Relationships between Objective and Subjective Performance Ratings

Kenneth A. Merchant*  
University of Southern California

Carolyn Stringer  
University of Otago

Paul Theivananthampillai  
University of Otago

Preliminary first draft.  
October 2009

* Corresponding author: 3660 Trousdale Pkwy, Los Angeles, CA 90089-0441. E-mail: kmerchant@marshall.usc.edu.
Relationships between Objective and Subjective Performance Measures

ABSTRACT: This study explored relationships between objective and subjective performance measures in a company setting where both were weighted equally in importance for allocating bonuses to high- and mid-level managers and professionals. We found that the correlation between the objective and subjective performance ratings was positive but small, which suggests no performance evaluation halo effect. Contrary to most prior studies, we found no evidence of the subjective rating centrality bias; the subjective performance ratings were more highly differentiated than were the objective ratings. We found no evidence of a leniency bias in the subjective performance ratings, apparently because the company’s mandated maximum average subjective performance rating was rigorously enforced. However, we did find evidence of leniency in the objective measure portion of the system, and that leniency was particularly high in a time of economic stress. Finally, we found that the objective performance ratings were higher both for managers, as compared to professionals, and for employees at higher ranks, but they were lower for employees working at corporate level rather than in one of the SBUs. We discuss the possible causes and implications of these findings.

Keywords: performance measurement, performance evaluations, objective measures, subjective measures, halo effect, centrality bias, leniency bias, incentive compensation.

Data availability: Confidentiality agreements prevent the authors from distributing the data.

I. INTRODUCTION

While objective and subjective measures can be either substitutes for each other (e.g., Kren and Tyson 2009), both analytical and empirical research has suggested that in most cases the two forms of measurement should be used as complements (Rajan and Reichelstein 2009, 2006; Zhao and Yu 2007; Van der Stede et al. 2006; Moers 2005; Ittner et al., 2003; Baiman & Rajan 1995; Baker et al. 1994). Each type of measure has its advantages and disadvantages. For the important purposes of evaluating performance and allocating performance-dependent compensation, the combined use of both objective and subjective measures can often exploit the advantages of the each while minimizing the disadvantages.
Much is yet to be learned, however, about how these two types of measures work together in a performance measurement system (Bol 2008). For example, do the two types of measures reinforce each other or do they reflect different aspects of performance? Are the subjective rating biases of compression and leniency commonly found in experimental settings evident in actual company settings, and are they mitigated or exacerbated by the objective performance measures? Do the objective and subjective evaluations and the relationships between them vary with the situation, such as the employee’s job or rank and/or the economic conditions being faced?

Subjectivity, in particular, has not received much attention in the incentives literature because of the difficulty in gaining access to real world data that managers generally consider to be highly sensitive and confidential (Bol 2008; Prendergast 1999; Indjejikian 1999; Harris 1994). As a consequence, much of the early evidence about subjective performance evaluations came from experimental studies. Our study, which explores answers to the questions posed above, was made possible when we gained access to one large Australasian company’s archival database that included both objective and subjective annual performance ratings for the approximately 700 highest ranked managers and professionals over a two-year period. In this company, the objective and subjective performance ratings were each weighted 50% in importance for the purpose of allocating bonus awards.

Beyond the equal use of objective and subjective performance ratings in the allocation of bonus monies, our research setting had two unique features that were useful for research purposes. First, the two years of data that we acquired spanned a crisis period, which company managers called a “performance shock”—the first operating loss in the company’s
history. As a consequence, we were able to study the extent to which managers adjusted both the objective and subjective elements of their bonus system in response to the crisis.

Second, at our site, the subjective performance ratings were constrained to a maximum mean at the entity (corporate and SBU) level. The constraint, which is used by only a small minority of firms (Bol 2008), was applied to try to ensure consistency and fairness across the business units or, in other words, to minimize or avoid the leniency bias that has been demonstrated in prior research (e.g., Moers 2005; Murphy, 1992). We were able to study whether the subjective performance rating constraint was rigorously and consistently enforced, and hence effective, even in the period with difficult economic conditions. And we were able to explore whether leniency was exercised in the objective portion of the performance evaluations in a situation where leniency was discouraged in the subjective portion of the incentive plan.

We found that the correlation between the objective and subjective performance ratings was quite small, which suggests the two forms of ratings are assessing substantially different aspects of performance. In other words, we observed little or no performance evaluation halo effect. Contrary to most prior studies, we found no evidence of compression in the subjective performance evaluations. To the contrary, the subjective performance ratings were more highly differentiated than were the objective ratings. We found no subjective performance rating leniency bias, which suggests that the company-imposed maximum for the subjective performance ratings was rigorously enforced. However, we did find that leniency was exercised in the objective portion of the system. We found that managers adjusted for the effects of negative uncontrollable factors by lowering the performance targets after they were set or by making rating adjustments *ex post*. We found
that the leniency allowed in the objective performance ratings was significantly higher in
times of economic stress than in normal operating conditions. And we found that the
subjective performance ratings, but not the objective ratings, were systematically higher for
managers, as compared to professionals, and for both types of employees at higher ranks.

To supplement the analyses of archival data, we conducted a number of interviews,
both formal and informal, with employees of the firm. The formal interviews were recorded
and transcribed. We also observed some management meetings, including those involving a
review of the company’s incentive compensation plan. The on-site visits helped us both in the
understanding of the details of the system implementation and the interpretation of the
findings.

II. PRIOR LITERATURE

When quantitative performance measures are effective, firms are likely to assign a
high proportion of bonuses based on objective formulas (Gibbs et al. 2004; Hayes and
Schaefer 2000; Bushman et al. 1996). In most managerial situations, however, the
quantitative performance measures are not perfect. They often distort incentives because, for
example, they are inappropriate, incomplete or prone to manipulation and/or because they
impose undue risk on the employee (i.e., they are affected by uncontrollable factors).

Evaluating performance subjectively is one possible solution to the performance
measurement problems (e.g., Bol 2008; Nisar 2007; Rajan and Reichelestein 2006; Van der
Stede et al. 2006; Moers 2005; Gibbs et al. 2004; Baker et al. 1988). Evidence suggests that
subjective performance evaluations are actually even more common than those involving
objective measurement (Bol 2008). This fact suggests both that the performance
measurement problems described above are pervasive and that subjective evaluations are a viable solution to those problems.

Subjective performance evaluations are based on personal impressions and judgments. They are difficult to anticipate *ex ante*, and they are nonverifiable *ex post*, so they are said to be non-contractible. Subjective performance evaluations and the rewards linked to them are sometimes referred to as *implicit contracts*. But even so, subjective judgments can be used as complete substitutes for objective formulas (e.g., Kren and Tyson 2009; Gibbs et al. 2004). More often, however, subjectivity is used to modify or supplement the objective formula results.

Subjectivity can be used in any of a number of ways, and it can provide multiple benefits. Subjectivity can be used to compensate for measures that, because they are inappropriate or incomplete, might cause agents to take actions that will not increase long-term firm value. It can be used to take into consideration other hard-to-quantify information to make the evaluation conclusions more complete and accurate, and it can be used to filter out the effects of some uncontrollable factors. When this is done, the firm benefits from improved incentive alignment and lower compensation cost because the employees are subjected to reduced risk. And subjectivity can be used to adjust for the obsolescence of preset performance standards based on flawed planning assumptions. Evaluators can exploit new, relevant information that becomes apparent during the measurement period. This is sometimes referred to as an *ex post* “settling up.” In effect, subjectivity allows for the recalibration of incentives during or after the period, which can be especially important if there are costs to changing or renegotiating formal, formula-based bonus contracts. But
allowing performance evaluation subjectivity has its own sources of errors and biases, as is discussed below.

**Differentiation between objective and subjective performance ratings**

If both objective and subjective performance evaluations are used, how are they related? One possibility is that they are highly correlated. The subjective evaluators might be highly influenced by the objective performance ratings, causing them not to differentiate individuals’ performances on any other basis. This is the so-called *halo effect* that has been observed in some, particularly experimental, research settings (e.g., Jennings et al. 2004; Balzer and Sulsky 1992; Fox et al. 1983). Bol and Smith (2009) found that when objective measures were deemed to be highly controllable, subjective evaluations were highly correlated with the objective measures. Ittner et al. (2003) found that supervisors place greater weight on objective/quantitative measures than they do on subjective/qualitative measures.

At low levels of the firm, objective measures can sometimes provide good indicators of an individual’s performance. Number of good units assembled is probably a good indicator of the performance of an employee on an assembly line. At higher levels of the firm, however, the objective measures are usually not as effective. For managers, for example, some commonly used objective measures reflect the performance of the entire organizational entity, not the manager’s individual performance, the factors over which they have substantial influence. Subjective ratings can be more focused on individual performance, can consider a broader range of factors, some of which are difficult or impossible to quantify, and can usually filter out some aspects of uncontrollability (Bol and Smith 2009).
Particularly in higher organization-level situations, then, we would expect the two objective and subjective performance measures to be significantly different from each other because they are based on different information sets. Indeed that is what much prior research has found. Heneman (1986) analyzed the results of 23 independent studies containing both objective and subjective ratings of employee performance. Even after adjusting for the effects of unreliability in the measures, he found low, positive correlations (.27) between subjective performance ratings and the objective measures (e.g., sales volume, complaint letters, output) that were deemed relevant in those situations. A later, similar meta-analysis of such studies found in a mean correlation between objective and subjective evaluations of .39 (Bommer et al. 1995). While these correlations are significantly greater than zero, they indicate that objective and subjective performance measures should not be used interchangeably. Bommer et al. did not find evidence in any subgroup that suggested convergent validity. They concluded that there is no unidimensional performance factor. Objective and subjective measures each give a different perspective on performance.

Our data reflect evaluations of employees at relatively high organization levels where objective performance measures are known to be generally imperfect, so we believe that our findings will be consistent with those of the studies reviewed by Heneman and Bommer et al. That is the expectation expressed in our first hypothesis.

**H1: Objective and subjective performance evaluation elements are significantly different from each other.**

**Subjective evaluation biases**

Allowing subjectivity in performance evaluations has some significant potential advantages, but it also has some risks and costs. Subjectivity has the potential of working
well only if the evaluators make fair, informed, unbiased judgments. But it has been
documented that subjective performance evaluations are laden with errors. Viswesvaran et
al.’s (1996) meta-analytic study found a mean interrater reliability of only 0.52 for
performance ratings.

There are many potential causes of subjective performance rating errors (Bol 2008,
2009). Evaluators can be seen as agents with own utility functions that can deviate from the
principal’s objectives. These utility function deviations can be the source of any of a number
of problems. If supervisors are not rewarded for accurate ratings, or punished for error-laden
or biased ones, they may have insufficient motivation to invest time in gathering information
(Bol 2009; Fox et al. 1983). But direct rewards or punishments are difficult to administer
because one inherent feature of subjective performance evaluations is that their results cannot
be verified by third parties. Also contributing to errors and bias in subjective performance
evaluations are cognitive limitations that cause evaluators to tend to focus on only a limited
set of performance dimensions (Ittner et al. 2003). Some evaluators arbitrarily favor certain
employees, perhaps to encourage loyalty or otherwise serve their self-interest (Ferris & Judge
1991). Raters have a natural self-serving bias, a tendency to inflate their subordinates’ ratings
so that they themselves appear to be successful (Greenberg 1991). And most evaluators have
a preference for a pleasant relationship with their subordinates, which is sometimes referred
to as a desire to minimize confrontation costs (Bol 2009; Varma et al. 1996). This preference
can lead to a centrality bias, a leniency bias, or just ambiguity, a lack of feedback.

The centrality bias (also sometimes referred to as a compression bias) occurs where
evaluators differentiate only slightly between employees (e.g. Bol 2009; Grund and Przemeck
2008; Bretz et al. 1992; Murphy & Cleveland 1991; Landy and Farr 1980). Lack of
discrimination makes it easy for the evaluators to tell their subordinates that their performance was rated similarly to those of other employees. The motivation for the centrality bias stems from an inherent inequality aversion that is said to be an important force driving human behavior in many situations (Grund and Prezemeck 2008; Bolton and Ockenfels 2000). High differentiation in evaluations can have adverse consequences for management employee relations and workgroup cohesion. Thus, as one review article (Rynes et al., 2005, p. 583) concluded, “Pay is rarely seen by employees as strongly differentiated across employees when it is tied to subjective measures of performance.” But the centrality bias has adverse organizational consequences because, among other things, it discourages high performers and causes lack of attention to those aspects of performance rated subjectively (Ahn et al. 2008).

The leniency bias, which is also common, occurs where a large majority of employees get “above average” ratings (e.g. Bol 2009; Levy and Williams 2004; Jawahar & Williams 1997; Bretz et al. 1992). This bias stems from supervisor averse to offering poor ratings to subordinates, especially where these ratings are used to determine pay, both because bad evaluations may be demotivating rather than motivating and because it is uncomfortable to deliver bad news.

Prior studies have found evidence of both the centrality and leniency biases. In the subjective evaluation system used at Merck & Co., Murphy (1992) found that more than 70 percent of Merck employees were rated in only three of 13 performance categories, and only about five percent were given ratings below the mid-point of the rating scale. Bretz et al. (1992) found that it was common for 60-70% of firms’ employees to be rated in the top two
Objective performance criteria are often thought to be less susceptible to rater biases such as centrality and leniency (Moers 2005; Rynes, et al. 2005). Studies have shown that usually pay is seen as more differentiated—in other words there is less centrality bias—when it is based on objective performance evaluations rather than subjective ones (Ahn et al. 2008; Moers 2005; Rynes et al. 2005). This is the expectation we express in hypothesis two.

**H2: Subjective evaluations have less dispersion than do objective evaluations.**

Would we expect leniency bias at our research site which, as described above, prescribes a SBU-level rating maximum for its subjective ratings? The prescribed mean is specifically designed to limit the leniency bias. If the prescribed mean is enforced, we would expect that the subjective ratings would exhibit no leniency bias. This is the expectation we lay out as hypothesis 3.

**H3: When the mean is prescribed for subjective performance ratings, there is no subjective performance rating leniency bias.**

In most situations, though, both the evaluators and the employees being evaluated have a positive utility for leniency. If H3 is supported, and leniency cannot be exercised in the subjective ratings, the evaluators could only provide leniency through the objective portion of the bonus plan. In a system with a strict objective formula-based allocation of incentive awards, leniency can only be exercised by making performance targets easier to achieve (Merchant 1989; Merchant and Manzoni 1989). In a less strict objective system, sometimes subjectivity can be used *ex post* to make variance analysis-type adjustments to performance evaluations or simply to make bonus grants that are not closely linked to the performance
Department of Accountancy and Business Law, Working paper series number 17, July 2010 evaluations (Merchant 1989). Using the subjective performance ratings as a standard against which to evaluate the degree of leniency in the objective performance ratings, we propose hypothesis four for test.

**H4: When the mean is prescribed for subjective performance ratings, the mean objective performance rating will be higher than the mean subjective rating.**

**Effects of Deteriorating Economic Conditions**

What happens to performance evaluations when economic conditions deteriorate? In such conditions, incentive plans are often stressed. When results deteriorate and payouts decline, perhaps for reasons that may be totally beyond the control of the employees affected, feelings of injustice can be intense. Many incentive plans are abandoned in such times because of employee pressure (Petty et al. 1992).

If the incentive plans are continued, the difficult economic times increase the demand for leniency in evaluations, both to ease the compensation burden on the employees and to minimize employee turnover. But objective financial performance targets have been shown to be generally more challenging to achieve in difficult times (Indjejikian et al. 2009; Merchant 1989; Merchant and Manzonic 1989). To meet the demand for leniency, particularly in the face of more challenging financial targets, most firms use some form of evaluation and/or reward allocation subjectivity to allow the payout of some bonuses. For example, Matejka et al. (2009) found that loss-making firms made significantly higher use of nonfinancial performance measures and tended to evaluate performance more in a subjective manner. Indjejikian et al. (2009) found that in the recession year of 2009 the reported probabilities of achieving earnings targets were significantly lower than the chances of achieving
nonfinancial performance targets, presumably because many of the latter were evaluated subjectively. Managers can also allow subjective overrides of the objective formula linking performance with rewards, such as by changing the performance targets ex post, or they can grant special, discretionary bonuses to personnel who are believed to especially deserving (Merchant 1989).

At our research site, however, the mean subjective ratings and, hence, the bonuses paid based on those evaluators are fixed, evaluators have to exercise their leniency through the objective performance ratings. Thus we predict that in difficult economic situations when subjectivity cannot be used for leniency purposes, performance target difficulty should be lower in times of difficult economic conditions. The leniency in the objective portion of the incentive plan should be even more pronounced than in the good economic times. These expectations are expressed in hypotheses five and six.

H5: When a mean is prescribed for the subjective performance ratings, financial target difficulty will decrease in times of financial stress.

H6: Objective performance rating leniency will be greater in difficult economic conditions than in normal operating conditions.

III. RESEARCH DESIGN

Site

The company from which we acquired data, which we will call Sensol, is a privately-owned Australasian corporation with revenues of approximately US$1 billion and 9,000 employees. Sensol’s large core business used standardized technologies and had high market share. However, the core business was in a declining industry due to the effects of technological substitution. It had become the cash cow used to fund Sensol’s diversification
strategy into new business areas (e.g., high-end technologies). The volume declines in Sensol’s core business had a large impact on the company’s profitability because a high percentage of its costs were fixed.

**Sensol’s Incentive Program**

Sensol began offering bonuses in the late 1980s. Their plan at that time created a bonus pool based on corporate performance. The bonus pool was allocated to managers totally subjectively. Over time, the plan became more formalized.

The incentive plan we studied had been used since the mid 1990s. A total of approximately 700 employees, those in the top 12 management grades and top eight professional grades, were offered the opportunity to earn annual performance-dependent bonuses. The bonus payouts were based half on an objective rating determined by a set formula and half on a subjective performance rating. Both the objective and subjective performance ratings were recorded on a 200-point rating scale.

The objective evaluations were based on the measures in included in a “value scorecard,” which was based loosely on the balanced scorecard concept. The key financial measure, EVA®, a stylized type of residual income measure developed by the consulting firm Stern Stewart & Co., was weighted 80% in importance. For employees working in a strategic business unit (SBU), the 80% figure was comprised of the corporate EVA measure, weighted 70% in importance, and the business unit EVA with a 30% weighting. The other 20% of the objective evaluation was placed on other quantified measures, including service performance (5%), external revenue (5%), and social measures (10%), which included employee

---

1 For employees at working at corporate level, only corporate EVA was considered.
satisfaction and indicators of health and safety. The set of objective measures and their importance weightings were identical for all employees included in the incentive program.

While the objective portion of the incentive program was intended to be strictly formula-based, Sensol allowed the possibility of some flexibility in adjusting the targets set during the annual planning process. These adjustments were made to adjust for the effects of factors or events appearing after the targets were set that were deemed to be outside the managers’ control. The adjustments had to be approved by the senior management team and the Board of Directors. As one manager explained:

The board is fairly accommodating [but] they will not go for small changes. They are willing to make adjustments only for those types of external influences that can put your target way out [of reach].

The subjective portion of the evaluations was based on performance in a short list (typically five or six) of individual performance areas that were linked to the value scorecard (e.g., percent cost reductions, success in meeting project milestones) and other, softer factors (e.g., teamwork, accountability). Superiors had discretion over both the choice of measures and the weighting of importance of the measures. To ensure consistent ratings, company policy required that the subjective evaluations average no more than 120 on the 200-point evaluation scale in each reporting entity (i.e., SBU or corporation). Performance ratings and bonus payouts were based on annual performance.

The target bonus payouts varied both with job and grade level. Each plan participant was given a “base bonus,” an amount to be paid if both objective and subjective performance ratings were exactly equal to 100. The approximate average base bonuses for senior managers, middle managers, and professionals were, respectively, 15%, 10%, and 6%. In
normal times, the actual average bonus payouts for the three employee groups were approximately 20%, 10% and 6%, respectively.

Performance thresholds and maximums were also set. The threshold was the level of performance below which no bonuses would be paid. The maximum bonus payouts were four times the base bonus level. Thus, for example, a manager with a base bonus of $10,000 who received a scorecard rating (objective measure) of 200% (maximum), and an individual subjective rating of 200% (“outstanding”) would earn a cash bonus of $40,000. The function linking performance with cash payouts was linear between the threshold and the target and between the target and the maximum.

The rough rule of thumb was that the threshold was at 80% of budget, and the maximum was at 120% of budget. But the actual thresholds and maximums deviated from that rule of thumb because the risks and opportunities varied significantly across SBUs. So in practice, the threshold levels of performance were described to us as being set at a level of performance deemed to be “not worth paying bonuses.” We were not provided access to all the target details because they were considered highly sensitive, but we were told that in some SBUs operating in uncertain environments the threshold was set at 80% of budget, while in some stable SBUs, the threshold was set at 95% of budget. Similarly, judgment was involved in the setting of the performance maximums. The intent was to set the maximums at a performance level involving “considerable stretch.” The performance targets, thresholds and maximums were reviewed by the entire senior management team. As one senior manager explained:

We must come to a consensus at the top level. There is a healthy debate and discussion before we agree for each different business on their stretch and target ranges.
Sensol also offered a long-term incentive plan for its top 30 managers. But in the period of our study, the long-term performance targets were generally seen as unachievable, so it seemed to have little or no behavioral effects. Sensol offered no equity-based compensation.

**The “Performance Shock”**

At the start of the period for which we acquired data, Sensol’s fiscal year ended March 31. Our first annual set of data is for the fiscal year ended March 31, 2000. But then Sensol managers decided to change the company’s fiscal year to end on June 30 to be consistent with the financial reporting period used by most firms in Sensol’s industry. The quarter April 1, 2000 to June 30, 2000 was an “orphan” quarter, not included in any fiscal year. Our second annual set of data is for the following fiscal year, ended June 30, 2001. Throughout the remainder of the paper, we will refer to FY 2000 as Year 1 of our study and FY 2001 as Year 2.

Coincidentally, the orphan quarter turned out to be what Sensol managers called a “shock loss” quarter, the first time in its history that the company had experienced a loss on a consolidated basis. No bonuses were paid during this quarter. The company’s prospects were seen to have become seriously harmed. Our interviews revealed that during this period “doom and gloom” abounded, as employees worried about the impact of technical substitution for the core product. There was considerable speculation that the organization would never be able to pay a bonus again. Because of this shock loss quarter, our data set turned out to contain “before and after” scenarios. The first year contained performance ratings in

---

2 Sensol managers considered these data to be extremely sensitive. We were able to obtain the data only after passage of a considerable amount of time.
Department of Accountancy and Business Law, Working paper series number 17, July 2010

economic conditions that were seen as good; the conditions in the second year were seen as much more challenging.

During the shock loss quarter, budgeted targets were set for Year 2. But soon after they were approved, these targets were considered to be unachievable in the harsher environment then seen to exist. The Year 2 targets were reduced across the business units.

Data

Our database include objective and subjective performance ratings and bonus payout information for each of the two fiscal year periods for the top 12 grades of managers (below the CEO level) and the top 8 grades of professionals (specialists who do not manage subordinates) within Sensol. These groups total 720 employees for Year 1 and 743 for Year 2. We do not have individual record identifiers that would allow us to match the individual employee records across the two measurement periods.

Measures

The core variables used in the empirical analyses include the following:

SUBJECTIVE RATING. A score between 0 and 200 assigned by the employee’s immediate superior.

OBJECTIVE RATING. A score between 0 and 200 based on that unit’s value scorecard performance as compared to preset targets.

TARGET DIFFICULTY. Calculated as the difference between the maximum target minus the threshold target divided by the actual performance target.

Also included in some analyses are a number of control variables:

EMPLOYEE LOCATION. A dummy variable was used to distinguish between employees working at the corporate level of the organization and those working in a strategic business unit (0=corporate; 1=SBU).
STRATEGY. A dummy variable was used to distinguish between Sensol’s core business, which was declining, and the newer businesses in which it was investing (coded as 0 = decline; 1 = maintain).

EMPLOYEE TYPE. Dummy variables were used to distinguish managers vs. professionals and the ranks of each (high vs. low).

YEAR. A dummy variable was used to distinguish between the two years of data (Year 1 = 0; Year 2 = 1).

IV. RESULTS

4.1 Descriptive Statistics

Table 1 shows descriptive statistics for the objective and subjective performance ratings and target difficulty variables in each of the performance measurement periods. Table 2 shows statistics descriptive of the employee types and employee locations.

----- Insert Tables 1 and 2 here -----

4.2 Differentiation of objective and subjective ratings

Hypothesis 1 stated the expectation that the objective and subjective performance ratings would be significantly different from each other. The correlation between the objective and subjective performance ratings is .16. This coefficient is somewhat lower than those found in the surveys by Heneman (1986) and Bommer et al. (2005).

While the .16 correlation is significantly larger than zero, showing some positive relationship between the objective and subjective ratings, it provides support for H1 because .16 is significantly less than a perfect correlation. It is obvious that the objective and subjective dimensions of Sensol’s performance evaluations provide different perspectives.
The objective measures are not dominating the subjective ratings as they might if there was a strong evaluation halo effect.

As a sensitivity test, we calculated the objective/subjective rating correlations for the different categories of employees. We found no significant difference in the correlations for employees working at corporate level vs. those working in an SBU. But we did find the objective/subjective rating correlation for the group of high ranking managers (.07) to be significantly lower than for other employee groups (t = -3.11; p<.001). The correlations for the other groups were as follows: middle managers (.19), high-level professionals (.16), and low-level professionals (.18).

4.2 Centrality Bias

Hypothesis 2 set up the test for the centrality bias. As is shown in Table 1, we found that, contrary to H2, the range and standard deviations for the subjective ratings were both higher than those for the objective ratings in both normal Year 1 and post-performance-shock Year 2. That is, the subjective ratings were less compressed than the objective ratings. This may be because in the setting we studied the level of analysis for the objective measures is at the entity (corporate or SBU) level, while it is at the individual level for the subjective ratings. But the same is true in other studies that have found evidence of the centrality bias (e.g., Moers 2005). In any case, our finding show that there is more performance differentiation in the subjective ratings than in the objective ratings. We conclude that there is no evidence of compression, or the centrality bias, in the subjective performance ratings.
4.3 Leniency bias

Hypotheses 3 and 4 set up the tests for the leniency effect. Hypothesis 3 stated the expectation that because of the forced mean for the subjective ratings, no leniency would be found in the subjective ratings. The results support this hypothesis. As is shown in Table 1, the mean subjective rating is 111 in year 1 and 115 in year two, which means that those rating performance subjectively in Sensol are complying with the company mandates. These average subjective performance ratings are both below the mandated maximum of 120 on a 200-point scale.

Hypothesis 4 stated the expectation that, assuming support for H3, if there was any leniency in the evaluations, it would take place with the objective ratings. Using the fixed subjective ratings as a benchmark, which is consistent with the leniency comparisons made in prior studies (e.g., Bol 2009; Moers 2005), the objective ratings are significantly higher than the subjective ratings by an average of 18 points (on the 200-point rating scale) \( (t = 19.1; p < .000) \) across the two-year period. The results show evidence of leniency in the objective portion of the system and provide support for hypothesis 4. The results, though, appear to be driven solely by an interactive effect with economic stress, as is discussed below.

4.4 Effects of economic stress

Hypotheses 5 and 6 predicted that economic stress would create an extra pressure for leniency. This leniency could only be exercised through the objective ratings, so we expected to see both easier objective performance targets and higher objective ratings in the stressed times (Year 2) as compared to more normal times (Year 1). As was described above, Sensol managers did use subjectivity to adjust the objective performance targets for Year 2.
downward when details of the performance shock became known, but before we did our
analyses we did not know how significant those adjustments were.

We found that as compared to the year before the performance shock (Year 1), the
mean performance target difficulty was significantly lower in the year of stressed economic
conditions (Year 2) \( (t = 13.5; \ p < .001) \) (Table 1). This result provides strong support for H5.
The variance in target difficulty was also significantly lower in Year 2 \( (F = 946.0; \ p < .001) \).
Consistent with the decreased target difficulty, the objective performance ratings were
significantly higher in Year 2 as compared with Year 1 \( (60 \text{ points on the } 200\text{-point rating scale}; \ z = 32; \ p < .000) \), as is shown in Table 1. These results provide strong support for H6.

### 4.5 Multivariate Tests

To understand better the factors affecting each of the performance ratings and, hence,
the leniency bias, we ran regressions using the objective ratings, target difficulty, subjective
ratings, and the difference between the objective and subjective ratings (as an indicator of
leniency) as dependent variables, with all of the control variables—employee location,
employee type, and strategy—as independent variables. The results are shown in Table 3.

---- Insert Table 3 here ----

The Table 3 results show that the objective performance ratings are significantly
affected by both employee location and strategy. The objective performance ratings are
lower, and target difficulty is significantly higher for employees located at corporate, rather
than in one of the SBUs, and for employees working in a unit with a decline, rather than a
maintain, strategy. The subjective ratings are not affected by employee location or strategy.
However, they are higher for managers, and particularly high-ranking managers, and higher-
ranked professionals. The differences between the two types of ratings, and hence leniency,
stem from the same factors discussed above because none of the control variables has a significant effect on both the objective and subjective performance ratings.

We also ran the multiple regressions adding year as a dummy variable. The results are shown in Table 4. Year is highly significant. All ratings were higher in the year with difficult economic conditions. The results for the other variables were not materially affected by the addition of the year dummy, although some of the significance levels changed.

---- Insert Table 4 here ----

5. Insights from field visits

We followed up our empirical analyses with some field interviews to supplement our analyses of the company’s archival data. We also observed some management meetings, including some involving reviews of the company’s incentive compensation plan.

In our field visits, we wanted to understand was the extent to which employees included in the incentive plan paid attention to the amounts of money being offered and the bases for earning the incentive awards. Everyone we talked with was aware that they had what they considered to be significant amounts of pay at risk. Here is a representative quote:

It’s a significant amount of money at stake. For me, at risk is more than 30 percent of my salary. That’s just with a multiplier of 1, and for this year there’s a reasonable chance if things carry on, the multiplier could roll out to be over 2, so it could end up being 60 percent of my salary again. So that does get your attention, and funny enough it gets the attention of your partner.

We were particularly interested in learning more about the use of leniency in the objective portion of the incentive plan. As discussed above, one way that leniency was exercised was through the downward revision of targets for Year 2 from those that had been
agreed to earlier. A senior business unit manager explained the thinking behind this revision as follows:

We felt that we were obliged at the time to put in a fairly big stretch target. But when we sat down with [senior managers], they said, Well, look give us your worst case scenario rather than what the full stretch is and we will accept that. So that was when our original EBIT target was revised down. When you look at the politics of it all, basically [Sensol] had just had a shocking quarter … so [senior management] decided that if they were going to revise the budget, they would set it so that the actual result would never be any worse [than the budget]. The board, after that dismal quarter would have been happy with anything that involved a profit for the whole of [Sensol]. So, you know, they put a very achievable figure in there.

Another explained:

They accepted much lower targets. Now as it happened this year [Year 2], our performance has been way above target, and we’re projecting to continue similar levels of performance going forward. So we’ve gone from one extreme where set the targets too high to the other extreme where they’re perhaps too low.

It is possible that variance analysis-type accounting adjustments were also made in Year 2 ex post for similar reasons. We found evidence that these types of adjustments were made in other years, but we were not able to collect any stories of such adjustments being made in Year 2. Perhaps they were not deemed necessary because the leniency in the re-set performance targets for Year 2 was sufficient.

We explored our finding of positive relationships between the employees’ ranks and the subjective performance ratings. Only a few Sensol employees, mostly in the corporate human relations (HR) department, were aware of this relationship. One plausible explanation posited for this relationship was that when comments are made in meetings from a higher-level person, they are taken more seriously. The higher ranking people were sometimes seen as making better contributions, but perhaps only because they spoke with “more stature.” To
minimize this potential bias, HR had suggested that subjective rating averages be published for each employee grade, but senior management rejected that idea.

We did find some evidence of the self-serving bias at top management levels. One top manager explained to us that “If my direct reports are not outstanding, then something is wrong with the way I am managing them.” This kind of bias combined with a less than rigorous enforcement of the required mean subjective performance rating at higher levels of the organization could lead to the pattern of results we found.

V. SUMMARY AND CONCLUSIONS

The purpose of this study was to explore the relationships between objective and subjective measures of performance in a performance measurement/incentive compensation system. It is well known that both objective and subjective measures are often important parts of such systems, but because of data access limitations much of the knowledge about the existence, causes, and effects of these two forms of measurement has come from experimental studies. This study exploited the opportunity to analyze archival data from a company setting that weighted objective and subjective measures equally in importance for performance evaluation and bonus allocation purposes. We obtained data about top- and middle-level managers’ and professionals’ objective and subjective performance ratings for two quite different years, one representing normal operating conditions and one a time of economic stress, which allowed us to exploit a natural experimental situation. And an interesting twist in this company’s system was a mandated maximum subjective performance rating. This mandated maximum was designed to eliminate the leniency bias that is said to be common with subjective performance evaluations. We were able to study whether the
mandated maximum was enforced, and if it was, whether leniency was exercised in the objective performance evaluation portion of the system.

In our empirical analyses, we found that the correlation between the objective and subjective performance ratings was positive but small (.16), which suggests the two forms of ratings are assessing substantially different aspects of performance, or maybe one or both are subject to significant measurement error. In any case, this finding suggests there is little or no performance evaluation halo effect.

The objective/subjective performance rating correlation we observed was lower than the averages found in two literature reviews in this area (Bommer et al. 1995; Heneman 1986). This may be caused by the fact that our sample of employees was, on average, at higher organizational ranks that most of the prior studies. We found the correlations between the objective and subjective performance ratings to be particularly low for higher-ranking managers, which suggests that the correlations might be negatively associated with rank. More research is needed in this area, but this finding could suggest that the objective measures are less informative of overall performance at higher organizational levels. At higher levels, more subjective judgment is needed.

We found no evidence of centrality bias in the subjective performance ratings. On the contrary, the subjective performance ratings were more highly differentiated than were the objective ratings. This finding, which is contrary to that of other studies, that deserves further study. It might be due to the fact that the measures reflect different performance levels of analysis; the objective measures reflect organizational performance, while the subjective measures reflect individual performance, or it might be an unintended result of the mandated
We looked for a leniency bias in the subjective performance ratings, but unlike many prior studies, such as Bol (2009), Moers (2005) and Ittner et al. (2003), we did not find it. The subjective ratings were, on average, lower than the objective ratings. The most important factor affecting our results in this area is probably the company’s mandated maximum SBU-level average subjective performance rating, which seems to have been rigorously enforced.

However, our findings also suggest that demands for leniency are strong, particularly in times of economic crisis. At our site leniency was granted where it could be—in the objective portion of the system. In the stressed year we studied, performance targets were lowered after they had been finalized to adjust for the negative effects of uncontrollable factors that became known and were not impounded in the original plan. At our site, variance analysis-type accounting adjustments were made ex post in some situations also. The total leniency exercised in the objective performance ratings portion of the system was significantly higher in the time of economic stress than it was in normal operating conditions. The evidence contradicts, or at least calls into question, the assumption commonly made by other researchers (e.g., Bol 2009; Moers 2005) that “objective performance ratings are unbiased or at least significantly closer to the true performance value than the subjective ratings” (Bol 2009, p. 15).

In our multivariate regression analyses we found that the objective performance ratings were lower for the corporation than for the SBUs, a result that deserves further study. We also found that the subjective performance ratings were higher for managers, as compared to professionals, and higher for employees at higher ranks. We do not know if
these contextual differences were caused by evaluation biases or if they simply reflect the fact that employees at higher ranks merely tend to perform better than those at lower ranks.

In total, our findings have important implications for the design of performance evaluation and incentive systems. For example, while numerous papers have shown that subjective evaluator bias can be reduced by training evaluators or using improved rating scales (Bol 2009), this paper shows that an administrative solution—enforced mandated rating means—can essentially eliminate the leniency bias. Leniency has its own set of advantages and disadvantages, and it is not always advantageous to avoid it (Bol 2009). But our findings show that if managers want to avoid leniency, they have to be careful to make sure it is not provided elsewhere, in portions of the system that are ostensibly totally objective. They also suggest a number of questions worthy of further research. For example, why do ratings differ systematically between corporate and business unit employees, between managers and professionals, and between employees of high and low rank? Are these differences caused by real performance differences, errors or biases?

Our findings, however, should be interpreted with the following caveats in mind. This was a one-company study. We do not know the extent to which our findings are generalizable to other companies. For sensitivity and confidentiality reasons, we were able to obtain the data only after a significant period of time had passed, although there is no reason to believe that our findings were affected by the specific time period we studied. For similar sensitivity reasons, our data set did not include identifiers for individual managers and professionals, so we could not compare specific individuals’ ratings across the two time periods. We had access to only two years of data, so our explorations of the effects of variables that change between time periods were limited. As with prior archival research in this topic area, the tests
for leniency assume that employees’ abilities to meet the performance standards for the dimensions of performance evaluated objectively and subjectively are equal. Bol (2009) found evidence that this assumption was warranted in her study, but of course our setting is somewhat different. And we did not have access to any data that would show any of the performance effects of the rating choices studied here, so we are unable to make any tentative normative suggestions. Despite these limitations, however, we believe that our findings contribute to a better understanding of how objective and subjective performance evaluations are used and how they are related.
References


### Table 1
Descriptive statistics for performance rating and target difficulty variables

<table>
<thead>
<tr>
<th>Performance ratings:</th>
<th>Mean</th>
<th>Range</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OBJECTIVE:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1</td>
<td>101</td>
<td>84-163</td>
<td>16</td>
</tr>
<tr>
<td>Year 2</td>
<td>161</td>
<td>109-169</td>
<td>16</td>
</tr>
<tr>
<td>2-year period</td>
<td>131</td>
<td>84-169</td>
<td>34</td>
</tr>
<tr>
<td><strong>SUBJECTIVE:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1</td>
<td>111</td>
<td>40-180</td>
<td>22</td>
</tr>
<tr>
<td>Year 2</td>
<td>115</td>
<td>50-180</td>
<td>17</td>
</tr>
<tr>
<td>2-year period</td>
<td>113</td>
<td>40-180</td>
<td>20</td>
</tr>
<tr>
<td><strong>TARGET DIFFICULTY</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1</td>
<td>2.41</td>
<td>.14-7.11</td>
<td>2.92</td>
</tr>
<tr>
<td>Year 2</td>
<td>.87</td>
<td>.15-3.65</td>
<td>1.08</td>
</tr>
<tr>
<td>2-year period</td>
<td>1.63</td>
<td>.14 – 7.11</td>
<td>2.32</td>
</tr>
</tbody>
</table>

### Table 2
Employee descriptive statistics

<table>
<thead>
<tr>
<th>E_TYPE (employee type)</th>
<th>Year 1</th>
<th>Year 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Managers</td>
<td>343</td>
<td>337</td>
</tr>
<tr>
<td>- senior</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>- middle</td>
<td>310</td>
<td>303</td>
</tr>
<tr>
<td>- Professionals</td>
<td>377</td>
<td>406</td>
</tr>
<tr>
<td>- high</td>
<td>195</td>
<td>211</td>
</tr>
<tr>
<td>- low</td>
<td>182</td>
<td>195</td>
</tr>
<tr>
<td>E-LOCATION (location of employee)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>- Corporate</td>
<td>197</td>
<td>120</td>
</tr>
<tr>
<td>- Business unit</td>
<td>523</td>
<td>623</td>
</tr>
</tbody>
</table>
Table 3
Effects on performance ratings, target difficulty, and the difference between objective and subjective performance ratings

<table>
<thead>
<tr>
<th></th>
<th>Objective performance rating</th>
<th>Target difficulty&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Subjective performance rating</th>
<th>Difference between objective and subjective performance ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.36 (.020) ***</td>
<td>1.05 (.089) ***</td>
<td>1.06 (.010) ***</td>
<td>0.30 (.020) ***</td>
</tr>
<tr>
<td>EMPLOYEE LOCATION&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.13 (.023) ***</td>
<td>4.11 (.104) ***</td>
<td>0.02 (.012)</td>
<td>-0.02 (.024) ***</td>
</tr>
<tr>
<td>STRATEGY&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-0.05 (0.020) *</td>
<td>-0.25 (.095) **</td>
<td>-0.01 (.011)</td>
<td>-0.04 (.022)</td>
</tr>
<tr>
<td>EMPLOYEE TYPE:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior manager&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.04 (.045)</td>
<td>-0.47 (.204) *</td>
<td>0.37 (.024) ***</td>
<td>-0.33 (.047) ***</td>
</tr>
<tr>
<td>Middle manager&lt;sup&gt;e&lt;/sup&gt;</td>
<td>-0.02 (.022)</td>
<td>-0.40 (.100) ***</td>
<td>0.11 (.012) ***</td>
<td>-0.12 (.023) ***</td>
</tr>
<tr>
<td>High professional&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.00 (.024)</td>
<td>-0.20 (.110)</td>
<td>0.04 (.013) **</td>
<td>-0.04 (.025)</td>
</tr>
<tr>
<td>F = 6.69</td>
<td>F = 380.15</td>
<td>F = 57.46</td>
<td>Adj. R² = .02</td>
<td>Adj. R² = .07</td>
</tr>
<tr>
<td>Adj. R² = .57</td>
<td>Adj. R² = .16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OLS regressions with dependent variables as shown. The reported statistics include the coefficient estimates with standard errors in parentheses. Significance levels are indicated as *** p<.001; ** p<.01; * p<.05 (two-tailed).

---

<sup>a</sup> the difference between the maximum target minus the lower target limit divided by the actual performance target

<sup>b</sup> corporate = 1; SBU = 0

<sup>c</sup> grow = 1; maintain = 0

<sup>d</sup> senior manager = 1; else 0

<sup>e</sup> middle manager = 1; else 0

<sup>f</sup> high ranking professional = 1; else 0
Table 4
Effects on performance ratings, target difficulty, and the difference between objective and subjective performance ratings, adding year a dummy explanatory variable

<table>
<thead>
<tr>
<th></th>
<th>Objective performance rating</th>
<th>Target Difficulty&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Subjective performance rating</th>
<th>Difference between objective and subjective performance ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.03 (.010) ***</td>
<td>1.66 (.093) ***</td>
<td>1.03 (.011) ***</td>
<td>0.00 (.014)</td>
</tr>
<tr>
<td>EMPLOYEE LOCATION&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.03 (.011) *</td>
<td>3.92 (.098) ***</td>
<td>0.03 (012)*</td>
<td>-0.06 (.015) ***</td>
</tr>
<tr>
<td>STRATEGY&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-0.05 (.010) ***</td>
<td>-0.24 (.089) **</td>
<td>-0.01 (.011)</td>
<td>-0.04 (.014)**</td>
</tr>
<tr>
<td>EMPLOYEE TYPE:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior manager&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.02 (.021)</td>
<td>-0.44 (.191) *</td>
<td>0.37 (.024) ***</td>
<td>-0.34 (.030)***</td>
</tr>
<tr>
<td>Middle manager&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.00 (.010)</td>
<td>-0.43 (.094) ***</td>
<td>0.11 (.012) ***</td>
<td>-0.11 (.015)***</td>
</tr>
<tr>
<td>High professional&lt;sup&gt;f&lt;/sup&gt;</td>
<td>-0.01 (.011)</td>
<td>-0.182 (.103)</td>
<td>0.04 (.013) **</td>
<td>-0.05 (.016)**</td>
</tr>
<tr>
<td>YEAR&lt;sup&gt;g&lt;/sup&gt;</td>
<td>0.60 (.008)***</td>
<td>-1.10 (.076)***</td>
<td>0.05 (.009)***</td>
<td>0.55 (.012)***</td>
</tr>
<tr>
<td>F = 870.23</td>
<td>Adj. R&lt;sup&gt;2&lt;/sup&gt; = .78</td>
<td>F = 397.49</td>
<td>Adj. R&lt;sup&gt;2&lt;/sup&gt; = .62</td>
<td>F = 53.33</td>
</tr>
<tr>
<td>Adj. R&lt;sup&gt;2&lt;/sup&gt; = .62</td>
<td>F = 412.09</td>
<td>Adj. R&lt;sup&gt;2&lt;/sup&gt; = .63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OLS regressions with dependent variables as shown. The reported statistics include the coefficient estimates with standard errors in parentheses. Significance levels are indicated as *** p<.001; ** p<.01; * p<.05 (two-tailed).

---

<sup>a</sup> the difference between the maximum target minus the lower target limit divided by the actual performance target
<sup>b</sup> corporate = 1; SBU = 0
<sup>c</sup> grow = 1; maintain = 0
<sup>d</sup> senior manager = 1; else 0
<sup>e</sup> middle manager = 1; else 0
<sup>f</sup> high ranking professional = 1; else 0
<sup>g</sup> year 1 = 0 else year 2 = 1

37