short-term incentives is that the payment of the reward spans multiple periods in long-term incentives. Once determined, long-term incentives, such as promotion or pay-raises, are difficult to eliminate through demotions or pay cuts. In addition, promotion involves job assignment and is a significant factor, because decision-making for a managerial position is crucial for company performance (Bertrand 2003). In contrast, the cost of incorrectly giving long-term incentives to low-skilled employees is much higher than when short-term incentives are given, because the sum of payments for long-term incentives is often greater than the one-time payment of short-term incentives. To determine who obtains the long-term incentives, companies have to distinguish between high-skilled and low-skilled employees. This study argues that subjective evaluation is more suitable for long-term incentives than an objective assessment. To achieve that, the study empirically considers two characteristics of subjective evaluation: (1) historical information and (2) difference between high-skilled and low-skilled employees by accumulating historical information. Personnel data from a listed Japanese IT Company have been used to clarify the above-mentioned characteristics. For the latter hypothesis, this study visualizes and statistically tests a series of changes for subjective evaluation to identify

high- and low-performing employees over time. To ensure that cognitive bias and favoritism does not significantly influence the results, this study tests whether subjective evaluations are sufficient for predicting the future performance of employees. The results show that employees with high subjective evaluation in the first years of the analysis tend to achieve higher objective performances in later years. These analytical results show that (1) evaluators should incorporate historical information in subjective evaluation; (2) subjective evaluation is gradually established over time, by incorporating historical information; and (3) these characteristics of evaluation show that on average, subjective evaluations are adequate for predicting future outcomes.

This study makes the following three contributions. First, it contributes to a series of studies about subjective evaluation (Bol, 2008). Unlike previous studies, it empirically demonstrates a new role of subjective evaluation in determining the performance of employees by accumulating historical information. Secondly, the study focuses on time-series transition to demonstrate the difference between objective and subjective evaluations. In reality, evaluations are carried out many times over a long period. Some research methods, such as laboratory experiments or surveys have, time constraints, and therefore ineffective in capturing the long-term aspects of this performance evaluation. In this study, it was possible to visualize and test the long-term fluctuation patterns of objective and subjective evaluations using archival analysis. Thirdly and lastly, the study contributes to a series of studies on long-term incentives. Previous studies, such as Grabner and Moers (2013), do not fully explain the management practices that require the adoption of subjective evaluation for long-term incentives such as promotion or salary increases. Grabner and Moers (2013) posit that subjective evaluation is more suitable for promotions, but only for positions with different operations. Nevertheless, they do not explain why subjective evaluation is employed in long-term incentives without necessarily changing the job placement. To fill that gap, this study shows that subjective evaluation can determine performance of an employee over time. Such a role of subjective evaluation is important for long-term incentives.

The paper is divided into seven sections. The next section presents the relevant previous literature, and then Section 3 contains the hypothesis development. Section 4 describes the performance evaluation system in the research site and the methods of testing the hypothesis. Section 5 presents the results while Section 6 explores additional analysis to show that effects such as cognitive bias or favoritism do not strongly influence the results in the previous section. Finally, Section 7 describes the contribution and limitations of this study.

2. Previous Literature

2.1. Long-term incentive

Previous studies have clarified many aspects of the relationship between the various performance evaluations and bonus plans. Banker et al. (1996) and Banker (2000) use the retail industry data to test the introduction effects of an incentive plan based on financial indicators. Retail industry has introduced an incentive plan bonus, which depends on the sales

of each store. These studies show that the effect of the incentive plan increases over time, and the performance of employees who stay in the store with the new incentive plan is higher than the performance of the employees who leave the store. Their results indicate that the combination of performance evaluation and incentives has a significant impact on a company. Other studies focus on different measures or evaluations in addition to financial measures. Banker (2000) analyzes the effect of the incentive plan based on non-financial measures such as customer satisfaction. Banker (2000) shows that non-financial measures improve after introducing the incentive plan, which leads to an improvement in financial performance. Matsumura and Shin (2006) analyze the data of a postal business company, the company that introduced an incentive plan based on the relative performance evaluation. Their results show that common uncertainty, such as weather and geography, increases the effect of the incentive plan. In addition, some studies focus on subjective evaluation, such as assessment and discretionary evaluation by the superiors. Gibbs et al. (2004) undertake a questionnaire survey on subjective evaluation in the automobile industry. Their results show that subjective evaluation is often used in some situations, and so it becomes difficult to achieve the budget goals because of the changes in environments. These studies examine and explain many aspects of the relationship between short-term incentives and some performance measures or evaluations, financial measures, non-financial measures, relative performance evaluation and subjective evaluation.

However, companies employ not only short-term incentive, but also long-term incentive systems, such as promotion and salary increases; they pay these long-term incentives over a long period. Why do companies use long-term incentives in addition to the short-term incentives? Long-term incentives also play an important role in companies. Among others, they substitute short-term incentives and incentivize white-collar employees whose performance is difficult to measure.

One of the implications of long-term incentives is that if current performance can determine compensations in the next periods, the compensations after the next periods can have incentive effects for the agents. Therefore, a long-term incentive can substitute a short-term incentive agreement. According to Gibbons and Murphy (1992), the incentive coefficient of bonus contract in the early career is theoretically smaller than late in the career, because performance in the early career affects compensations in the later career and imposes more risks on agents in their early careers. Ederhof (2011) empirically shows that the incentive coefficient of employees in their early career is smaller than the incentive coefficient of employees later in their career.

Some previous studies note that long-term incentives are used for white-collar employees whose performances are difficult to measure. A long-term incentive is more suitable for white-collar employees because they often multitask, and that makes it hard to link their performance to a bonus (Holmström and Milgrom, 1991); it takes a long time to observe the results of their tasks. From this perspective, Grabner (2014) and Baik et al. (2016) study the relationship between performance evaluation and incentives. Grabner focuses on the trade-off between incentives and creativity. In companies where creativity is strategically important, performance-based-pay can harm intrinsic motivation and prompt employees to take only short-term actions with outcomes observable in short-term periods. In contrast, if

incentive-pay is entirely not used, it can cause the problem of producing products and disregard profit. Grabner refers to this as "art for art's sake"). The author empirically shows that companies that place strategic importance on creativity can break this trade-off by combining assessment with subjective evaluation and performance-based compensation; this proves that subjective evaluation complements the performance-based compensation. Baik et al. (2016) focus on the relationship between explicit incentives (e.g., bonuses or stock option) and implicit incentives (e.g., possibility of promotion) [1]. They show that the implicit and explicit incentives are weaker and stronger, respectively, in higher hierarchal position because the information value of the financial measures increases. A higher hierarchal position allows the employee to have more authority. This finding is consistent with related research results about delegation, which posited that more delegation leads to an increase in the proportion of incentive compensation (Nagar, 2002; Abernethy et al., 2004; Moers, 2006).

2.2. Performance evaluation and long-term incentive

Long-term incentives play some important roles in the company. Some studies have presented some empirical evidences about the combination of performance measures and long-term incentives; such as Cichello et al. (2009), Campbell (2008), Grabner (2014), and Moers (2013) examine the relationship between the promotion and performance evaluation.

Cichello et al. (2009) analyze the relationship between promotion or turnover of divisional managers and the divisional return on assets (ROA). Their results show that the divisional ROA strongly correlates with turnover than promotion, indicating that the effect of the accounting indicator on promotion is relatively small. Campbell (2008) focuses on the incentive effect of promotion and non-financial measures. The author shows that the impact of non-financial measures for the promotion is significantly positive even after controlling the impact of the financial measures, and the action to improve non-financial measures becomes lower after promotion. These results indicate that non-financial measures can provide useful information on promotion and can extract employee performance information with promotion incentives. Cichello et al. (2009) and Campbell (2008) provide some useful knowledge about the relationship between financial or non-financial measures and long-term incentives. Subjective evaluation is often used to determine promotion or salary increases.

In this regard, Dohmen (2004), DeVaro and Waldman (2012) have revealed that subjective performance evaluation can predict promotion. Regardless, there is little focus on the role of subjective evaluation in long-term incentives. Grabner and Moers (2013), among a few studies, focus on the relationship between promotion and subjective evaluation; they show

---

[1] Baik et al. (2016) refer to stock option compensation schemes as long-term incentives in a different context to long-term incentives covered in this study. While the long-term incentive in this study is based on the assumption that payments will last for a long time, the stock option type of exchange does not mean that payments themselves will last for more than one period.

that objective evaluation strongly correlates with the same promotion (i.e., the same job is retained after promotion). However, they also show that subjective evaluation strongly correlates with different promotions (i.e., the job is changed after promotion) than with financial measures. These results show that the objective evaluation can provide useful information about the employee for the current job, while subjective evaluation can be more useful for providing information about the employee for a new job. Grabner and Moers focus on the function of job placement rather than the function as a long-term incentive of promotion. Therefore, their results do not explain why subjective evaluation is used for long-term incentives without a change in job placement.

2.3. Roles of subjective evaluation for long-term incentive

Long-term incentive differs from short-term incentive in that it is paid over multiple periods. Once promotion or salary increases are determined, demotion or pay-cut becomes difficult. Baker et al. (1994a) indicate that demotion is rarer than promotion; therefore, it is important to evaluate employee productivity or ability before promotions (Becker, 1962; Gibbons and Waldman, 1999). Labor economics empirically show that some indicators that reflect the ability of the employee "seniority (length of service)" and "education" are important factors in determining the wage (Lazear, 2009; Lazear and Oyer, 2013).

Using time-series variation of subjective evaluation, this study empirically shows that subjective evaluation is more suitable than objective evaluation for long-term incentives; it can measure an employee's performance better than objective evaluation. Unlike objective evaluations, subjective evaluations consider historical information and has some functions due to the characteristics of considering past information. These functions include (1) downward revision of goals in situations where it is not possible to achieve the goal owing to unexpected factors (Gibbs et al., 2004; Merchant and Manzoni, 1989) and (2) referrals to historical performance in old incentive systems in new incentive system (Woods, 2012). If the subjective evaluation properly accumulates the historical information, it can update information about the productivity or ability of employees, meaning that it gains accuracy over time [2].

---

[2] If we are only interested in accumulating past information, we can consider making an objective evaluation using formulas, including past objective results. However, to evaluate employee productivity and capability properly based on a formula that includes past objective results, we should consider indicators that reflect productivity and capabilities in advance. Some of these indicators are easy to predict; they include educational background and length of service. On the other hand, many indicators are more difficult to predict and they include economic conditions, relationships with competition, differences between departments, and individual employee situations. Therefore, subjective evaluation is adopted to provide flexibility in evaluation. However, the analysis results of this study do no support this prediction.

3. Theory Development and Hypotheses

3.1. Historical information

Subjective evaluation requires some historical information, such as previous performance and promotion in previous year, about an employee as the benchmark. This historical information reflects the ability of the employee and so it is used to gauge performance or ability. Gibbons and Waldman (1999) also used the same method in their analytical studies. Similarly, this study considers the performance and promotion in previous years.

To assess the ability of the employee, the evaluator may refer to performance in the previous year. This prediction is based on a premise that a high-performing employee can retain it in other periods. Analytical studies such as those by Becker (1962), Prendergast (1999), Gibbons and Waldman (1999, 2006), and Lazear (2009) used the same assumption. Past performances are often considered as the expected value of the fiscal year results. In fact, some studies empirically show that the outcome in the previous year can affect the subjective evaluation in the current period (Thorsteinson et al., 2008; Woods, 2012). Therefore, we develop the following hypothesis based on this information.

Hypothesis 1.1
Ceteris paribus, subjective evaluation increases with the performance in a previous year.

Promotion in the previous year can also affect subjective evaluation in the current year because it can indicate that the employee has achieved high performance until the previous year (Lazear and Rozen, 1981). In particular, if another person has determined the promotion, the ability of the promoted employee is predicted to be higher than the ability of the non-promoted employees. The fact of promotion shows that other evaluators have praised the ability of the employee (DeVaro and Waldman, 2012). Therefore, we develop the following hypothesis.

Hypothesis 1.2
Ceteris paribus, subjective evaluation increases if the employee were promoted in the previous year.

3.2. Discrimination of subjective evaluation

The payment of long-term incentive is over multiple periods. If a poor-performing employee is accidentally promoted or given a salary increase, the cost of such a mistake will be very high. Therefore, it is important to identify employee ability. This study focuses on the role of subjective evaluation; it shows that it can identify high- and poor-performing employees based on their historical information. This improvement in the discrimination of the subjective evaluation is attributed to the fact that it updates and improves the mean and accuracy of evaluation, respectively.

Updating the mean of evaluation indicates that an employee's subjective evaluation improves/falls if the employee performed above/below the expected level. If there is a shift in the evaluation, the subjective evaluation will shift to a value reflecting the ability of the employee if it is properly managed[3]. Over time, therefore, the changes in subjective evaluations will result in higher subjective ratings for high-performing employees and lower subjective ratings for low-performing employees. As a result, the difference in subjective evaluations between high-performing and low-performing employees is likely to increase over time. Therefore, we develop the following hypothesis 2.1.

Hypothesis 2.1
  Over time, the difference between the subjective evaluation of high- and low-performing employees increases.

The improvement in accuracy of evaluation owing to an increase in information available indicates that subjective evaluation is becoming more accurate. These effects are best described the following analogy. When an employee achieves an expected level of performance, the mean of evaluation is not updated. However, the expected level of subjective evaluation is ensured if the employee achieves the expected performance. This would increase the accuracy of subjective evaluation[4]. Therefore, we develop the following hypothesis 2.2.

Hypothesis 2.2
  Over time, the accuracy of subjective evaluation increases.

From the above discussion, this study considers the effect of time to increase subjective evaluation discernment as the following two effect: (1) updating the mean of evaluation and (2) improving the accuracy of evaluation. This idea is consistent with Holmström's (1999) model, a leading theoretical study by Career Concern. The model relies on the agent's ability to produce results; thus, evaluators need to estimate the agent's ability from the results the agent has produced to determine the amount of compensation. Holmström analytically demonstrates the relationship between the signal z_t for the agent's ability obtained only from the current period's performance and the estimated value m_t for the agent's ability obtained from cumulative performance up to the previous period. According to Holmström, the

[3] Owing to cognitive biases, subjective evaluations may not be properly managed; it may not move to the values that reflect the inherent abilities of employees. This effect will be discussed in an additional analysis section.

[4] The concept of accuracy is consistent with the accuracy in Banker and Datar (1989), and it is considered as the degree of variation in subjective evaluation (reciprocal of variance). However, in the theoretical model of Banker and Datar (1989), the context is slightly different in that the parameter to be estimated is the agent's effort, whereas the parameter to be assumed in this study is the agent's ability.

estimated value in the next period m_{t+1} is a weighted average of z_t and m_t. Therefore, when z_t is higher than m_t ($m_{t+1} > m_t$), the estimate for the agent's ability moves upward; the converse holds true. Furthermore, the signals (z_t) for agent's ability reduces the variance. This means that additional signal increases the accuracy of estimates for the agent's ability m_{t+1}. In Holmström's model, h_{t+1} is the sum of the accuracy h_t of the estimates up to the previous period and the accuracy h_ϵ of the current period signal. It represents the accuracy (reciprocal of variance) of the estimate m_{t+1} for the agent's ability. These results hold true not only under the setting of a specific principal-agent model, but also in the context of general statistical estimation as follows. In this context, the signal z_t estimates the posterior distribution of η from the signal ($z_1, \ldots, z_t$) obtained so far according to the normal distribution in which the parameter η is an average and $1/h_\epsilon$ is the variance. As the sequence of signals ($z_1, \ldots, z_t$) increases, the average m_t of the posterior distribution shifts to the true value of the parameter η and the variance $1/h_t$ of the posterior distribution approaches 0. Therefore, the discrimination power of subjective evaluation increases as time passes and information is accumulated, subjective evaluation shifts, and the accuracy of subjective evaluation increases. Therefore, we consider the discriminative power to be the combined effect of the transition of evaluation and the improvement of accuracy. We then develop the following hypothesis.

Hypothesis 2.3
   Over time, the effect of the combined effect of both increases.

4. Research Method

4.1. Research site

This study tests the developed hypothesis in the previous section using a dataset of personnel data, from 2004 to 2013, from a Japanese Company (referred to as Company A in this study). Company A belongs to the IT industry and it is composed of two divisions that provide different products and services. Each division has two different types of jobs: selling and technical jobs. In addition, the company has adopted a job grade system, which has no strict correspondence between the grades and job assignments. However, some particular job assignments with higher hierarchal positions require higher grades than some standard jobs. Company A determines salary by grades and it has nine grades.

The company has two types of assessments: assessment for bonus (nine-stage evaluation), which is performed twice a year in summer and winter, and assessment for salary increase (six-stage evaluation), which is performed once a year after the bonus assessment. The boss in some groups (7–8 people in each group) performs two types of assessments, based on both divisions and grades. In the bonus assessments of summer and winter, the policy is to assess the employee's performance in each year. In contrast, in the salary increase assessment, the policy is to include the bonus assessment at the rate of 40%. The other 60% is subjectively determined by other aspects such as employee ability and working attitude. Owing to data constraints, this study cannot use data on employee salaries. Nonetheless, these two types of

assessments are still important variables for determining employee salaries. The relationships between each assessment and salary are as follows, respectively. The relationship between the bonus and bonuses assessment in Company A is as follows. First, each bonus pool is set each year and then the bonus amount for employees in the group is determined and distributed in response to the bonus assessment. Therefore, amounts of bonuses and bonuses assessment are indirectly linked via the bonus pools. The relationship between base salary and salary rise assessment in Company A is as follows. First, the base salary is largely determined in response to the employee's grade (the nine steps described above) and then the amount of the base salary is adjusted based on more detailed levels in each grade. Rise assessment is an evaluation of whether to raise the more detailed levels in each grade. Note that Company A has used this system since 2004.

4.2. Variable Measurement

4.2.1. Subjective evaluation

This study calculates the variable for subjective evaluation using two types of assessments in Company A: bonus and rise assessments. The boss determines the rise assessment under the company's personnel system. This rise assessment is determined by the employee's performance in this year (*i.e.* bonus evaluation) at the rate of 40% and the other subjective factor at the rate of 60%. Therefore, this study operationalizes a subjective evaluation of employee i in period t through the following equation (as $subjective1$). This is multiplied by the formula $\frac{6}{7}$ to adjust the difference in the scale.

$$subjective1_{i,t} = rise_{i,t} - 0.4 \times \frac{6}{7} \times performance_{i,t}$$

$rise_{i,t}$: rise assessment of employee i in year t (6 stages)
$performance_{i,t}$: bonus assessment of employee i in year t (7 stages)

In contrast, the evaluation policy of Company A's rise assessment is not enforceable. This assessment can be determined at the discretion of the evaluator. To avoid the cognitive load on the evaluation, the evaluator may not aggregate the two types of assessment, but roughly determine the rise assessment by anchoring it (Kahneman et al., 1974). The rise assessment still includes the subjective evaluation by the evaluator, even if it is not completely determined in accordance with the policy. However, the variable of subjective evaluation in the above could not be properly measured, because the operationalization is based on the company's policy. For the robustness of the analysis at this point, this study also operationalizes the other variable for subjectivity ($subjective2$) as the following equation. The variable of $Mrise_p$ is the average value of the rise assessment within each bonus assessment (6 stages). Moreover, the subscript "p" represents the stage of the bonus assessment (*i.e.* $p \in \{1,2,3,4,5,6,7\}$).

Note that this variable of subjectivity is the same as the residual of the regression model, with the rise assessment as the dependent variable and dummy variable for bonus assessment as independent variables. In this sense, the variable of $subjective2$ can be interpreted as the variable of the rise assessment after controlling the bonus assessment as dummies.

$$subjective2_{i,t} = rise_{i,t} - Mrise_{p(i,t)}$$

$rise_{i,t}$: the rise assessment of the employee i in year t
$Mrise_{p(i,t)}$: the mean of the rise assessment in samples, which have bonus assessment p
$p(i,t)$: the bonus assessment of employee i in year t (7 stages)

Höppe and Moers (2011) describe two types of the subjectivity. One type of subjectivity is to be incorporated separately from the objective evaluation. The other type is the subjectivity, which weighs different objective measures. This study focuses on the former type of subjectivity.

4.2.2. Performance in the previous year

Under Company A's assessment policy, bonus assessment, based on the current year's performance, is carried out twice a summer and winter of the year. This study uses the bonus assessment as a variable for the employee's performance ($performance$). However, bonus assessment is evaluated by superiors and not officially or mechanically. Consequently, the concern is that the evaluation of bonuses used as a proxy variable for results in this study includes the subjectivity of superiors, and thus, it does not appropriately reflect the results of the employees. We interviewed a research site about this. Company A has a policy that bonuses are evaluated based on 100% of the current year's performance, unlike salary increases. Therefore, although supervisors evaluate bonuses, they tend to rely more on objective figures, such as the operating results and the number of program bugs. Therefore, there is very little room for superiors' subjectivity to be involved in the evaluation of bonuses. [5]

4.2.3. Promotion in the previous year

Company A has adopted the job grade system described above. This study considers a promotion as a raise in this grade ($promotion$).

---

[5] Therefore, in this study, we treat bonus assessment ($performance$) as a proxy variable for objective assessment and compare it with subjective evaluation ($subjective1$ and $subjective2$) defined above. In the context of comparison with subjective evaluation, the objective is to consider the role of subjectivity in appraisal by interpreting bonus evaluation as "a more objective evaluation" rather than "complete objective evaluation."

4.2.4. Control variables

The study uses the number of years ($career$) the employee has worked for as the control variable. Many studies, such as Becker (1964) and Lazear and Oyer (2013), in labor economics assume that the ability of an employee increases by the time elapsed. Based on this assumption, this study adopts the number of years of service in Company A as a proxy for career lengths ($career$). However, Company A also employs workers who have left other companies. In this case, the career length of this worker does not correspond to the length of service. Therefore, this study also uses a dummy variable that takes 1 if the employee has no working experience in another company, to distinguish between these types of employees (new).

Then, the performance in the previous year ($performance_{i,t-1}$) is incorporated in the regression model. Notably, the performance in the current year is used as a control variable ($performance_{i,t}$). The grades in Company A's system are also controlled ($grade$). According to Gibbons and Waldman (2006) in their theoretical research on the internal labor market, it is economically efficient that an employee with high-ability is assigned to a high hierarchal position. Some classic empirical studies of the internal labor market, such as those by Medoff and Abraham (1980) show a positive correlation between grade and performance. In addition, other control variables are job types, divisions, and year. These effects are controlled as dummy variables in the regression model.

4.3. Regression model

4.3.1. Historical information

Hypothesis 1.1 and hypothesis 1.2 are tested using the following linear model.

$$subjective_{i,t} = \beta_0 + \beta_1 performance_{i,t-1} + \beta_2 promotion_{i,t-1} + controls + \epsilon_{i,t}$$

4.3.2. Discrimination

To test hypothesis 2.1, it is necessary to identify the high-ability and low ability employees in a sample. To identify the two groups of employees, this study uses the average value of the subjective evaluation ($subjective1$ and $subjective2$) for each employee. In particular, the top 10 percent of employees of this value is labeled as "$High$" and the lower 10 percent "Low." However, if some employees have not worked for all nine years (from 2004 to 2013), then it may be difficult to compare employees properly because the average values of subjective evaluation have different terms among them. To handle this problem for the analysis hypothesis 2.1 and later, this study excludes some samples that have one or more missing values in these nine years. Note that this operation can cause some problems related to survival bias. In particular, Low groups can be affected by such a survival bias because

employees who continued to receive low subjective evaluations would quit. Therefore, there is a high possibility that such employees are not enrolled until 2013, which is the final year of the analysis period. This point will be described later.

The hypothesis is tested using the following regression model.

$$mean_{t,j} = \beta_0 + \beta_1 time_{t,j} + \beta_2 High_{t,j} + \beta_3 time_{t,j} \times High_{t,j} + \epsilon_{t,j}$$

$mean_{t,j}$: The mean of evaluation within group $j (= High, Low)$ in year t

$time_{t,j}$: Year
$High_{t,j}$: Dummy variables if the sample belongs to group $High$

If the subjective evaluation is updated by accumulating past information, then the difference in subjective evaluation between $High$ and Low may increase over time. Therefore, if β_3 is significantly positive, hypothesis 2.1 is supported.

Next, hypothesis 2.2 is tested using the following regression model. If subjective evaluation becomes more accurate by accumulating historical information, the standard deviations of the groups are considered to reduce over time. This study operationalizes the accuracy of subjective evaluation as the reciprocal of the standard deviation within the group. Under hypothesis 2.2, both standard deviations (within $High$ and Low) could equally become accurate. Therefore, the hypothesis can be validated if the regression models include each single effect only. However, the survival bias can exist in the Low group and can affect the estimation of the effects of time to subjective evaluation. In particular, the effect can be underestimated, because employees who continue to receive low evaluations would quit and become unobservable in the dataset. The interaction terms enable the observation and focus on the effect in the $High$ group, which has little survival bias. From the above discussion, if β_1 is significantly positive, hypothesis 2.2 is supported in both groups. In contrast, if only β_3 is significantly positive, hypothesis 2.2 is supported only in the $High$ group.

$$1/sd_{t,j} = \beta_0 + \beta_1 time_{t,j} + \beta_2 High_{t,j} + \beta_3 time_{t,j} \times High_{t,j} + \epsilon_{t,j}$$

$sd_{t,j}$: The standard deviation of evaluation within group $j (= High, Low)$ in year t
$time$: Year
$High$: Dummy variables if the sample belongs to group $High$

Finally, the test of hypothesis 2.3 is undertaken using the following regression model. From the previous discussion, hypothesis 2.3 is supported if β_3 is significant.

$$mean_{t,j} \times 1/sd_{t,j} = \beta_0 + \beta_1 time_{t,j} + \beta_2 High_{t,j} + \beta_3 time_{t,j} \times High_{t,j} + \epsilon_{t,j}$$

5. Results

5.1. Sample selection

    The first step in the sample selection process is to keep all samples without missing values in all variables used in this study. The second step is to exclude samples with special cases about bonus assessment or rise assessment. In Company A, some assessment (for bonus and rise) is determined in advance. A special example of such a case is when ensuring a fixed allowance for mid-career employees. Because the assessment of these samples is not really evaluated, they were excluded from the analysis. The third step is to exclude samples of employees with grades 1 or 2. In this company, all employees are automatically promoted up to grade 3 and promotions up to this grade are different from promotions to other levels they are not associated with the evaluation of employees. The purpose of this study is to verify the role of subjective evaluation for promotion. Thus, promotions up to grade 3 are beyond the research purpose. Therefore, it excludes samples of employees under grade 2. However, with respect to this operation, the main results of the present study were not affected. After these procedures, the total sample becomes 2980 (employee × year).

5.2. Summary statistics

    Table 1 shows the descriptive statistics of variables used in this study. According to Table 1, the standard deviations of $subjective1_{i,t}$ and $subjective2_{i,t}$ are 0.66 and 0.62, respectively. Then, the standard deviation $performance_{i,t}$ is 0.69. The results indicate that the variables of subjective evaluation vary in the same extent as the variable of bonus assessment. Further, considering the variable of subjective evaluation based on Company A's personnel policy $ubjective1_{i,t}$, its maximum and minimum values are 4.46 and 0.31, respectively. For the variable of the subjective evaluation obtained by subtracting the average value of the bonus rating $subjective2_{i,t}$, its maximum and minimum values are 2.17 and -1.78, respectively. These results indicate that even if bonus assessment were the same, the difference in rise assessment can be near 4 at most.

Figure 1 shows the heat map of rise assessment ($rise$) and bonus assessment ($performance_{i,t}$). The heat map describes the distribution as high (low) and observation frequency in the sample becomes closer to red (blue). For example, if bonus assessment ($performance_{i,t}$) was 4.0, the rise assessment ($rise$) is likely to get an evaluation of 4, with a color close to red in Figure 1. In contrast, the color of rise assessment 5 or 3 in bonus assessment 4.0 is intermediate between red and blue. These results indicate that evaluators' discretion can affect rise assessment in giving incentives. In other words, variations in the vertical side show the degree of discretion by the subjectivity of evaluators. To visualize the operation of the subjective evaluation in this study, $subjective1_{i,t}$ and $subjective2_{i,t}$ are described by straight lines (colored brown or green) in Figure 1. Each line indicates a point that becomes zero in the bonus assessment. According to these lines, subjective evaluation

TABLE 1: Summary Statistics

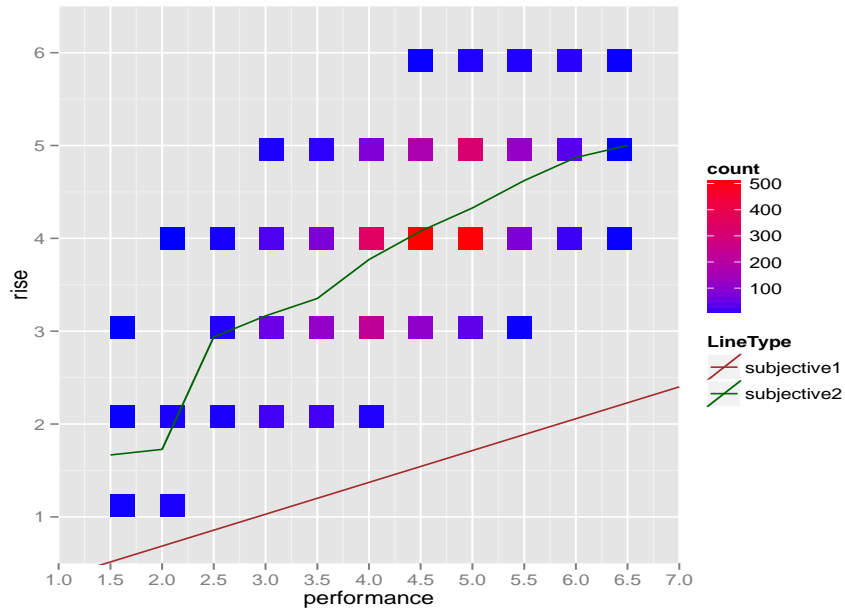|	mean	sd	min	1Q	2Q	3Q	max
$subjective1_{i,t}$	2.49	0.66	0.31	2.29	2.46	2.97	4.46
$subjective2_{i,t}$	0.00	0.62	-1.78	-0.32	-0.08	0.37	2.17
$performance_{i,t}$	4.48	0.69	1.50	4.00	4.50	5.00	6.50
$rise_{i,t}$	4.03	0.76	1.00	4.00	4.00	5.00	6.00
$performance_{i,t-1}$	4.52	0.66	1.50	4.00	4.50	5.00	6.50
$promotion_{i,t-1}$	0.09	0.28	0.00	0.00	0.00	0.00	1.00
$career_{i,t}$	12.97	6.15	2.00	8.00	12.00	17.00	33.00
$grade_{i,t}$	4.36	1.22	3.00	3.00	4.00	5.00	8.00

FIGURE 1: Rise assessment and bonus assessment

in this study will have been operationalized by the vertical distance between the rise evaluation and each line.

5.3. Historical information

Table 2 shows the result of hypothesis 1.1 and 1.2. It shows both the results of ordinary least squares (OLS) and the fixed effects model ($panellinear$). As seen in the table, the coefficient of $performance_{i,t-1}$ is significantly positive in all models. This result supports hypothesis 1.1. The coefficient of $promotion_{i,t-1}$ is also significantly positive in all models and this result supports hypothesis 1.2. In addition, these variables are not only significant, but also have sufficiently large impact on rise assessment ($rise$). The variables of subjective evaluation $promotion_{i,t-1}$ and $subjective2_{i,t}$ is calculated from the rise evaluation. Therefore, the scale of these measures is the same as rise evaluation. In other words, the increase of one unit of $subjective1_{i,t}$ or $subjective2_{i,t}$ means increase of one unit of rise assessment. Thus, the coefficient of $performance_{i,t-1}$ indicates that rise assessment increases by 0.5 on average, if performance in the previous year increases by 1. This interpretation can adapt the variable of promotion in the previous year ($promotion_{i,t-1}$). In other words, if an employee is promoted in the previous year, the rise evaluation of that employee is on average 0.5 higher. These results indicate that even after controlling the fixed effect for each employee, the performance and promotion in the previous year have a sufficient effect on the determination of subjective evaluation.

The coefficients of career ($career$) are -0.007 and -0.008; they were significantly negative. This result is consistent with that of Medoff and Abraham (1980), which states that workers with short careers achieve higher performances after controlling the same grade. However, considering the size of the coefficient, an increase of 1 career will only be about 0.01 in rise assessment. There is little effect of a career to subjective evaluation, at least from this result. Performance in the current year ($performance_{i,t}$) is also significantly negative in all models. Finally, the coefficients of graduates (new) were significantly positive at 0.045 and 0.046, respectively. That is, after controlling the performances in the previous and current years, the subjective evaluation of employees who have not worked in other companies is high on average. This result indicates that such a type of employee has a longer term to accumulate firm-specific skills than a mid-career employee (Prendergast, 1992), or finds it easier to develop a close relationship with the boss, which leads to an increase in subjective evaluation (Bol, 2011).

5.4. Discrimination

Hypothesis 2.1, 2.2, and 2.3 are tested. As described in the previous section, the analyzed sample is only complete samples for all years (from 2004 to 2013). This procedure decreases the sample size from 2980 to 1800.

TABLE 2: The tests of hypothesis 1.1 and 1.2

	OLS	Panel linear	OLS	Panel linear
$performance_{i,t-1}$	0.523***	0.491***	0.517***	0.483***
	(0.018)	(0.020)	(0.018)	(0.020)
$promotion_{i,t-1}$	0.551***	0.548***	0.546***	0.548***
	(0.037)	(0.039)	(0.036)	(0.039)
$career_{i,t}$	−0.007*		−0.008**	
	(0.003)		(0.003)	
$performance_{i,t}$	−0.039*	−0.100***	−0.299***	−0.354***
	(0.017)	(0.019)	(0.017)	(0.019)
$new_{i,t}$	0.045		0.046	
	(0.025)		(0.025)	
$grade$	Yes	Yes	Yes	Yes
$depart$	Yes	Yes	Yes	Yes
job	Yes	Yes	Yes	Yes
$year$	Yes	Yes	Yes	Yes
Fixed Effect	No	Yes	No	Yes
N	2,980	2,980	2,980	2,980
Adj R^2	0.391	0.257	0.336	0.285

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Two-tailed test. Standard errors are in parentheses. In this table, the words of "Yes" in the column of grade, depart, job, and year means that each variable is controlled as a dummy variable. Fixed Effect means the fixed effect of each employees.
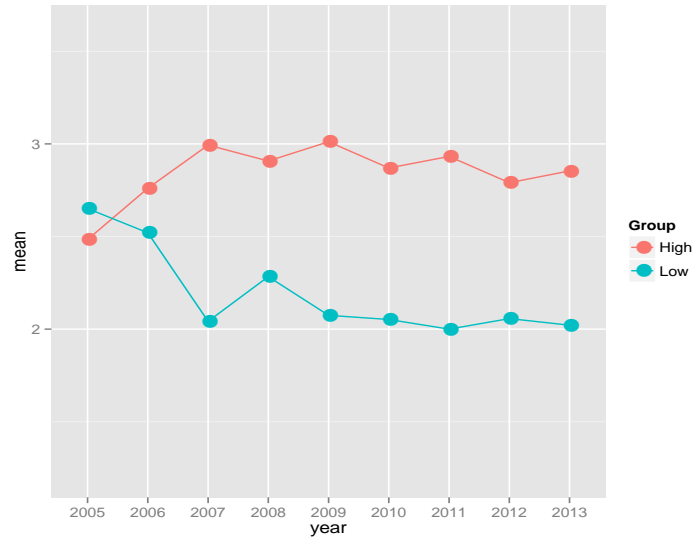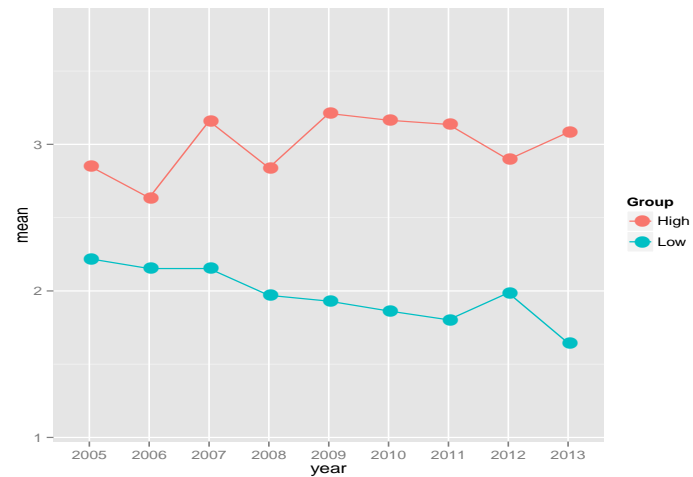
FIGURE 2: The plot of $subjective1$

FIGURE 3: The plot of $subjective1$

In Figure 2, the average and standard deviation of the variable for each evaluation ($performance_{i,t}$, $subjective1_{i,t}$ and $subjective2_{i,t}$) within each groups ($High$ and Low) are plotted for each year. The circle in the figure represents the average value of the evaluation within a group on that year; the length of bar represents the standard deviation of the evaluation within the group on that year. First, the average of objective evaluation ($performance_{i,t}$) in the $High$ group is higher than the Low group in almost all years, except in 2005. The average value reverses in 2005 because the evaluation could not be stable in the nearest year to introduce the HR system. In addition, the upper and lower error bars overlap in all years. This result indicates that the difference in objective evaluation between

high and low groups is small to the extent that it would not be possible to distinguish the evaluation in both groups if the objective evaluation is shifted one standard deviation. On the other hand, Figure 3 shows that the error bar of subjective evaluation within each group does not overlap since 2009. Therefore, even if the subjective evaluation is shifted one standard deviation, the evaluation could be distinguishable since 2009. The difference between objective evaluation (Figure 2) and subjective evaluation (Figure 3) arises from the following two factors of subjective evaluation. The first is that the difference between the average value of the upper and lower group is expanding with time. The second is that the standard deviation of the upper group is decreasing with time. These two trends are consistent with the hypothesis that has been set in this study.

In Table 3, hypotheses 2.1, 2.2, and 2.3 are tested. First, with regard to subjective evaluation, the interaction effects of $High$ and $time$ to $mean$ are significantly positive in the two models (0.093 and 0.105, respectively). This result supports hypothesis 2.1. However, the interaction effects of $High$ and $time$ to $mean$ are also significantly positive in objective evaluation. Possibly, this result of objective assessment due to the average value of the upper and lower groups is reversed, as described above, in 2005. Next, in the regression model with the accuracy of evaluations $1/sd$ as the dependent variable, the interaction terms of $time$ and $High$ are not significant, but are still positive in both models (0.071 and 0.041, respectively). Notably, these values are larger than the coefficient in the objective evaluation model and this partially supports hypothesis 2.2. Finally, in the regression model with the interaction term of $mean$ and $1/sd$ as the dependent variable, the interaction terms of $time$ and $High$ are significantly positive (0.248 and 0.170, respectively). In contrast, in the regression model of objective evaluation, this interaction term is not significant. These results show the differences between subjective and objective evaluation. That is, subjective evaluation can include historical information by evaluators' subjectivity, whereas objective evaluation cannot. In fact, the average and standard deviation of the objective evaluation in each group since 2007 has become constant and it is not affected by time. On the other hand, the subjective evaluation continues incorporating historical information. In Figure 3 and 4, the differences between the average values of the evaluation are updated in both groups for employees and the standard deviation within the high group becomes smaller, even since 2007. These results support the argument that subjective evaluation is better than objective evaluation for long-term incentives.

6. Additional Analysis

6.1. Needs of the additional analysis

The results of the previous chapter show that the subjective evaluation is correlated with historical information (e.g., performance or promotion in previous year) even after controlling the fixed effect of employees. Furthermore, subjective evaluation identifies the upper and lower group over time. However, these results not immune to influential factors such as favoritism (Prendergast and Topel, 1996) and cognitive bias (Tversky and Kahneman,

TABLE 3: The tests of hypothesis 2.1, 2.2, and 2.3

	$mean_{t,j}$			$1/sd_{t,j}$			$mean_{t,j} \times 1/sd_{t,j}$		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$time_{t,j}$	-0.060^{**}	-0.069^{**}	-0.070^{**}	-0.039	-0.031	-0.025	-0.186	-0.182^{*}	-0.165
	(0.019)	(0.020)	(0.020)	(0.039)	(0.029)	(0.028)	(0.110)	(0.077)	(0.091)
$High_{t,j}$	0.563^{**}	0.458^{*}	0.186	-0.386	-0.118	0.167	-0.133	0.501	0.824
	(0.152)	(0.161)	(0.160)	(0.312)	(0.232)	(0.222)	(0.877)	(0.613)	(0.722)
$time_{t,j}$ $\times High_{t,j}$'	0.093^{**}	0.105^{**}	0.094^{**}	0.139^{*}	0.084	0.033	0.551^{**}	0.409^{**}	0.229
	(0.027)	(0.029)	(0.028)	(0.055)	(0.041)	(0.039)	(0.156)	(0.109)	(0.128)
N	18	18	18	18	18	18	18	18	18
Adj R^2	0.930	0.917	0.842	0.353	0.357	0.340	0.758	0.846	0.680

On each line, (1) means subjective1, (2) means subjective2 and (3) means objective. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. A two-tailed test. Standard errors are in parentheses. The subscript of t means the year in which the sample is. The subscript of j means which group the employee belongs to.

1974). For example, the correlation of subjective evaluation with historical information may have been influenced by the evaluator's continued excessive reference to such information.

Previous studies suggest three reasons why evaluators often refer to some measures as a benchmark: a signal of employee ability, the burden to evaluate, and the halo effect.

First, one reason why the evaluator refers to some criteria (e.g., performance in the previous year or career) is that these benchmarks reflect employee ability. Performance in the previous year could be a function of the employee's ability. Moreover, an employee with significant working experience would have high-ability. Many analytical studies such as Gibbons and Waldman (1999) have subscribed to this assumption. The evaluator could refer to these benchmarks, which reflect ability, to improve the estimation accuracy of the employee's ability. This study stands in this aspect.

The second reason is the burden to evaluate. If it does not fit these benchmarks, an evaluatee would complain against the low evaluation because it is not guaranteed by these objective indicators. Therefore, the evaluators lose valuable time in persuading these complaining employees. The evaluator may suffer psychological stress from these

complaints (Bol et al., 2016). To avoid such burden of evaluating, evaluators adjust their evaluation to fit these objective indicators. From this view, the evaluator can use such information to avoid these burdens, although the evaluator can evaluate the employee's ability better.

Finally, the evaluator may overly depend on these benchmarks by cognitive bias. For example, the halo effect or anchoring effect can prompt the evaluator to use these benchmarks excessively (Ittner et al., 2003; Thorsteinson et al., 2008). If these measures reflect the employee's ability and are used properly, these effects (burdensome to evaluate or cognitive biases) do not have a serious effect on the conclusions of this study. However, a problem arises if these measures are excessively (or too little) used. If the subjective evaluation cannot reflect the employee's ability because of the excessive use of these measures, the results in the previous section could not support the assertion in this study. This study asserts that subjective evaluation is better than objective evaluation for long-term incentives, because it can identify high-ability employees. This assertion is based on the assumption that subjective evaluation reflects the employee ability.

In contrast, the results of hypotheses 2.1, 2.2, and 2.3 can be due to favoritism rather than the identification of high-performing employees over time. For example, Prendergast (1996), in a typical theoretical model for favoritism, also assumed uncertainty about the preferences of evaluators. As this study uses the archival analysis research method, it is impossible to identify whether this result is due to favoritism. However, the critical point for this study is whether the influence of favoritism is sufficiently large to impair the correspondence between employee ability and subjective evaluation.

From these discussions, additional analysis is necessary to increase the robustness on the conclusions of this study. The purpose of additional analysis is to test whether subjective evaluation reflects employee ability. In particular, it tests if subjective evaluation is evaluated properly, by verifying if it can predict the future performance of the employee.

6.2. Method

The procedure of additional analysis is follows.

> Step1: Set a base year between 2004 and 2013.
> Step2: Divide all samples of the base year into two groups (former group is from 2004 to the base year; latter group is from the base year to 2013).
> Step3: Calculate the mean of $subjective$ for each employee in the former group.
> Step4: Calculate the mean of $performance$ for each employee in the latter group.
> Step5: Run the regression with the mean in step 3 as a dependent variable, and the mean in step 4 as an independent variable.

In the above procedure, the goal of this additional analysis is to test whether an employee who received a high subjective evaluation in the former period can achieve high performance

in the latter period. The bonus evaluation ($performance$) is adopted, because the variable of the performance is the dependent variable. Therefore, in the regression analysis in Step 5, the coefficient can be interpreted as an impact of subjective evaluation on future average bonus evaluation. Notably, some employees are promoted during the analysis period. For these employees, the standards of subjective evaluation can become severe in the latter period, because they would be evaluated in comparison to employees with an elevated grade after promotion. Therefore, to control the effect of the difference in each of this class, the average grade of the latter of the fiscal year is controlled.

6.3. Result of additional analysis

Table 4 describes the results of the additional analysis. The coefficients of the average subjective evaluation in the former year ($subjective$) are 0.617, 0.707, 0.684, and 0.746, respectively. These coefficients are significantly positive. Regardless of the models with different base year, employees who received high subjective evaluation in the former year can achieve high performance in the latter half of the year, on average. Furthermore, the coefficients of the grade ($grade$) are -0.029, -0.038, -0.035 and -0.029. These coefficients are negative as expected, but not significant.

TABLE 4: The results of additional analysis for $subjective1$

	2008	2009	2010	2011
$subjective_i1$	0.617***	0.707***	0.684***	0.746***
	(0.114)	(0.114)	(0.118)	(0.126)
$grade_i$	−0.029	−0.038	−0.035	−0.029
	(0.032)	(0.032)	(0.033)	(0.037)
N	200	200	200	200
Adj R^2	0.124	0.158	0.139	0.147

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. A two-tailed test. Standard errors are in parentheses. Future performance$_i$ is the mean of performance$_{i,t}$ within each employee from the base year to 2013. Subjective is the mean of subjective1$_{i,t}$ within each employee from 2005 to 2013. Grade is the mean of grade$_{i,t}$ within each employee from the base year to 2013.

TABLE 5: The results of additional analysis for $subjective2$

	2008	2009	2010	2011
$subjective_i2$	0.371**	0.487**	0.516**	0.683***
	(0.138)	(0.146)	(0.156)	(0.167)
$grade_i$	−0.001	−0.008	−0.010	−0.012
	(0.034)	(0.034)	(0.035)	(0.038)
N	200	200	200	200
Adjusted R^2	0.029	0.048	0.046	0.074

*p < 0.05, **p < 0.01, ***p < 0.001. A two-tailed test. Standard errors are in parentheses. Future performance$_i$ is the mean of performance$_{i,t}$ within each employee from the base year to 2013. Subjective is the mean of subjective2$_{i,t}$ within each employee from 2005 to 2013. Grade is the mean of grade$_{i,t}$ within each employee from the base year to 2013.

7. Conclusion

The purpose of this study was to demonstrate the role of subjective evaluation in long-term incentives. Using the data of a Japanese IT Company, this study shows that subjective evaluation helps accumulate historical information to identify with ease high-ability employees. The first result is that the more the performance or promotion in the previous year, the higher the subjective evaluation. The second result is that discrimination of subjective evaluation increases over time. In addition, this study undertakes additional analysis for robustness. The additional analysis shows that higher performance in former years can increase performance in later years, on average. The results support that subjective evaluation in the company is appropriate for predicting at least future outcomes.

This study makes the following contributions. First, it contributes to a series of studies on subjective evaluation. Many studies on subjective evaluation often focus on its negative aspects, such as favoritism (Prendergast and Topel, 1994), leniency (or centrality), bias (Moers, 2005; Bol, 2011), halo effect (Thorsteinson et al., 2008) or a tendency to focus on only a particular measure (Lipe and Salterio, 2000; Ittner et al., 2003; Woods, 2012). Few studies empirically show the positive points of subjective evaluation. However, although subjective evaluation is widely adapted, there is little empirical evidence on why subjective evaluation is so popular in reality. This study also contributes to a series of studies on the relationship between long-term incentives and performance evaluation.

Previous studies do not give empirical evidence for why companies use subjective evaluation for long-term incentives, even when employees do not change jobs. This study provides proves that subjective evaluation is better than objective performances for long-term incentives. For long-term incentives, it is important to identify high-ability employees, because long-term incentives are paid over multiple periods and the cost to promote a low ability employee is higher compared with short-term incentives. The results of this study empirically show that subjective evaluation has higher capacity to identify high-ability employees than objective evaluation.

However, the study also has some limitations. First, if a company wants to incorporate historical information to evaluate long-term incentives, they can use the aggregation of some objective measures, but this study cannot explain why companies do not adapt such an aggregation of objective measures. In addition, the results neglect some effects such as motivation or growth of employees. Finally, survival bias can affect the result of this study, although this effect is partially considered in the interpretation of the result. This survival bias is caused by using only samples that have complete dataset over all periods (from 2004 to 2013). Therefore, this study disregards samples of fired employees or employee turnover. Such an effect of survival bias may give a result that standard deviation in the low group is not decreasing.

References

Abernethy, M. A., Bouwens J., and Van Lent L. (2004). Determinants of Control System Design in Divisionalized Firms. *The Accounting Review* 79, 545–570.

Anthony, R. N., and Govindarajan. V. (2007). *Management Control Systems*, 12th edition (McGraw-Hill).

Baik, B., Evans J. H., Kim K., and Yanadori Y. (2016). White-Collar Incentives. *Accounting, Organizations and Society* 53, 34–49.

Baker, G. P., Gibbs M., and Holmström B. (1994a). The Internal Economics of the Firm-Evidence from Personnel Data. *The Quarterly Journal of Economics* 95, 881–919.

Baker, G. P., Gibbons R., and Murphy K. J. (1994b). Subjective Performance Measures in Optimal Incentive Contracts. *The Quarterly Journal of Economics* 109, 1125–1156.

Banker, R. D. (2000). An Empirical Investigation of an Incentive Plan that Includes Non-financial Performance Measures. *The Accounting Review* 75, 65–92.

Banker, R. D., and Datar S. M. (1989). Sensitivity, Precision, and Linear Aggregation of Signals for Performance Evaluation. *Journal of Accounting Research* 27, 21–39.

Banker, R. D., Lee S. Y., and Potter G. (1996). A Field Study of the Impact of a Performance-Based Incentive Plan. *Journal of Accounting and Economics* 21, 195–226.

Becker, S. (1962). Investment in Human Capital: A Theoretical Analysis. *The Journal of Political Economy* 70, 9–49.

Bertrand, M., and Schoar, A. (2003). Managing with Style: The Effect of Managers on Firm Policies. *The Quarterly Journal of Economics* 118, 1069–1208.

Bol, J., Kramer S., and Maas V.S. (2016). How Control System Design Affects Performance Evaluation Compression: The Role of Information Accuracy and Outcome Transparency. *Accounting Organizations and Society* 51, 64–73.

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 Managers' Performance Evaluation Biases. *The Accounting Review* 86, 1549–1575.

Campbell, D. (2008). Non-financial Performance Measures and Promotion-Based Incentives. *Journal of Accounting Research* 46, 297–332.

Campbell, D, Epstein M. J., and Martinez-Jerez, F.A. (2011). The Learning Effects of Monitoring. *Accounting Review* 86, 1909–1934.

Chow, C. W. (1983). The Effects of Job Standard Tightness and Compensation Scheme on Performance: An Exploration of Linkages. *The Accounting Review* 58, 667–685.

Cichello, M. S., Fee C. E., Hadlock C. J., and Sonti R. (2009). Promotions, Turnover, and Performance Evaluation: Evidence from the Careers of Division Managers. *The Accounting Review* 84, 1119–1143.

DeVaro, J., and Waldman M. (2012). The Signaling Role of Promotions: Further Theory and Empirical Evidence. *Journal of Labor Economics* 20, 91–147.

Dohmen, T. J. (2004). Performance, Seniority, and Wages: Formal Salary Systems and Individual Earnings Profiles. *Labour Economics* 11, 741–763.

Ederhof, M. (2011). Incentive Compensation and Promotion-Based Incentives of Mid-Level Managers: Evidence from a Multinational Corporation. *The Accounting Review* 86, 131–153.

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

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

Gibbons, R., and Murphy K. J. (1992). Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence. *Journal of Political Economy* 100, 468–505.

Gibbons, R., and Waldman M. (1999). A Theory of Wage and Promotion Dynamics Inside Firms. *The Quarterly Journal of Economics* 114, 1321–1358.

Gibbons, R., and Waldman M. (2006). Enriching a Theory of Wage and Promotion Dynamics inside Firms. *Journal of Labor Economics* 24, 59–107.

Gibbs, M., Merchant K. A., Van der Stede W. A., and Vargus M. E. (2004). Determinants and Effects of Subjectivity in Incentives. *The Accounting Review* 79, 409–436.

Grabner, I. (2014). Incentive System Design in Creativity-Dependent Firms. *The Accounting Review* 89, 1729–1750.

Grabner, I., and Moers F. (2013). Managers' Choices of Performance Measures in Promotion Decisions: An Analysis of Alternative Job Assignments. *Journal of Accounting Research* 51, 1187–1220.

Hayes, R. M., and Schaefer S. (2000). Implicit Contracts and the Explanatory Power of Top Executive Compensation for Future Performance. *The RAND Journal of Economics* 31, 273–293.

Holmström, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics* 10, 74–91.

Holmström, B. (1982). Moral Hazard in Teams. *The Bell Journal of Economics* 11, 74–91.

Holmström, B. (1999). Managerial Incentive Problems—A Dynamic Perspective. *Review of Economic Studies* 66, 169–182.

Holmström, B., and Milgrom P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization* 6, 24–52.

Höppe, F., and Moers F. (2011). The Choice of Different Types of Subjectivity in CEO Annual Bonus Contracts. *The Accounting Review* 86, 2023–2046.

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

Jensen, M. C., and Meckling W. H. (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics* 3, 305–360.

Kahneman, D., Slovic P., and Tversky A. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 1124–1131.

Lambert, R. A. (2001). Contracting Theory and Accounting. *Journal of Accounting and Economics* 32, 3–87.

Lazear, E. P. (2004). The Peter Principle: A Theory of Decline. *Journal of Political Economy* 112, S141–S163.

Lazear, E. P. (2009). Firm-Specific Human Capital: A Skill-Weights Approach. *Journal of Political Economy* 117, 914–940.

Lazear, E. P., and Oyer P. (2013). Personnel Economics. In Gibbons R. and Milgrom P. (Eds.). *Handbook of Organizational Economics* 479–519.

Lazear, E. P., and Rosen S. (1981). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy* 89, 841–864.

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

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

Medoff, J. L., and Abraham K. G. (1980). Experience, Performance, and Earnings. *The Quarterly Journal of Economics* 95, 703–736.

Merchant, K. A., and Manzoni J. (1989). The Achievability of Budget Targets in Profit Centers a Field Study. *The Accounting Review* 64, 539–558.

Milgrom, P., and Roberts J. (1990). *Economics, Organization, and Management* (Prentice-Hall, Inc.).

Moers, F. (2005). Discretion and Bias in Performance Evaluation: The Impact of Diversity and Subjectivity. *Accounting, Organizations and Society* 30, 67–80.

Moers, F. (2006). Performance Measure Properties and Delegation. *The Accounting Review* 81, 897–924.

Moers, F. (2007). Doing Archival Research in Management Accounting. In C. S. Chapman, Hopwood A. G. and Shields M. D. (Eds.). *Handbook of Management Accounting Research* 1, 399–413.

Nagar, V. (2002). Delegation and Incentive Compensation. *The Accounting Review* 77, 379–395.

Prendergast, C. (1992). Career Development and Specific Human Capital Collection. *Journal of the Japanese and International Economies* 6, 207–227.

Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature* 37, 7–63.

Prendergast, C., and Topel R. H. (1993). Discretion and Bias in Performance Evaluation. *European Economic Review* 37, 355–365.

Prendergast, C., and Topel R. H. (1996). Favoritism in Organizations. *Journal of Political Economy* 104, 958–978.

Saal, F. E., Downey R. G., and Lahey M. A. (1980). Rating the Ratings: Assessing the Psychometric Quality of Rating Data. *Psychological Bulletin* 88, 413–428.

Sprinkle, G. B., and Williamson M. G. (2007). Experimental Research in Management Accounting, In Chapman C. S., Hopwood A. G. and Shields M. D. (Eds.). *Handbook of Management Accounting Research* 1, 415–444.

Thorsteinson, T. J., Breier J., Atwell A., Hamilton C., and Privette M. (2008). Anchoring Effects on Performance Judgments. *Organizational Behavior and Human Decision Processes* 107, 29–40.

Woods, A. (2012). Subjective Adjustments to Objective Performance Measures: The Influence of Prior Performance. *Accounting Organizations and Society* 37, 403–425.