

**EFFECTS OF SUBORDINATE LIKEABILITY AND BALANCED SCORECARD  
FORMAT ON PERFORMANCE-RELATED JUDGMENTS**

Steven E. Kaplan  
School of Accountancy  
W.P. Carey School of Business  
Arizona State University  
CI Business Box 873606  
Tempe, AZ 85287-3606

Michael J. Petersen  
School of Accountancy  
W.P. Carey School of Business  
Arizona State University  
CI Business Box 873606  
Tempe, AZ 85287-3606

Janet A. Samuels  
Accounting Department  
School of Global Management and Leadership  
Arizona State University, West Campus  
Business Box 37100  
Phoenix, AZ 85069-7100

July, 2006

**Data Availability:** Data gathered in this study are available from the authors upon request

# **EFFECTS OF SUBORDINATE LIKEABILITY AND BALANCED SCORECARD FORMAT ON PERFORMANCE-RELATED JUDGMENTS**

## **ABSTRACT**

Previous appraisal research has found that subordinate manager likeability biases appraisal related judgments (Cardy and Dobbins, 1986; Tsui and Barry, 1986; Turban et al., 1990). We investigate whether the format and structure of the Balanced Scorecard (BSC) moderates subordinate likeability bias on evaluators' performance-related judgments (overall performance rating, strength of conviction in rating, and bonus money allocation). Following Lipe and Saltario (2002), we characterize the evaluation of subordinate managers using a multidimensional performance measurement system as a cognitively difficult task. We hypothesize that, in comparison to when a BSC format is not used, use of a BSC format fosters a 'divide-and-conquer' approach by evaluators and that the structure of the BSC will guide the application of this approach, which, in turn, is expected to mitigate the influence of subordinate likeability on evaluators' performance-related judgments. Unexpectedly, we find that the biasing effect of subordinate manager likeability on performance-related judgments is not moderated by the format and structure of the BSC. We also provide supplemental mediation analysis which demonstrates that likeability has a direct effect on bonus allocations in addition to the indirect effect it conveys through improved overall performance ratings.

## INTRODUCTION

The balanced scorecard (BSC) was developed by Kaplan and Norton (1992, 1996a, 1996b) for multidimensional performance measurement. A BSC typically contains a set of measures spanning four broad performance categories: financial performance, customer relations, internal business processes, and the organization's learning and growth activities. Collectively, these measures are intended to represent "an integrated set of leading and lagging performance measures designed to capture the organization's strategy" (Lipe and Salterio 2000, p. 285). Firms that implement multidimensional performance measurement systems often use them for performance evaluation and compensation purposes. A critical issue facing these firms is whether to assign explicit weights to each measure or to allow subjectivity in weighting each measure (Ittner et al. 2003). Using explicit weights may cause dysfunctional behavior if the weights are not appropriately assigned (Krishnan et al. 2005). However, allowing subjectivity in weighting may allow performance evaluations to be subject to biases. To inform this issue, research has examined the extent to which attributes of the BSC performance measures (e.g., common vs. unique) (Banker et al. 2004; Dilla and Steinbart 2005; Libby et al. 2004; Lipe and Salterio 2000; Roberts et al. 2004) and the organizing features of the BSC (Lipe and Salterio 2002) influence subordinate performance-related judgments. This work portrays evaluators as information processors and examines how their cognitive capabilities and shortcomings may systematically influence their performance-related judgments.

Previous experimental research on the use of multiple performance measures has not explored the role, if any, of affect toward subordinate managers on overall performance evaluations. Because of their ongoing professional relationship, evaluators typically will have formed an impression of subordinate managers' likeability. In this regard, subordinate managers

possess positive likeability when their personal attributes are generally regarded as having favorable implications. Alternatively, subordinate managers possess negative likeability when their personal attributes are generally regarded as having unfavorable implications. To the extent that performance evaluations are more valid when based exclusively on behaviors and associated performance measures (Balzer 1986; Dipboye 1985; Lefkowitz 2000), consideration of a subordinate's likeability when evaluating the subordinate's performance represents a source of bias and may be dysfunctional. Performance evaluation researchers (Bates 2002; DeNisi et al. 1997; Robbins and DeNisi 1994, 1998) have shown that subordinate likeability influences evaluators' processing of information about subordinates' performance and their subsequent evaluations of subordinates. For example, research in this area has found that the evaluator's personal liking for a subordinate is associated with both rating accuracy (Cardy and Dobbins 1986) and rating errors (Tsui and Barry 1986).

Within accounting, the role of affect on judgment and decision making has been relatively limited (Moreno et al. 2002). However, Kida and Smith (1995) recognize that affect is likely to influence cognitive processes of decision makers using accounting based information. Both Kida et al. (2001) and Moreno et al. (2002) report that managers' affective reactions significantly influence capital budgeting decision-making.

In this paper we propose that the structure of the BSC mitigates the influence of subordinate likeability on evaluators' performance-related judgments of subordinates (i.e., overall performance rating, strength of conviction in subordinate's rating, and bonus allocation to subordinate). Thus, the purpose of our research is to complement and extend research on whether and how the structure of the BSC affects evaluators' subordinate related performance judgments (Lipe and Salterio 2002). In contrast to previous BSC studies, our study includes a strength of

conviction judgment. Subordinate likeability is expected to influence strength of conviction judgments under an affect-consistency heuristic, which is discussed below. BSC research indicates that when faced with a complex task involving many performance measures, evaluators will simplify their task through the use of heuristics. For example, Lipe and Salterio (2002) contend that the structure of the BSC triggers the use of a divide and conquer approach (Shanteau 1988) and that this approach will be guided by the structure of the BSC (Lipe and Salterio 2002). Alternatively, in the absence of the BSC structure, the entire set of performance measures are presented as a single category. We expect evaluators will use an affect-consistency approach to simplify and guide their judgments. Under this approach, evaluators attend to and weight information that is consistent with subordinate's likeability (Robbins and DeNisi 1994). Thus, in a setting where two subordinates perform similarly within each BSC category, we contend that this similarity in performance will be made more apparent through the use of a divide and conquer approach under a BSC format and, consequently, mitigate the influence of subordinate likeability on subordinate performance-related judgments. Alternatively, in the absence of the BSC format, subordinate likeability is expected to have a greater influence on evaluators' judgments through the use of the affect consistency heuristic in which measures that are consistent with the subordinate's likeability are more heavily weighted and attention is primarily directed to the likeable subordinate's performance.

Using an experimental approach similar to previous BSC studies (Banker et al. 2004; Dilla and Steinbart 2005; Libby et al. 2004; Lipe and Salterio 2000, 2002; Roberts et al. 2004), evaluators' rated the performance of two division managers, indicated their strength of conviction in their ratings for each manager, and allocated a performance bonus to each manager based upon background information about the company as well as performance information for

multiple performance measures. The study manipulated performance measures organization (e.g., presence or absence of the BSC format) and information about the subordinates' likeability (e.g., Like Manager A / Dislike Manager B and vice versa) in a fully-crossed design.

Unexpectedly, the results do not support our expectation that the structure of the BSC mitigates the influence of subordinate likeability on evaluators' performance evaluation judgments. Instead, subordinate likeability was found to significantly influence evaluators' subordinate performance-related judgments (performance rating, strength of conviction and bonus) regardless of the performance measure format. Additionally, as part of a supplemental analysis, we find that while performance ratings partially mediate the effect of likeability on the bonus allocated to each manager, likeability still has a significant direct effect on the bonus. These results indicate that the likeability bias is robust to various performance evaluation formats.

These results have several important implications. First, prior research has found that the BSC organizes measures to reduce information processing complexity (Lipe and Salterio 2002). We suggest that this organizing framework will help mitigate the biasing effect of likeability on performance-related judgments. However, we find that the likeability bias is not influenced by the organizing format of the performance measurement system. This suggests that designers of performance measurement systems may not be able to reduce the effect of likeability on performance-related judgments via the format of the performance measurement system. Also, in a supplemental analysis, we find that in addition to likeability's significant influence on overall performance ratings which, in turn, influence compensation (bonus), likeability also directly influences compensation over and above the indirect effect through performance ratings. Therefore, the importance of likeability is significant on overall performance ratings and is

magnified on compensation. Over time, subordinates may become aware of the importance of likeability in performance-related judgments. In an effort to improve these judgments, they may direct effort away from activities that improve performance measure outcomes and toward efforts that improve their perceived likeability. This re-direction of effort may be dysfunction to an organization.

Finally, this study contributes to the evaluation literature by examining the effect of likeability within a formal performance measurement system. While many of the performance measurement systems currently advocated (e.g. the BSC) or currently used in practice have these features, prior evaluation research that has investigated the likeability bias has not provided evaluators with specific performance measures, targets and actual outcomes. This study finds that even providing specific performance measures and targets does not eliminate the likeability bias on performance evaluation.

The rest of the paper is organized as follows. The next section reviews relevant literature. In the subsequent sections we present our hypotheses, describe our research method, and present our results, including supplemental analysis. In the last section of the paper we then discuss the implications and limitations of our research, and offer suggestions for further research.

## **BACKGROUND**

Affect refers to feeling states and encompasses two broad categories: undifferentiated and differentiated (Park et al. 1987). Undifferentiated affect refers to mood states and is not directed at a particular entity. For example, feelings of joy or sadness are mood states. An evaluator experiencing one of these moods might be expected to inflate or deflate evaluations of any subordinate. Thus, while undifferentiated affect might bias evaluations, this form of bias is

expected to affect evaluations of all subordinates equally (Cardy and Dobbins 1994). Alternatively, differentiated affect constitutes a like-dislike reaction to an individual and has been characterized as “the major currency in which social intercourse is transacted” (Zajonc 1980). In contrast to undifferentiated affect, this form of affect differentiates among subordinates. In this study we examine the effect of differentiated affect between two subordinate managers on evaluators’ performance-related judgments in a multiple performance evaluation setting.

Early research by Alexander and Wilkins (1982) found that subordinate likeability had a greater effect when subjective rather than more objective performance measures were used. Commonly, no explicit weights are assigned to each measure included as part of a multiple performance measurement system such that forming an overall performance evaluation is largely subjective. That is, while many, if not all, of the individual measures may be relatively quantifiable and considered “objective”, discretion is left on how to combine measures into an overall performance evaluation of subordinate managers. Therefore, without explicit weights, performance evaluation judgments in a multiple performance measurement system (such as the BSC) are subjective and likely to be affected by subordinate likeability.

Findings from Lipe and Salterio (2000) suggest that when evaluating two subordinates, performance evaluation judgments are likely to be based on the relative performance of each subordinate. Further, given the ambiguous setting and the complexity of the task, performance evaluations are likely to be biased. For example, experimental BSC research has generally found that given the subjective nature of the process, the complexity of the task, and cognitive shortcomings of individuals, when evaluating two managers, unaided performance ratings reflect a common-measures bias (Banker et al. 2004; Dilla and Steinbart 2005; Libby et al. 2004; Lipe

and Salterio 2000). That is, a performance measure that is common between two managers is weighted more heavily than a performance measure this is unique to each manager.

However, use of strategy maps (Banker et al. 2004), a “disaggregated BSC” (Roberts et. al. 2004), or process accountability or third-party assurance about performance measures (Libby et al. 2004) have been found to mitigate the common measures bias. These studies are set exclusively within a BSC framework and do not consider subordinate manager likeability. However, the existence of the common measures bias would not appear to be due to the structure of the BSC. Instead, this bias more generally represents evaluators’ attempt to simplify the complex task of evaluating performance in a setting with multiple performance measures.

Lipe and Salterio (2002) examine the conditions under which the structure imposed by the BSC (e.g., grouping measures into one of four categories) affects performance evaluation judgments. Based on research in cognitive psychology (Baddeley 1994; Miller 1956), Lipe and Salterio (2002), consistent with their earlier work (Lipe and Salterio 2000), assert that the use of a large number of performance measures to evaluate performance represents a complex task. This assertion is based on research that has found that a positive relationship between task complexity and number of cues or attributes (Bonner 1994; Wood 1986). More specifically, Lipe and Salterio (2002, p. 532), state that separately processing, weighing, and combining a large number of performance measures into an overall evaluation “is cognitively, a very difficult thing to do.” Within cognitive psychology, individuals are portrayed as having limited information processing capacity. For example, research findings indicate that individuals are only able to simultaneously process approximately seven to nine information cues (Miller 1956). A broad range of research supports this view (Baddeley 1994).

Lipe and Salterio (2002, p. 533) contend that when faced with a complex evaluation task, evaluators will invoke a heuristic such as a divide and conquer information processing approach and that “the organization of the BSC lends itself quite naturally to this kind of mental approach.” Under a divide and conquer approach, each of the performance outcomes is assigned to one of a small number of groups. Next, performance is assessed for each group based upon the measures that have been assigned to the group. Because each group contains a relatively small number of measures, evaluation at the group level represents a less cognitively demanding task. Finally, separate group evaluations are combined into an overall evaluation.

Lipe and Salterio (2002, p. 533) contend that whether the structure of the BSC influences overall performance evaluations depends upon the specific pattern of outcomes among performance measures. In particular, Lipe and Salterio (2002) predict that the structure of the BSC will influence performance evaluation judgments when the measures within a BSC category consistently reflect performance by one subordinate was better than another subordinate. Based on a divide and conquer approach, simultaneously processing a small number of measures that are grouped together should facilitate evaluators’ ability to see relations among the measures. Such perception, in turn, will influence the subjects’ weighting of the measures. Specifically, two (or more) measures that are perceived to be related will be weighted less than two measures that are perceived to be unrelated. Thus, when all the measures that are consistently favorable towards one subordinate are included within a single BSC category, these measures will be weighted less than when these measures are not grouped together. Alternatively, when the measures that are grouped together do not consistently favor one subordinate, then Lipe and Salterio (2002) contend that the structure of the BSC will not influence overall performance ratings. The results of two experiments provide support for their predictions. However, their

study on the structuring role of the BSC on performance evaluation judgments does not consider the potential influence of subordinate manager likeability. As discussed below, we contend that when subordinate manager likeability is considered, the organizing structure of the BSC will influence performance evaluation judgments even when performance measures do not consistently favor one manager over another.

## **HYPOTHESES**

Performance appraisal researchers (Cardy and Dobbins 1994; Dipboye 1985; Lefkowitz 2000; Murphy and Cleveland 1991; Turban et al. 1990; Varma et al. 1996) recognize that performance evaluation is a judgment involving both affective and cognitive aspects. In a seminal paper, Zajonc (1980, p. 154) contends that affective reaction “is capable of influencing the ensuing cognitive process to a significant degree.” That is, affect influences cognitive processes, which in turn, influence performance ratings (Antonioni and Park 2001). For example, affect may impact what and how a manager observes the work of a subordinate, and consequently, the manager’s memories about the subordinate. Evidence consistent with this view is presented by Murphy et al. (1986). They report that evaluators’ recall of behavioral information tends to be consistent with their general impressions of a subordinate. Additionally, research (Cardy and Dobbins 1986; Tsui and Barry 1986; Turban et al. 1990) investigating the role of affect on appraisal judgments generally occurs in a setting without performance criterion (e.g., targets). In such a setting, Baltes and Parker (2000) contend that subordinate affect may induce bias in evaluators’ implicit performance targets. That is, evaluators may use a relatively lower performance standard for a likeable subordinate and a relatively higher performance standard for an unlikeable subordinate. Formal performance measurement systems would not be

subject to these cognitive biases for two reasons. First, rather than having managers recall behavioral information, specific performance measures and outcomes are used to assess performance. Second, evaluators are not able to use differing relative performance standards for more or less likeable subordinates as specific targets are typically established for each performance measure. Therefore, formal performance measurement systems, which include specific performance measures, pre-determined targets and actual outcomes, are not likely to be subject to cognitive biases of likeability due to subjective recall and relative performance standards.

However, in addition to influencing observations of a subordinate's behavior and relative performance standards, researchers contend that affect also influences cognitive processing of performance related information during the performance evaluation task. In this regard, Robbins and DeNisi (1994, 1998) integrate affective and cognitive aspects by proposing an affect-consistency bias. Under an affect-consistency bias, evaluators attend to and emphasize information that is consistent with their affective reaction. Thus, based on this bias, evaluators are expected to seek out and elevate the importance of favorable performance information about a well-liked subordinate and are expected to discount poor performance information "as not meaningful or an aberration" (Robbins and DeNisi 1994, p. 343). This tendency reverses for a subordinate who is unlikeable such that evaluators will seek out and elevate the importance of unfavorable performance information and discount favorable performance information. Varma et al. (1996) further hold that evaluators' judgments are most likely to exhibit an affect-consistency bias when performance information is ambiguous, as is commonly the case with multidimensional performance measurement. As described, although not recognized by Robbins and DeNisi (1994, 1998), the affect-consistency heuristic may be viewed as a specific

application of motivated reasoning (Kunda 1990). Kunda (1990, p. 480) proposes that when motivated “to arrive at a particular, directional conclusion” individuals will rely on a “biased set of cognitive processes.” Presumably, an evaluator is motivated to evaluate a likeable subordinate favorably, creating a directional conclusion.

Robbins and DeNisi (1994) test their model in the context of undergraduate management students evaluating the performance of their instructors. This task was selected because of its familiarity to undergraduate students. Nine behavioral incidents were developed to reflect performance information for each of the three instructors. Approximately one-third of the incidents represented above-average, average, and below-average levels of performance, respectively. Several weeks earlier, these students had completed a questionnaire to assess interpersonal affect towards each of the three instructors. The results of the study indicate that affect consistency was not associated with the information acquisition stage. However, affect consistency was significantly associated with the weighting stage. As expected, affect consistent incidents were assigned a higher weight than either affect neutral or affect inconsistent incidents. Finally, overall instructor ratings exhibited an affect consistency bias. Subsequently, Robbins and DeNisi (1998) provide evidence that affect consistency had a stronger influence on the weighing of instructor incidents and instructor ratings than mood consistency. However, similar to other evaluation research, Robbins and DeNisi (1994; 1998) did not provide participants with performance benchmarks or targets. Therefore, their results could be due to the affect-consistency bias or to participants using differing targets or benchmarks for more or less likeable instructors.

The current study examines the influence of subordinate likeability in an organizational environment involving a formal multiple performance measure system (PMS). With a formal

PMS, specific performance measures are identified along with pre-determined targets and actual outcomes are reported. Therefore, with the exception of the affect-consistency bias, prior explanations for the likeability bias are controlled for. We contend that the extent of the affect-consistency bias is likely to depend on performance measure organization (e.g., whether or not the BSC is used to organize the multiple measures). As discussed above, performance evaluation judgments based on a large number of different performance measures represent a complex evaluation task, and consequently, evaluators are likely to rely on heuristics to simplify the task. First, consider a setting in which the BSC is not used as a basis to organize the multiple measures. Instead, the many measures are presented as one large group of performance measures for each of two managers, one likeable and one unlikeable. In such a setting, based on prior appraisal research (Robbins and DeNisi 1994, 1998) we expect evaluators to simplify their information processing by adopting the affect-consistency heuristic. Under this heuristic evaluators direct their attention and differentially weight those measures that favor a likeable manager. In this context, a performance measure favors a likeable manager when a likeable manager outperforms an unlikeable manager on the performance measure. Thus, evaluators are expected to differentially attend to and place greater weight upon those measures favoring a likeable manager and either ignore or discount those measures that do not favor a likeable manager. By using this heuristic, a likeable manager's performance rating and, consequently, bonus allocation will increase. That is, by selectively weighting performance measures favoring a likeable manager, the likeable manager is expected to receive an inflated performance rating, and consequently, a higher bonus allocation. Also, under this heuristic, once an evaluator recognizes that the outcomes of a performance measure favors a likeable manager, we also expect evaluators to differentially attend to outcome information specifically about a likeable

manager relative to the outcome information specifically about an unlikeable manager. We expect differential attention to occur because the evaluator, in simplifying the task, is primarily developing a favorable rating for the likeable manager, and consequently will specifically seek out and attend to performance measures and outcomes that favor the likeable manager. Overall, in a non-BSC setting, in comparison to an unlikeable manager, evaluators are expected to inflate the performance ratings and bonus assignments for a likeable manager, and exhibit greater conviction in the performance rating of a likeable manager.

Alternatively, consider a setting in which the BSC is used as a basis to organize the multiple measures about each of two managers, one likeable and one unlikeable. Given the complexity of the task and structure provided by the BSC, evaluators are expected to invoke a divide and conquer heuristic to simplify the task and guide their information processing (Lipe and Salterio 2002). That is, because the format of BSC naturally disaggregates performance measures among four categories, evaluators are expected to initially evaluate managers within a BSC category. By dividing the overall task into several smaller tasks, evaluators are expected to attend to each of the relatively few performance measures within the category (Lipe and Salterio 2002), including those that do not favor a likeable manager. Thus, the structure of the BSC, by fostering attention and consideration of all performance measures within and across performance categories, is expected to moderate the influence of manager likeability on performance evaluation judgments. That is, the BSC format is expected to decrease the influence of manager likeability by increasing the likelihood that measures both favorable and unfavorable to a likeable manager are considered by an evaluator. To the extent that the format BSC successfully guides and focuses attention to all performance measures, manager likeability should have less of an influence of performance-related judgments, and the performance evaluation judgments

should be based more on the information content of the performance measures. Overall, in a BSC setting, manager likeability is expected to have less of an influence on performance evaluation ratings, strength of conviction in the performance rating, and bonus compensation allocations. This discussion leads to the following hypotheses:

**Hypothesis One (a):** The likeability of subordinate managers and performance measure organization will interact to influence overall performance evaluations. Specifically, the influence of subordinate manager likeability on performance ratings will be significantly diminished when a BSC is used compared to when a BSC is not used.

**Hypothesis One (b):** The likeability of subordinate managers and performance measure organization will interact to influence strength of conviction in performance evaluations. Specifically, the influence of subordinate manager likeability on strength of conviction will be significantly diminished when a BSC is used compared to when a BSC is not used.

**Hypothesis One (c):** The likeability of subordinate managers and performance measure organization will interact to influence bonus allocations. Specifically, the influence of subordinate manager likeability on bonus allocations will be significantly diminished when a BSC is used compared to when a BSC is not used.

## METHOD

### Overview and Task

The task and experimental materials were based on those developed by Lipe and Salterio (2000, 2002). Participants were presented with a case involving the Women's Clothing Stores, Incorporated (WCS), a retail firm specializing in women's apparel. While WCS is described as having multiple divisions, participants were told that the case focuses on the two largest divisions. In their assigned role as a senior executive of WCS, participants were instructed that their task was to evaluate the performance of the division managers for the two largest divisions.

The case indicated that the company has been using multiple performance measures for several years. The case included a discussion of how company strategy and customer attributes were considered in the development of a common set of multiple measures and targets that were established for all divisional managers for the current year. Thus, the strategy, performance measures, and targets were the same for both divisions. This experimental design was used to ensure that the common measures bias (Lipe and Salterio 2000) did not confound our results.

Table 1 presents targets and actual performance levels for the multiple performance measures for both division managers being evaluated using the BSC format. As shown, a total of sixteen performance measures are included, divided equally among the four BCS categories. The case was designed such that each manager outperformed the other manager on two of four measures within each BSC category and that their overall performance based on an equal weighting of each measure was quite similar.

-----  
Insert Table 1 about here  
-----

For each division manager, participants made an overall performance rating. Similar to Lipe and Salterio (2000), participants were asked to place an “X” on a 101-point scale anchored by 0 (labeled “reassign”) and 100 (labeled “excellent”). Next, participants were asked to indicate their strength of conviction in their overall performance rating. Again, participants were asked to place an “X” on a 101-point scale. For this scale, 0 was labeled “very weak” and 100 was labeled “very strong.” After evaluating both managers, participants were asked to allocate a \$20,000 bonus between the two divisional managers. The bonus was included to provide evidence on potential economic consequences associated with performance ratings. A similar bonus measure

has been included in previous BSC research (Dilla and Steinbart 2005, Roberts et al. 2004).

Lastly, participants completed a debriefing questionnaire that collected demographic and other perceptual information about the task.

## **Design**

The study employed a 2 x 2 between subjects design. The first between subjects factor manipulated performance measure organization. Under the BSC format condition, the BSC concept was described and four BSC categories were defined. Further, the sixteen performance measures were classified into the appropriate BSC category – financial, customer related, internal business processes, and learning and growth. Thus, each BSC category contained four different performance measures. Under the NOFORM condition, the sixteen measures were presented without the BSC format. That is, the sixteen performance measures were listed under a single grouping. Similar to Lipe and Salterio (2002), under the NOFORM condition, performance measures were listed either alphabetically or based upon a random ordering.<sup>1</sup>

The second between subjects variable manipulated manager likeability. Likeability information about each division manager was provided in addition to a short professional biography describing each division manager's background. A set of five personal attributes was included to convey the likeability of each division manager. These personal attributes were based on work from Anderson (1968).<sup>2</sup> He identified 555 personal attributes that had positive, negative, or no association with the formation of likeability perceptions. The set of five personal attributes used to portray positive likeability included four positive attributes and one neutral attribute from Anderson (1968). The five attributes are cheerful, loyal, tactful, wholesome, and methodical. The last item represents an attribute with neutral likeability implications. The set of five personal attributes used to portray negative likeability included four negative attributes and

one neutral attribute from Anderson (1968). The five attributes are boastful, gossipy, self-centered, superficial, and systematic. The last item represents an attribute with neutral likeability implications. Our approach is similar to the approach used by Cardy and Dobbins (1986) and Dobbins and Russell (1986). In each of these studies, subordinate manager likeability was established by providing a set of five personal attributes from Anderson (1968).

Under the “Like Manager A” condition, the set of five positive likeability attributes were used to describe Division Manager A, Chris Peters, and the set of five negative likeability attributes were used to describe Division Manager B, Taylor Graham. Specifically, under this condition, the description read, in part, “In working with Chris over the last two years, you have found him to be loyal, cheerful, methodical, wholesome, and tactful.” Also, under this condition, the description read, in part, “In working with Taylor over the last two years, you have found him to be boastful, gossipy, systematic, self-centered, and superficial.” Alternatively, under the “Like Manager B” condition, the set of five positive likeability attributes were used to describe Division Manager B and the set of five negative likeability attributes were used to describe Division Manager A.

### **Dependent Measures**

Hypotheses one (a) through (c) focus on overall performance rating, strength of conviction, and compensation for the two division managers. To test the effect of subordinate manager likeability on overall performance ratings and strength of conviction in performance ratings a difference score is determined for each judgment and used as a dependent measure. The performance rating difference score is calculated by subtracting the overall performance rating of division manager B from the overall performance rating of division manager A. Thus, a positive value indicates that division manager A was rated higher than division manager B. Alternatively,

a negative value indicates that division manager B was rated higher than division manager A. A conviction difference score is calculated by subtracting the strength of conviction score of division manager B from the strength of conviction score of division manager A. Thus, a positive value indicates that the strength of conviction in the rating of division manager A was stronger than the strength of conviction in the rating of division manager B. Alternatively, a negative value indicates that the strength of conviction is stronger in the rating of division manager B than division manager A. Regarding tests of the effect of relative manager likeability on compensation, participants allocated a bonus pool between manager A and manager B. Therefore, the total bonus is constant and manager B's bonus is a function of the bonus allocated to manager A. Consequently, we use the bonus assigned to division manager A as the dependent measure.

### **Participants**

Evening MBA students at a major metropolitan state university were used as participants for the study. The questionnaire was administered after the topic of the BSC had been discussed in the course. Ninety-nine students enrolled in a managerial accounting course completed the questionnaire and passed the manipulation check, discussed below. Background information about these participants is presented in Table 2. As shown, the majority of participants were male and had evaluated individuals in the past. The mean age of participants was approximately 29 years, the mean years of work experience was approximately 6 years, and the mean number of accounting classes was approximately 4.

-----  
Insert Table 2 about here  
-----

## **RESULTS**

## **Liking Manipulation Check**

After responding to the case, participants answered two manipulation check questions to assess the extent to which likeability was perceived as intended. As part of the debriefing questionnaire, participants rated the likeability of each division manager. Each statement read, in part, “Based on the information presented, please indicate your impression of how likeable \_\_\_\_\_ is.” The first statement referred to the manager of division A and the second statement referred to the manager of division B. The end points on a seven-point scale for each statement were “Completely Unlikeable” (1) and “Completely Likeable” (7). In the “Like manager A” and “Like manager B” conditions of manager likeability, responses to these two items were used to determine whether any of the participants did not interpret or attend to the likeability manipulation. For the “Like manager A” condition, participants who did not rate the relative likeability of manager A higher than manager B were assumed either to misinterpret or not attend to the manager likeability manipulation, and consequently, were dropped from the analysis. Out of a total of 69 participants assigned to the “Like manager A” condition, twenty participants were dropped. For the “Like manager B” condition, participants who did not rate the relative likeability of manager B higher than manager A were assumed either to misinterpret or not attend to the manager likeability manipulation, and consequently, were dropped from the analysis. Out of a total of 67 participants assigned to the “Like manager B” condition, seventeen participants were dropped. Participants responding “incorrectly” to the two likeability manipulation check items were dropped because of a lack of “inclusion importance” (Yates 1990, p. 376). In this regard, Tan and Yates (1995, p. 315) contend that “if a decision maker never even acknowledges the existence of a particular dimension, the decision maker cannot possibly respond to that dimension.” In the context of the current study, we restricted the sample

to participants who attended to and interpreted explicit attempts to manipulate the relative likeability of managers A and B as intended.<sup>3</sup>

Under the “Like manager A” condition, the mean likeability ratings (standard deviations) of manager A and manager B, were 6.3 (0.7) and 3.1 (1.2), respectively. Under the “Like manager B” condition, the mean likeability ratings (standard deviations) of manager A and manager B were 2.6 (1.1) and 6.1 (0.5), respectively. A one-way ANOVA was performed using manager likeability as the independent variable and likeability ratings of manager A as the dependent measure. Manager likeability condition was significant ( $F = 406.42, p < .0001$ ) and the means were in the expected direction. A second one-way ANOVA was performed using manager likeability as the independent variable and likeability ratings of manager B as the dependent measure. Manager likeability condition was significant ( $F = 242.63, p < .0001$ ) and the means were in the expected direction. Overall, these results indicate that the manipulation of manager likeability was successful.

### **Hypothesis Testing**

In the context of overall performance evaluation, hypothesis one (a) predicts that performance measure organization moderates the relationship between division manager likeability and rating difference scores. A 2x2 analysis of variance (ANOVA) was conducted using the rating difference score as the dependent measure. The independent variables were manager likeability at two levels and performance measure organization at two levels. Our hypothesis would be supported by a significant interaction between manager likeability and performance measure organization. Descriptive statistics for the rating difference scores are presented in Table 3, Panel A. The ANOVA results (see Table 3, Panel B) indicate that manager likeability ( $F = 14.30, p < .001$ ) is significant. As shown in Table 3, Panel A, the rating

difference score was higher in the “Like manager A” condition (Mean = 5.37) than under the “Like manager B” condition (Mean = -1.22). However, neither the performance measure organization main effect nor the interaction between manager likeability and performance measure organization is significant. Thus, hypothesis one (a), predicting a significant interaction, is not supported. Instead, likeability significantly influences performance ratings and the BSC organization does not appear to mitigate the effects of likeability on performance ratings.

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

For strength of conviction judgments, hypothesis one (b) predicts that performance measure organization moderates the relationship between manager likeability and conviction difference scores. A 2x2 analysis of variance (ANOVA) was conducted using the conviction difference score as the dependent measure. The independent variables were manager likeability at two levels and performance measure organization at two levels. Again, our hypothesis would be supported by a significant interaction between manager likeability and performance measure organization. Descriptive statistics for the conviction difference scores are presented in Table 4, Panel A. The ANOVA results (see Table 4, Panel B) indicate that manager likeability ( $F = 18.68$ ,  $p < .001$ ) is significant. As shown in Table 4, Panel A, the conviction difference score was higher in the “Like manager A” condition (Mean = 5.12) than under the “Like manager B” condition (Mean = -2.82). This indicates that strength of conviction is positively associated with manager likeability. However, neither the performance measure organization main effect nor the interaction between manager likeability and performance measure organization is significant. Thus, hypothesis one (b), predicting a significant interaction, is not supported.

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

-----

In the context of compensation, hypothesis one (c) predicts that performance measure organization moderates the relationship between division manager likeability and allocation of bonus. A 2x2 analysis of variance (ANOVA) was conducted using the bonus allocated to manager A as the dependent measure. The independent variables were manager likeability at two levels and performance measure organization at two levels. Descriptive statistics for the bonus assigned to manager A are presented in Table 5, Panel A. The ANOVA results (see Table 5, Panel B) indicate that manager likeability is significant ( $F = 18.77, p < .001$ ). As shown in Table 5 Panel A, the bonus assigned to manager A is higher in the “Like manager A” condition (Mean = \$10,929) than in the “Like manager B” condition (Mean = \$9,556). However, neither the main effect for performance measure organization nor the interaction between performance measure organization and manager likeability is significant. Thus, hypothesis one (c), predicting a significant interaction, is not supported.

-----

Insert Table 5 about here

-----

### **Supplemental Analysis**

In this supplemental analysis, we seek to determine whether rating difference scores mediate the effect of manager likeability on compensation. Mediating variables do not change the relationship between other variables, but increase one’s understanding of the process by which evaluators “transform the predictor or input variables,” (Baron and Kenny 1986, p. 1178). Baron and Kenny (1986) assert that several relationships must hold to demonstrate mediation. In the context of our study, an independent variable (subordinate manager likeability) must be found to have a direct effect on the mediator (rating difference scores) as well as the dependent variable

(manager A bonus). Both of these effects have been found. In addition, there must be a direct relationship between the mediator (rating difference score) and the dependent variable (manager A bonus). The correlation between these two variables is positive and highly significant ( $r = 0.75$ ,  $p < 0.0001$ ). Finally, when the mediator (rating difference score) is controlled for, the effect of the independent variable (manager likeability) on the dependent variable (manager A's bonus) should be eliminated or substantially lowered. To test this final aspect of mediation, analysis-of-covariance was conducted with manager A's bonus as the dependent variable and the rating difference score as a covariate in addition to the two independent variables. The results show that the rating difference score is highly significant ( $F = 103.2$ ,  $p < 0.0001$ ) and that while manager likeability remains significant ( $F = 4.6$ ,  $p > 0.04$ ), the F-value declines and the p-value increases. The overall pattern of results indicates that rating difference scores partially mediate the relationship between manager likeability and compensation. However, even with the rating difference score included in the ANCOVA, likeability is still significant. This suggests that manager likeability has a direct effect on compensation over and above the indirect effect resulting from elevated performance evaluations.

## **DISCUSSION**

In a multidimensional performance measurement environment, this study examined the role of the BSC format in mitigating the effect of subordinate manager likeability on overall performance ratings and compensation judgments. While previous appraisal research (Lefkowitz 2000) has found that subordinate manager likeability biases appraisal-related judgments, research has not explored the extent to which the format of the instrument may overcome this bias. We expected that the structure of the BSC would induce a divide and conquer processing strategy, which in

turn, would focus additional attention on affect-inconsistent performance measures and lessen the influence of subordinate likeability

Before discussing the results of the study, several limitations should be noted. First, the participants in the study were evening MBA students asked to evaluate the overall performance and allocate bonuses for two divisional managers. Generally, participants had substantial professional work experience and the majority had previously evaluated the work of others. Our participants may have limited experience using the BSC as an evaluation tool. However, as Lipe and Salterio (2000, p. 296) note, in explaining their decision to employ graduate business students as participants to examine real world phenomenon related to the BSC, “the theory does not suggest an optimal choice,” and consequently “it can be difficult to identify appropriate participants for the experiments.”

Secondly, the participants in this study did not actually know the subordinates, and thus, subordinate manager likeability was created through the use of descriptions of the subordinate managers. While such an approach has been used by appraisal researchers (Cardy and Dobbins 1986; Dobbins and Russell 1986; Robbins and DeNesi 1994), this may represent a relatively weak “treatment.” The advantage of this approach is that it allows one to study the influence of subordinate manager likeability during the evaluation process without being confounded by information an evaluator learned through previous interactions with the subordinate. In addition, by including targets for each performance measure, participants were presumably prevented from creating a favorable benchmark (e.g., creating a relatively low benchmark for a likeable subordinate and creating a relatively high benchmark for an unlikeable subordinate). This feature is in contrast to earlier managerial affect studies (Cardy and Dobbins

1986; Dobbins and Russell 1986; Robbins and DeNesi 1994) in which performance targets were not provided.

Turning to a discussion of the results, first, the results indicate that appraisal related judgments were not associated with the use of the BSC format to present and structure performance information. Our findings of a significant main effect of subordinate manager likeability on appraisal related judgments are consistent with previous appraisal research (Cardy and Dobbins 1986; Tsui and Barry 1986; Turban et al. 1990). However, our findings extend previous results to the BSC setting, which has not previously been considered by appraisal researchers. This extension is important for two reasons. First, the BSC organization should allow for easier processing of the measures and make it more difficult to ignore affect inconsistent measures. Second, the BSC setting typically includes performance targets, in contrast to previous evaluation research (Cardy and Dobbins 1986; Tsui and Barry 1986; Turban et al. 1990), which did not include performance targets. These findings suggest that it may not be possible to design and format ratings instruments to effectively overcome biases in the evaluation process associated with subordinate manager likeability. Further, our findings extend previous results by showing that subordinate manager likeability also influences strength of conviction judgments. Thus, evaluators not only evaluate managers who are likeable more favorably they are also more confident in these evaluations.

The results from the study did not conform to our expectations. We expected subordinate manager likeability would have less effect on appraisal related judgments under a BSC format. We conjecture that the BSC format did not mitigate the effect for manager likeability for either of two potential reasons. First, even though evaluators presumably used a divide and conquer heuristic to simplify the evaluation task, and this process was guided by the BSC, evaluators still

may have used an affect consistency heuristic to direct attention and evaluation among measures within each BSC category. That is, within category, combining measures within BSC remains a subjective task and subordinate manager likeability may have influenced this weighting process within category. Alternatively, it is possible that subordinate manager likeability is simply used to inflate ratings, regardless of the underlying performance measures. However, since the processing of performance measures would not be altered if likeability directly inflates evaluation judgments, it is not clear why strength of conviction judgments are significantly associated with subordinate manager likeability.

Second, we find that while performance ratings partially mediate the effect of likeability on the bonus allocated to each manager, likeability also influences bonus allocations beyond its direct effect on ratings. That is, evaluators allocate larger bonuses to more likeable managers above and beyond that justified by the higher performance evaluations. If encouraging managers to direct their efforts and resources towards the achievement of the entire set of multidimensional performance measures is in a company's best interest, then affect-based biases in rating and compensation judgments are sub-optimal. Presumably, over time, subordinate managers would recognize the importance of likeability in these judgments and disproportionately direct their time and resources towards impression management by specifically trying to create a likeable persona.

Finally, it is worth noting that our findings parallel those of Lipe and Salterio (2000), who found that the cognitive processing biases, such as the common-measures bias, persist using a BSC format. Subsequently, several approaches have been found to mitigate this bias. To the extent that the common-measures bias and the affect consistency bias are tied to one's information processing of multidimensional performance measures, approaches to mitigate the common measures bias presumably would also mitigate the bias due to affect consistency. Further

research is encouraged to explore the extent to which strategy maps (Banker et al. 2004; a “disaggregated BSC” (Roberts et al. 2004), or process accountability or third party assurance (Libby et al. 2004) serve to mitigate the influence of subordinate manager likeability on appraisal judgments of subordinates.

**TABLE 1****The Balanced Scorecard Employed in the Experiment**

Measure	Chris Peters – Division A			Taylor Graham – Division B		
	Target	Actual	% better than target	Target	Actual	% better than target
<i>Financial:</i>						
1. Return on sales	24%	26%	8.33%	24%	25%	4.17%
2. New store sales	30%	30%	0.00%	30%	31.2%	4.00%
3. Sales growth	34%	35.5%	4.41%	34%	36.5%	7.35%
4. Monetary market share relative to retail space	\$80	\$84.70	5.88%	\$80	\$82.55	3.19%
<i>Customer-Related:</i>						
1. Mystery shopper program rating	85	90	5.88%	85	92	8.24%
2. Repeat sales	25%	27%	8.00%	25%	26%	4.00%
3. Returns by customers as % of sales	12%	12%	0.00%	12%	11.5%	4.17%
4. Customer satisfaction rating	84%	86.2%	2.62%	84%	84%	0.00%
<i>Internal Business Processes:</i>						
1. Returns to suppliers	8%	7.7%	3.75%	8%	7%	12.50%
2. Average major brand names / store	32	34	6.25%	32	33	3.13%
3. Average markdowns	20%	18.5%	7.50%	20%	20%	0.00%
4. Sales from new market leaders	25%	25.6%	2.40%	25%	26.1%	4.40%
<i>Learning and Growth:</i>						
1. Stores computerizing	85%	88%	3.53%	85%	86%	1.18%
2. Hours of employee training / employee	12	13	8.33%	12	12	0.00%
3. Average tenure of sales personnel	1.4	1.5	7.14%	1.4	1.6	14.29%
4. Employee suggestions per employee	3.1	3.1	0.00%	3.1	3.2	3.22%

**TABLE 2**  
**Participant Demographic Data**  
**(N = 99)**

**Gender:**

Male - 75 participants (75.8%)  
Female- 24 participants (24.2%)

**Any evaluation experience:**

Yes – 68 participants (68.7%)  
No – 31 participants (31.3%)

**Age:**

Mean = 28.6 years  
Standard deviation = 5.0

**Years of Work Experience:**

Mean = 5.9 years  
Standard deviation = 5.3

**Number of Accounting Courses:**

Mean = 3.7 accounting courses  
Standard deviation = 2.0

**TABLE 3****Analysis of Manager Performance Ratings***Panel A: Rating difference score<sup>a</sup>*

Manager Likeability <sup>c</sup>	Performance measure organization <sup>b</sup>		Overall
	BSC Format	NOFORM Format	
Like Manager A	7.41 <sup>d</sup> (9.18)	3.70 (8.99)	5.37 (9.18)
Like Manager B	-0.89 (9.50)	-1.61 (7.78)	-1.22 (8.67)
Overall	2.84 (10.16)	1.26 (8.79)	

*Panel B: ANOVA with rating difference score as the dependent measure<sup>a</sup>*

Sources of Variation	SS	Df	MS	F	p <sup>e</sup>
Performance measure organization (PMO)	120.1	1	120.1	1.51	0.22
Manager Likeability (ML)	1,136.4	1	1,136.4	14.30	< 0.00
PMO x ML	54.7	1	54.7	0.69	0.41
Error	7,551.1	95	79.5		

<sup>a</sup> Rating difference score is calculated by subtracting the overall performance rating of division manager B from the overall performance rating of division manager A. A positive value indicates that division manager A was rated higher than division manager B. Alternatively, a negative value indicates that division manager B was rated higher than division manager A.

<sup>b</sup> Performance measure organization is a between subjects factor that was manipulated. In the BSC format condition, the sixteen performance measures were classified into the four BSC categories. In the NOFORM format condition, the sixteen performance measures were presented without the BSC format.

<sup>c</sup> Manager likeability is a between subjects factor that was manipulated. In the Like Manager A Condition, positive attributes were associated with manager A and negative attributes were associated with manager B. In the Like manager B condition, negative attributes were associated with manager A and positive attributes were associated with manager B.

<sup>d</sup> Panel values are the mean rating difference scores. Standard deviations are shown in parentheses below the means.

<sup>e</sup> All reported p-values are two-tailed.

**TABLE 4**

**Analysis of Strength of Conviction in Performance Rating**

*Panel A: Conviction Difference Score<sup>a</sup>*

	Performance Measure Organization		
	BSC	NOFORM	Overall
Manager Likeability	Format	Format	Overall
Like Manager A	5.27 <sup>d</sup> (10.17)	5.00 (11.51)	5.12 (10.82)
Like Manager B	-1.67 (6.81)	-4.17 (7.63)	-2.82 (7.23)
Overall	1.45 (9.08)	0.78 (10.85)	

*Panel B: ANOVA with conviction difference score as the dependent measure<sup>a</sup>*

<i>Sources of Variation</i>	<i>SS</i>	<i>Df</i>	<i>MS</i>	<i>F</i>	<i>p<sup>e</sup></i>
Performance measure organization (PMO)	47.41	1	47.41	0.56	0.46
Manager Likeability (ML)	1,592.81	1	1592.81	18.68	0.00
PMO X ML	30.63	1	30.63	0.36	0.55
Error	8,101.67	95	85.28		

<sup>a</sup> Conviction difference score is calculated by subtracting the strength of conviction in the performance rating of division manager B from the strength of conviction in the performance rating of division manager A. A positive value indicates that the strength of conviction in division manager A's rating was higher than the strength of conviction in division manager B's rating. Alternatively, a negative value indicates that the strength of conviction in division manager B's rating was higher than the strength of conviction in division manager A's rating.

<sup>b</sup> Performance measure organization is a between subjects factor that was manipulated. In the BSC format condition, the sixteen performance measures were classified into the four BSC categories. In the NOFORM format condition, the sixteen performance measures were presented without the BSC format.

<sup>c</sup> Manager likeability is a between subjects factor that was manipulated. In the Like Manager A Condition, positive attributes were associated with manager A and negative attributes were associated with manager B. In the Like manager B condition, negative attributes were associated with manager A and positive attributes were associated with manager B.

<sup>d</sup> Panel values are the mean conviction difference score. Standard deviations are shown in parentheses below the means.

<sup>e</sup> All reported p-values are two-tailed.

**TABLE 5****Analysis of Bonus Allocation***Panel A: Manager A Bonus<sup>a</sup>*

<u>Manager Likeability<sup>c</sup></u>	<u>Performance Measure Organization<sup>b</sup></u>		<u>Overall</u>
	<u>BSC Format</u>	<u>NOFORM Format</u>	
Like Manager A	\$11,341 <sup>d</sup> (1,409)	\$10,593 (1,494)	\$10,929 (1,489)
Like Manager B	\$9,519 (1,968)	\$9,599 (1,443)	\$9,556 (1,730)
Overall	\$10,337 (1,951)	\$10,136 (1,539)	

*Panel B: ANOVA with manager A bonus as the dependent measure<sup>a</sup>*

<u>Sources of Variation</u>	<u>SS</u>	<u>Df</u>	<u>MS</u>	<u>F</u>	<u>p<sup>e</sup></u>
Performance measure organization (PMO)	2,732,001	1	2,732,001	1.05	0.31
Manager Likeability (ML)	48,655,078	1	48,655,078	18.77	0.00
PMO X ML	4,211,530	1	4,211,530	1.62	0.21
Error	246,245,068	95	2,592,053		

<sup>a</sup> Participants allocated a \$20,000 bonus between manager A and manager B. Manager A bonus is the amount of the bonus that participants allocated to manager A. Because manager B bonus is a function of manager A bonus, it is not included in these tests.

<sup>b</sup> Performance measure organization is a between subjects factor that was manipulated. In the BSC format condition, the sixteen performance measures were classified into the four BSC categories. In the NOFORM format condition, the sixteen performance measures were presented without the BSC format.

<sup>c</sup> Manager likeability is a between subjects factor that was manipulated. In the Like Manager A Condition, positive attributes were associated with manager A and negative attributes were associated with manager B. In the Like manager B condition, negative attributes were associated with manager A and positive attributes were associated with manager B.

<sup>d</sup> Panel values are the mean bonus allocated to manager A. Standard deviations are shown in parentheses below the means.

<sup>e</sup> All reported p-values are two-tailed.

## REFERENCES

- Alexander, E. R., and R. D. Wilkins. 1982. Performance rating validity: The relationship between objective and subjective measures of performance. *Group and Organization Studies* 7(4): 485-496.
- Anderson, N. H. 1968. Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology* 9(3): 272-279.
- Antonioni, D., and H. Park. 2001. The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management* 27(4): 479-495.
- Baddeley, A. 1994. The magical number seven: Still magic after all these years? *Psychological Review* 101(2): 353-356.
- Balzer, W. K. 1986. Biases in the recording of performance-related information: The effects of initial impression and centrality of the appraisal task. *Organizational Behavior and Human Decision Processes* 37(3): 329-347.
- Banker, R. D., H. Chang, and M. J. Pizzini. 2004. The balanced scorecard: Judgmental effects of performance measures linked to strategy. *The Accounting Review* 79 (1): 1-23.
- Baron, R. M., and D. A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51(6): 1173-1182.
- Baltes, B. B. and C. P. Parker. 2000. Reducing the effects of performance expectations on behavioral ratings. *Organizational Behavior and Human Decision Processes* 82(2): 237-267.
- Bates, R. 2002. Liking and similarity as predictors of multi-source ratings. *Personnel Review* 31(5/6): 540-552.
- Bonner, S. 1994. A model of the effects of audit task complexity. *Accounting, Organizations and Society*. 19(3): 213-234.
- Cardy, R. L., and G. H. Dobbins. 1986. Affect and appraisal accuracy. Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology* 71(4): 672-678.
- Cardy, R. L., and G. H. Dobbins. 1994. Performance appraisal: The influence of liking on cognition. *Advances in Managerial Cognition and Organizational Information Processing* 5: 115-140.
- DeNisi, A. S., T. L. Robbins and T. P. Summers. 1997. Organization, processing, and use of performance information: A cognitive role for performance instruments. *Journal of Applied Social Psychology* 27(21): 1884-1905.

- Dilla, W. N., and P. J. Steinbart. 2005. Relative weighting of common and unique balanced scorecard measures by knowledgeable decision makers. *Behavioral Research in Accounting* 17: 43-53.
- Dipboye, R. L. 1985. Some neglected variables in research on discrimination in appraisal. *Academy of Management Review* 10(1): 116-127.
- Dobbins, G. H., and J. M. Russell. 1986. The biasing effects of subordinate likeableness on leaders' responses to poor performance: A laboratory and a field study. *Personnel Psychology* 39(4): 759-777.
- Ittner, C. D., D. F. Larcker, and M. W. Meyer. 2003. Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review* 78(3): 725-758.
- Kaplan, R., and D. Norton. 1992. The balanced scorecard – Measures that drive performance. *Harvard Business Review* 70(1): 71-79.
- Kaplan, R., and D. Norton. 1996a. *The Balanced Scorecard*. Boston, MA: Harvard Business School Press.
- Kaplan, R., and D. Norton. 1996b. Using the balanced scorecard as a strategic management system. *Harvard Business Review* 74(1): 75-85.
- Kida, T. E., K. K. Moreno, and J. F. Smith. 2001. The influence of affect on managers' capital-budgeting decisions. *Contemporary Accounting Research* 18(3): 477-494.
- Kida, T. E., and J. F. Smith. 1995. The encoding and retrieval of numerical data for decision making in accounting contexts: Model development. *Accounting, Organizations, and Society* 20(7/8): 585-610.
- Krishnan, R., J. L. Luft, and M. D. Shields. 2005. Effects of accounting-method choices and subjective performance-measure weighting decisions: Experimental evidence on precision and error covariance. *The Accounting Review* 80(4): 1163-1192.
- Kunda, Z. 1990. The case for motivated reasoning. *Psychological Bulletin* 108(3): 480-498.
- Lefkowitz, J. 2000. The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. *Journal of Occupational and Organizational Psychology* 73(1): 67-85.
- Libby, T., S. E. Salterio and A. Webb. 2004. The Balanced Scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review* 79(4): 1075-1094.

- Lipe, M. G., and S. Salterio. 2000. The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review* 75 (3): 283-298.
- Lipe, M. G., and S. Salterio. 2002. A note on the judgmental effects of the balanced scorecard's information organization. *Accounting, Organization and Society* 27(6): 531-540.
- Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81-97.
- Moreno, K., T. Kida., and J. F. Smith. 2002. The impact of affective reactions on risky decision making in accounting contexts. *Journal of Accounting Research* 40(5): 1331-1349.
- Murphy, K. R., and J. N. Cleveland. 1991. *Performance appraisal: An organizational perspective*. Boston, MA: Allyn and Bacon.
- Murphy, K. R., B. A. Gannett, B. M. Herr and J. A. Chen. 1986. Effects of subsequent performance on evaluations of previous performance. *Journal of Applied Psychology* 71(3): 427-431.
- Park, O. S., H. P. Sims, and S. J. Motowildo. 1987. Affect in organizations. In H. P. Sims and D. A. Gioia (Eds.), *The Thinking Organization*. San Francisco: Josey-Bass.
- Roberts, M. L., T. L. Albright, and A. R. Hibbets. 2004. Debiasing balanced scorecard evaluations. *Behavioral Research in Accounting* 16: 75-88.
- Robbins, T. L. and A. S. DeNisi. 1994. Interpersonal affect and cognitive processing in performance appraisal: Towards closing the gap. *Journal of Applied Psychology* 79(3): 341-350.
- Robbins, T. L. and A. S. DeNisi. 1998. Mood vs. interpersonal affect: Identifying process and rating distortions in performance appraisal. *Journal of Business and Psychology* 12(3): 313-325.
- Shanteau, J. 1988. Psychological characteristics and strategies of expert decision makers. *Acta Psychologica* 68(1-3): 203-215.
- Tan, H. T., and J. F. Yates. 1995. Sunk cost effects: The influences of instruction and future return estimates. *Organizational Behavior and Human Decision Processes* 63(3): 311-319.
- Tsui, A. S. and B. Barry. 1986. Interpersonal affect and rating errors. *Academy of Management Journal* 29(3): 586-599.
- Turban, D. B., A. P. Jones, and R. M. Rozelle. 1990. Influences of supervisor liking of a subordinate and the reward context on the treatment and evaluation of that subordinate. *Motivation and Emotion* 14 (3): 215-233.

- Varma, A., A. S. Denisi, and L. H. Peters. 1996. Interpersonal affect and performance appraisal. *Personnel Psychology* 49(2): 341-360.
- Wood, R. E. 1986. Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes* 37(1): 60-82.
- Yates, J. F. 1990. *Judgment and Decision Making*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Zajonc, R. B. 1980. Feeling and thinking: Preferences need no inferences. *American Psychologist* 35(2): 151-175.

## ENDNOTES

- 
- <sup>1</sup> Similar to Lipe and Salterio (2002), we included both a random and an alphabetical order to increase the generalizability of the results. For the random order, the order of the measures was chosen at random with the only constraint that adjacent measures should not come from the same BSC category. Within the NOFORM group, no differences were found for participants with the random versus the alphabetical order for any of the outcome measures. Therefore, these two conditions are collapsed together and analyzed as one condition.
- <sup>2</sup> Previous research investigating the effect of likeability on performance evaluations has also used this source (Cardy and Dobbins 1986; Dobbins and Russell 1986). We did not use identical personal attributes as previous research for two reasons. First, these previous articles only gave partial lists of the personal attribute words they used in their manipulations. Second, some of the previous articles used words that, in today's business environment, may have other behavioral implications (e.g. the dislikeable word 'greedy' may also invoke unethical implications).
- <sup>3</sup> We ran tests of H1(a) through H1(c) using all participants including those that failed the manipulation checks. All significant results remain significant (at  $p < .01$ ) and insignificant results remain insignificant.