

Missing non-strategic targets: Are managers penalized for emphasizing important measures?

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

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

October 2006

This paper has benefited from the helpful comments of Steve Kaplan and the workshop participants at Arizona State University – Tempe Campus and Arizona State University – West Campus.

Missing non-strategic targets: Are managers penalized for emphasizing important measures?

ABSTRACT

Surveys have found that, while the balanced scorecard (BSC) was designed to focus on strategic measures, many companies using a BSC include both strategically linked and non-strategically linked measures in their BSC. Strategy maps were developed to direct effort toward strategic goals. Strategy maps have been found to influence evaluation ratings in a BSC setting such that strategically linked measures are weighted more than non-strategically linked measures (Banker et al 2004). However, the effectiveness of strategy maps has not been tested in connection with below-target performance. Below-target performance is expected to trigger a negativity bias. We hypothesize that strategy maps do not mitigate the negativity bias. Therefore, evaluators will penalize managers with below-target performance even when the negative performance is on non-strategically linked measures and the manager greatly outperforms strategically linked targets. We use an experiment to test our hypotheses and find that when managers outperform all targets, a manager that dominates on strategically-linked measures is rated more highly than a manager who dominates on non-strategically-lined measures. However, these results don't hold when a manager misses targets. Instead, a manager with some below-target performance is rated more negatively than a manager with linearly equivalent total performance who slightly outperforms on all targets even when the manager with negative performance misses non-strategically linked measures and greatly outperforms strategically linked measures. Kaplan and Norton (2001) indicate that a key feature of the BSC is linking performance measures to strategy. While Banker et al (2004) found that evaluators could use a strategy map to focus on strategically linked measures, this study demonstrates that the effectiveness of strategy maps is

limited when managers have some negative performance. Our results support the viewpoint espoused by Kaplan and Norton (2001) that only strategically linked measures should be included on a BSC.

I. INTRODUCTION

Multi-attribute performance measurement systems such as the Balanced Scorecard (BSC) are increasingly used in business (Ittner and Larcker 2003; Silk 1998). While Kaplan and Norton (1996, 2000) suggest that a BSC should only contain measures linked to the organization's strategy, research has found that companies using a BSC are likely to include some measures that are more important to the organization's strategy (e.g. strategically linked measures) and others that are less important to the organization's strategy (e.g. non-strategically linked measures) (Reisinger, et al. 2003, Ittner and Larcker, 2003). Banker et al. (2004) advocate the use of strategy maps as a tool to communicate and direct managers' efforts towards the activities that are most critical to organization's success and away from those activities that are less critical to the organization's success. Banker et al., (2004) found that when strategy maps are available, individuals are able to identify strategically linked ("SL") measures as more important to the organization and, accordingly, weight those measures greater than non-strategically linked ("NSL") measures when evaluating performance. However, Banker et al., (2004) only consider the role of strategy maps to direct attention towards SL measures in situations where managers outperformed their targets on all measures.¹ In a real-life situation, managers constrained by time and resources may under-perform on some measures. This may be especially the case when managers devote time and attention toward achieving or exceeding targets on SL measures and therefore under-perform on less important (or NSL) measures.

In this study we propose that, even with the presence of strategy maps, managers who miss targets on less important measures while substantially outperforming targets on more important measures will be penalized relative to managers who slightly outperform on all targets.

¹ We are aware of two BSC articles that include at least one condition with negative performance. Lipe and Salterio (2002) & Haywood and Stuart (2006). However, these studies did not use strategy maps, did not have measures of varied importance, and did not directly test the effect of negative measures.

Penalizing managers who disproportionately direct their efforts towards achieving and/or exceeding SL measures may be dysfunctional for the organization. Managers are unlikely to continue to devote time and resources toward maximizing critical measures when performance evaluations do not align with the organization's critical performance measures.

Prior psychology research has documented a “negativity bias” wherein individuals overweight negative performance (Skowronski and Carlston 1989; Kanouse and Hanson 1972). We contend that strategy maps will not be sufficient to mitigate this negativity bias. That is, while strategy maps signal what measures are more/less important, we expect individuals to discount this information when managers do not meet some of their targets. We hypothesize that, compared to a manager with all small above-target performance, a manager with linearly equivalent total performance, but with both above- and below-target results, will be rated lower regardless of the strategic importance of the various below-target measures.

We gave all participants a strategy map and information on the performance of Manager A and Manager B. The BSC consisted of 16 measures with eight SL measures and eight NSL measures. Manager B slightly outperformed targets on all sixteen measures for all conditions. We manipulated the measures on which Manager A dominated (SL or NSL). We also manipulated whether Manager A underperformed on some measures (negativity absent or present).² We analyzed participants' evaluation ratings of the two managers and find that SL measures are weighted more heavily than NSL measures when both managers outperform their targets. However, a manager that underperforms on some targets is rated lower than the manager who slightly outperforms on all targets – even when the under-performing manager greatly outperforms on SL measures.

² “Underperformed” refers to performance that is below an established target. “Outperformed” refers to performance that is above an established target.

This research has implications for practice and theory. From a practice standpoint, this study finds that including NSL measures on the BSC could be dysfunctional to the organization. Managers often have limited time available to them. A purpose of the BSC is to focus managers' efforts on measures and actions which are SL to the organization as these measures are important to the financial success of the organization. However, if a BSC includes both SL and NSL measures and managers choose to focus their efforts on the SL measures they may miss NSL measures. As managers become aware of the negative effects of missing targets, managers will expend effort toward NSL measures at the expense of potentially maximizing the SL measures. Surveys have found that companies often include non-strategic measures on their BSC (Reisinger et al 2003; Ittner and Larcker 2003) and Banker et al (2004) showed the importance of strategy maps in focusing attention on the SL measures. However, this study demonstrates the potential problems associated with including NSL measures in the BSC. This study also contributes to the theoretical literature by demonstrating that the perceived importance of the measures moderates but does not eliminate the negativity bias.

The remainder of this paper is organized as follows. Section II develops the hypotheses. Section III describes the design and method for Experiment 1. Section IV presents the results of Experiment 1 and the design, method and results of Experiment 2. Section V summarizes and concludes.

II. BACKGROUND AND HYPOTHESIS DEVELOPMENT

Background

The BSC was developed to supplement the use of a single financial measure of performance with multiple financial and non-financial measures (Kaplan and Norton 1992). The BSC contains financial and non-financial measures from four perspectives: financial, customer, internal

process and learning and growth. However, other multi-attribute performance measurement systems exist that use a variety of financial and non-financial measures (e.g. KPI Scorecards, the Tableaux de Bord, etc). Kaplan and Norton (2001) note that a key difference from these other performance measurement systems is that the BSC links measures to strategy. In fact, the measures on a BSC “are derived from an organization’s vision and strategy” (Kaplan and Norton 1996, p8) and “every measure selected for a Balanced Scorecard should be an element of a chain of cause-and-effect relationships that communicates the meaning of the business unit’s strategy to the organization” (p149).

Research has investigated whether SL measures influence firm performance. Ittner and Larcker (1998) found that customer satisfaction, which was identified as a key strategic measure for the firm under study, was directly correlated with accounting performance. Banker et al (2000) document that SL nonfinancial measures of customer satisfaction are not only correlated with future financial performance, they contain additional information not reflected in past financial measures. Therefore, this prior research finds that strategic measures are positively associated with firm performance. This suggests that identifying SL measures for the BSC, as advocated by Kaplan and Norton (1996), can help firm performance by focusing managers’ efforts towards actions that result in superior performance on the SL measures.

However, while BSC advocates indicate that measures should be tied to strategy, surveys have found that companies may actually use some combination of strategic and non-strategic measures in their BSC (Reisinger, et al. 2003; Ittner and Larcker, 2003). For example, Reisinger, et al. (2003, 431) note that companies include many measures on a BSC, but “(t)here may be a need to emphasize certain activities over others and thus focus on a different subset of measures.” Therefore, some of the BSC measures may not actually be linked to the company’s

current strategy. Ittner and Larcker (2003,90) discuss survey results indicating that many managers, faced with the enormous task of choosing a few non-financial measures from the “hundreds, if not thousands” available, fall back on ‘boilerplate’ measures, which may or may not be linked to company strategy. Bryant et al. (2004) cite survey evidence that managers tend to use generic measures in their BSC systems or measures that are corporate-level or control-related measures. Therefore, in contrast to the original concept espoused by Kaplan and Norton (1992), a BSC may include measures that are NSL to the business’s objectives as well as SL measures.

Including SL and NSL measures in the same BSC may not be a problem if evaluators can differentiate between these measures and reward managers for performing on the SL measures. Banker et al. (2004) indicate that SL measures should be assigned greater weight as compared to NSL measures. However, they find that participants don’t weight SL measures more than NSL measures unless they are given strategy maps. Strategy maps are a graphical representation of an organization’s linkages between various activities and performance measures (Banker et al. 2004; Kaplan and Norton 2000; Oliva et al 1987). Cognitive research suggests that graphical representation makes it easier for individuals to understand linkages among the data (Glenberg and Langston 1992). Banker et al. (2004) suggest that using strategy maps will increase the “perceived relevance of strategically linked measures, and consequently their use in performance evaluation” (pg 4). At the same time, a strategy map should allow individuals to “recognize that non-linked measures are of little value and, therefore, they will rely less on them in evaluating performance” (Banker et al. 2004, p4). Therefore, individuals with an understanding of a company’s strategy and a strategy map should perceive the SL measures as more important and assign greater weight to these measures compared to NSL measures. Banker et al (2004) provide

evidence that, when strategy maps are used to help evaluators understand strategic linkages, evaluators will weight SL measures more heavily than NSL measures. This research considered situations where managers outperformed all of their targets. Consistent with Banker et al (2004), we propose the following hypothesis:

- H1: When both managers outperform all of their targets, a manager who dominates on SL measures will be evaluated higher than a linearly-equivalent manager who dominates on NSL measures.

However, research on strategy maps has only considered situations where managers outperformed their targets on all measures (Banker et al. 2004; Banker et al. 2006). In real life situations, it is likely that managers will miss their targets for at least some of their measures. While strategy maps focus attention on SL measures, missing targets may be viewed as negative performance. We suggest that this indicator of negative performance will adversely effect the evaluation regardless of whether the negative measures are SL or NSL.

Negativity Bias

Negative information is given more weight than positive information in a wide range of contexts. For example, in impression formation, negative traits are given more weight compared to positive traits (Peeters and Czapinski 1990). In risky decision making, individuals are loss averse and a one-unit loss is weighted more than an equal amount of gain (Kahneman and Tversky 1979; Einhorn and Hogarth 1981). Also, taxpayers make different ‘marginal’ decisions if they are in a ‘loss’ position as compared to a ‘gain’ position (Schepanski and Shearer 1997). Psychology literature refers to this general bias as the *negativity bias* (Skowronski and Carlston 1989; Kanouse and Hanson 1972). This bias is defined as the tendency for a unit of activation to result in a greater change in output in the negative motivational system as compared with the positive motivational system (Cacioppo et al. 1997). Skowronski and Carlston (1989, 131) state that “(i)n empirical tests of the equal-weight averaging model, researchers discovered some

systematic differences ... when negative or extreme information was included in the stimulus set. ... Mathematically speaking, negative and extreme cues appeared to have greater weight than did moderate and positive cues.”

Psychology researchers have suggested that the negativity bias operates during both intentional and automatic information processing (Pratto and John 1991). Recent psychology reviews have indicated that negative information influences intentional information processing by increasing the amount of conscious attention directed towards negative information and increasing recall of negative information (Baumeister et al 2001; Rozin and Roysman 2001; Ito et al 1998). Negative information also influences intentional information processing via the construction of more elaborate cognitive interpretations of negative information (Rozin and Roysman 2001; Baumeister et al 2001).

In addition to the intentional information processing effects of negative information, psychology research has indicated that there is an automatic or subconscious component to the negativity bias. Prior research has found that negative stimuli are identified more rapidly and attract more automatic attention compared to positive stimuli (Pratto and John 1991; Smith et al 2006). Negative information has also been found to be more perceptually salient and elicit a stronger physiological response than positive information (Peeters and Czapinski 1990). Consistent with the negativity bias, prior performance evaluation experiments and field studies have found that negative information is given greater weight in overall evaluations than is positive information (Ito et al. 1998; Ganzach 1995). However, this prior research has not included situations where some measures are perceived as more or less important than other measures. For example, in a BSC setting with both SL and NSL measures, the SL measures are perceived as more important than the NSL measures.

While Banker et al (2004, 2006) found that strategy maps focus evaluators' attention on the more important (SL linked) measures, they only considered situations with all positive performance. For evaluators to focus on SL measures, they must cognitively process the information and consciously weight the SL measures as more important than the NSL measures. The inclusion of below-target outcomes (negative performance) could change these results. Since the negativity bias has been suggested to be both an intentional and an automatic information processing bias, we expect that the negativity bias will dominate the evaluation process regardless of whether the manager with below-target measures outperforms on SL or NSL measures. Therefore, we propose that a manager who has missed some targets will be evaluated more negatively than a manager who has a linearly equivalent total performance but who has attained all of his/her targets. The below-target performance will be viewed negatively regardless of the importance of the measures. More formally, we propose the following hypothesis:

H2: A manager whose performance includes both below-target and above-target outcomes will be evaluated lower than a linearly-equivalent manager whose performance includes only above-target outcomes, regardless of whether that manager dominates on SL or NSL measures.

The first hypothesis suggests that evaluations will be higher for a manager who dominates on SL measures when both managers outperform all measures. However, when some of the outcomes are negative, hypothesis 2 predicts that the manager with both positive and negative measures will be evaluated more negatively than a linearly equivalent manager with all above-target measures regardless of the importance of the measures. However, evaluation ratings are only one aspect of the evaluation process.

In order to determine if the consequences of this performance evaluation difference is meaningful to the individuals being evaluated, we also consider the economic consequences of the decision. There are economic consequences that we consider: bonus pool allocations and

promotion decisions. First, employees may be given bonuses based on their performance. Typically, monetary bonuses are pooled into a separate account and distributed based on factors such as performance evaluations. We conjecture that performance on SL measures and presence/absence of negative evaluation information will affect the bonus pool allocation in a similar manner as the performance evaluation. More formally, we propose:

- H3a: When both managers outperform their targets, a manager who dominates on SL measures will be allocated a higher bonus compared to a linearly-equivalent manager who dominates on NSL measures.
- H3b: A manager whose performance includes both below-target and above-target outcomes will be allocated a lower bonus than a linearly-equivalent manager whose performance includes only above-target outcomes, regardless of whether that manager dominates on SL or NSL measures.

Another economic consequence of performance evaluations is the decision to promote one manager or another. In this decision, a split decision, or tie vote, cannot be accepted. If only one promotion is available, a manager must be ‘singled out’ for his/her performance. This is similar to a choice task where an evaluator must select one manager or another. Prior negativity research had found that negativity has a stronger effect for choice tasks compared to tasks such as matching or filling in a blank variable to set two alternatives equal to each other. Therefore, we propose that performance on SL measures and the presence/absence of negative performance measures will influence the promotion decision. More formally, we propose:

- H4a: When both managers outperform their targets, a manager who dominates on SL measures will be more likely to be promoted compared to a linearly-equivalent manager who dominates on NSL measures.
- H4b: A manager whose performance includes both below-target and above-target outcomes will be less likely to be promoted than a linearly-equivalent manager whose performance includes only above-target outcomes, regardless of whether that manager dominates on SL or NSL measures.

III. EXPERIMENTAL DESIGN AND METHOD

We examine the effects of below-target performance on evaluation judgments when a manager with below-target performance on some measures greatly outperforms on other measures of differing importance. We use a 2 x 2 between participants design with negativity (present or absent) and dominance (Manager A dominant on SL measures or NSL) as the experimental manipulations. Additionally, there is a within participants design such that each participant evaluated two managers (Manager A and Manager B).

Participants

102 evening MBA students from a major metropolitan university participated in the experiment. The students had recently studied the BSC and strategy maps. Participants' mean age was 30.8 years and mean work experience was 9.0 years. Seventy percent were male and 77 percent had previously evaluated employees.

Case Materials and Procedures

The materials described Smithson Stores, which has several retail chains and many stores. Participants assumed the position of evaluator of two regional managers (Manager A and Manager B) for the Women's Store Division. To avoid potential confounds, both managers had identical strategies, measures and targets. Participants received a description of the strategy of the division, a strategy map that highlighted the strategically linked performance measures and a sample BSC that listed the 16 measures including descriptions and targets. In addition to this, each performance measure in the sample BSC was labeled as 'Strategic' or 'Non-Strategic'. Next, they were given a BSC with the actual performance of managers A and B (see Appendix 1 for an example of the BSC). Participants rated the performance of each manager on a 13 point scale with endpoints of zero (Reassign) and 12 (Excellent). Participants then allocated a \$20,000 bonus pool between managers A and B. Finally, participants were told that an important vice

president job was becoming available and were asked to choose either Manager A or Manager B for promotion. The participants also completed a post-experimental questionnaire.

Dominance

In a between participants manipulation, participants were randomly assigned to one of two dominance conditions. In the SL measures dominant condition, Manager A dominated on the SL measures and Manager B dominated on the NSL measures. In the NSL measures dominant condition, Manager A dominated on the NSL measures and Manager B dominated on the SL measures. In all conditions, Manager A's total performance was linearly equivalent to Manager B's total performance.

Negativity

In a between participants manipulation, participants were randomly assigned to one of two negativity conditions. In the negativity absent condition, both managers outperformed on all targets. In the negativity present condition, Manager A greatly over-performed on the eight dominant-condition measures (SL or NSL) and slightly under-performed on the remaining eight measures. Manager B's performance was slightly above target for all 16 measures. Manager A's total performance was linearly equivalent to Manager B's performance in each condition. As a result, in the negativity present condition, Manager A's performance ranged from negative four percent to positive 15 percent with a total variance of 0.70 percent. Manager B's performance ranged from approximately one percent to 10 percent with a total variance of 0.09 percent. Therefore, Manager A had much greater variance than Manager B. To exclude this difference in variance as an alternative explanation for our results, Manager A had identical variance across both conditions as did Manager B. As a result, in the negativity absent condition, Manager A's performance ranged from one percent to 20 percent with a total variance of 0.70 percent

(identical to the negativity present condition). Manager B's performance ranged from approximately 6 percent to 15 percent with a total variance of 0.09 percent (identical to the negativity present condition).

Dependent Measures

Participants gave each manager a rating from zero (reassign) to 12 (excellent) with a higher number indicating a better evaluation rating. There was a description of the scale that expounded upon the meaning of each label (e.g. Excellent: far beyond expectations, manager excels). The dependent variable for hypotheses one and two was the evaluation rating of Manager A less the evaluation rating of Manager B (referred to as "evaluation difference"). A positive evaluation difference indicates that Manager A was evaluated higher than Manager B while a negative evaluation difference indicates that Manager A was evaluated lower than Manager B.

Participants next allocated a \$20,000 bonus between Manager A and Manager B. Because Manager B's bonus is a function of Manager A's bonus, the dependent variable for hypothesis 3 was the bonus allocated to Manager A. A bonus significantly greater than \$10,000 indicates that Manager A was allocated a greater bonus than Manager B while a bonus significantly less than \$10,000 indicates that Manager A was allocated a lesser bonus than Manager B.

Finally, participants chose one manager to promote to vice-president. The dependent variable for hypothesis four was the percentage of participants that selected Manager A for promotion.

IV. RESULTS

Manipulation Check

After responding to the scenario, participants answered questions to determine whether they were aware of the negativity manipulation. The first question read, “Chris Peters (Region A) exceeded his targets on all performance measures.” Participants were asked to select a response on a seven-point scale where one was “strongly disagree” and seven was “strongly agree”. For participants in the negativity present condition, Chris Peters (Manager A) did have below-target outcomes and participants that responded greater than the midpoint of the scale (the midpoint of four was ‘neither agree nor disagree’) were dropped. For participants in the negativity absent condition, Chris Peters (Manager A) did not have any below-target outcomes and participants that responded less than the midpoint of the scale were dropped. The second question read, “Taylor Graham (Region B) exceeded his targets on all performance measures.” Participants were asked to select a response on a seven-point scale where one was “strongly disagree” and seven was “strongly agree”. For all participants, Taylor Graham (Manager B) did not have any below-target outcomes and responses less than the midpoint of the scale were considered incorrect. Of the 102 participants, 21 responded incorrectly to at least one of the manipulation check questions and were dropped. Participants responding “incorrectly” to the two negativity 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 negativity (or lack thereof) of managers A and B as intended. The statistical analyses that follow are based upon the responses from the remaining 81 participants.³

³ Note that some tests only include 79 or 80 participants as a few of the 81 participants did not respond to all questions.

As part of the post-experimental questionnaire, participants responded to the statement “Strategically linked measures are more important than non-strategically linked measures.” Participants responded on a scale ranging from one (“strongly disagree”) to seven (“strongly agree”). A response of four was neutral (“neither disagree nor agree”). If participants believed that SL measures were more important than NSL measures, they should have responded significantly greater than the neutral response of four. The mean score on this question was 5.3 (std. deviation = 1.52). This is significantly different from the neutral mid-point of four (t-stat = 7.51, p-value < 0.0001). This indicates that, on average, participants believed that SL measures are more important than NSL measures. We also confirmed that the perceived importance of SL measures was not significantly associated with the manipulations (all p-values were > 0.18).

Test of Hypotheses

Hypothesis one predicted that when negativity is absent, a manager that dominates on SL measures would be evaluated higher than a manager that dominates on NSL measures. This hypothesis would be supported by an evaluation difference score that is significantly positive for the SL Dominates condition and significantly negative for the NSL Dominates condition. Evaluation results are summarized in Table 1. T-tests were used to test the significance of the differences. The evaluation difference of 1.15 for the SL Dominates condition is significantly greater than zero (t-stat = 4.52, one-tailed p-value <0.001), while the evaluation difference of -0.73 for the NSL Dominates condition is significantly less than zero (t-stat = -2.30, one-tailed p-value <0.017). These results support hypothesis one.

INSERT TABLE 1 ABOUT HERE

Hypothesis two predicted that, when negativity is present, Manager A would be rated more negatively than Manager B regardless of the importance of the below-target measures. This hypothesis would be supported by significantly negative evaluation difference scores in each Dominance cell for the “Negativity Present” condition. Table 1 summarizes the mean evaluation difference scores for each cell. As predicted, when the NSL measures are dominant and negativity is present for Manager A, the mean evaluation difference was -2.86 which is significantly less than zero (t-statistic of -6.37, one-tailed p-value < 0.0001). When SL measures are dominant and negativity is present, there was a mean evaluation difference of -1.28 which is significantly less than zero (t-statistic of -3.31, one-tailed p-value < 0.003). Therefore, we find that a manager with some below-target measures will be evaluated more negatively than a manager with all small above-target measures across both dominance conditions. This supports hypothesis two. Therefore, based on the previous tests, we confirm that managers that dominate on SL measures are evaluated higher compared to managers that dominate on NSL measures when both managers out-perform all of their targets. However, when one manager misses some targets, that manager is evaluated more negatively than a manager who outperforms all targets regardless of whether they dominated on SL or NSL measures.

Hypothesis three examines the bonus allocation. Table 2 reports the descriptive statistics of the portion of the \$20,000 bonus that was allocated to Manager A. Hypothesis 3a predicts that, when negativity is absent, a manager who dominates on SL measures will be allocated a larger bonus than a manager who dominates on NSL measures. This hypothesis would be supported by a bonus allocation to Manager A significantly greater than \$10,000 when SL measures dominate and significant less than \$10,000 when NSL measures dominate. T-tests were run to test whether the amounts were significantly different from \$10,000. The bonus allocated to Manager

A was \$11,925 for the SL Dominates condition, which is significantly higher than \$10,000 (t-stat = 3.71, one-tailed p-value < 0.0008), while the bonus allocated to Manager A was \$8,868 for the NSL Dominates condition, which is significantly lower than \$10,000 (t-stat = -2.74, one-tailed p-value < 0.006). This lends support for hypothesis 3a. Hypothesis 3b predicts that when negativity is present, a manager with both above- and below-target measures will be allocated a smaller bonus than a linearly-equivalent manager regardless of whether he/she dominates on SL measures. This hypothesis would be supported by a bonus allocation to Manager A significantly less than \$10,000 regardless of the SL / NSL dominates condition. The bonus allocated to Manager A was \$6,762 for the NSL Dominates condition, which is significantly lower than \$10,000 (t-stat = -6.38, one-tailed p-value < 0.001), while the bonus allocated to Manager A was \$9,139 for the SL Dominates condition, which was not significantly lower than \$10,000 (t-stat = -1.16, one-tailed p-value < 0.130). This lends only partial support for hypothesis 3b.

INSERT TABLE 2 ABOUT HERE

The promotion decision is the dependent variables in hypothesis four. The percentage of participants, by cell, that decided to promote Manager A is shown in Table 3. To test out hypotheses, we created a “promotion” variable and assigned a value of “1” if a participant chose to promote manager A and a value of “2” if a participant chose to promote Manager B. Thus, a mean score of 1.5 would indicate that the same number of participants decided to promote Manager A as decided to promote manager B. Hypothesis 4a predicted that, when negativity is absent, significantly more people would decide to promote the manager that dominates on SL measures. Therefore, this hypothesis would be supported by a “promotion” outcome significantly less than 1.5 when SL measures dominate and significantly greater than 1.5 when NSL measures

dominate. We find that when SL measures dominate, the “promotion” outcome was 1.17 which was significantly less than 1.5 (t-statistic -3.69, one-tailed p-value < 0.001) and when NSL measures dominate, the “promotion” outcome was 1.77 which was significantly greater than 1.5 (t-statistic 2.98, one-tailed p-value < 0.004). This supports hypothesis 4a. Hypothesis 4b predicated that, when negativity is present, significantly more people would decide to promote the manager with all above-target measures regardless of the strategic importance of the measures. This would be supported by the “promotion” outcome being significantly greater than 1.5 for both SL dominates and NSL dominates in the negativity present condition. We find that the “promotion” outcome was 1.72 when SL dominates and 1.95 when NSL dominates, both of which are significantly greater than 1.5 (t-statistics of 2.06 and 9.50, respectively, and one-tailed p-values < 0.03 and 0.0001, respectively). Thus, when negativity is absent, we find that a manager that dominates on SL measures is more likely to be promoted than a manager than dominates on NSL measures. However, when negativity is present, a manager with all above-target performance is more likely to be promoted than a manager with some below-target outcomes regardless of the importance of the measures.

INSERT TABLE 3 ABOUT HERE

V. SUMMARY AND DISCUSSION

Banker et al (2004) found that, with strategy maps, evaluators are able to identify SL measures and, in a performance evaluation, assign greater weight to these measures compared to NSL measures. We use an experiment to test whether below-target measures lead evaluators to discount the importance of measures and give unfavorable ratings to managers with below-target performance regardless of the importance of the various measures. The results of our experiment

suggest that, while SL measures are weighted more than NSL measures, managers with some below-target performance measures are rated lower than managers with only small above-target performance. This occurs even when the below-target measures are unimportant and the manager performs extremely well on more important measures. We also demonstrate that are results are not driven by variance.

This study makes several contributions. First, it demonstrates that the effectiveness of strategy maps has boundaries. While strategy maps have been shown to be effective at directing attention toward SL measures, this has only been shown when managers out-perform all targets. When there is below-target or negative performance, strategy maps do not lead to greater weighting of SL measures.

Second, we extend evaluation research in that we rule out variance as an alternate explanation for the negativity bias in an evaluation setting. Also, we find that the negativity bias is robust to situations where some measures are perceived as more important.

Finally, our results indicate that, even with strategy maps, evaluators pay more attention to negative performance and even penalize managers that under-perform on NSL measures and greatly over-perform on SL measures. In a real world scenario, managers are likely to under-perform on some measures. If both SL and NSL measures are included in a BSC, managers may sacrifice effort on a SL measure to meet a target on a less important measure. This supports the viewpoint espoused by Kaplan and Norton (2001) that only SL measures should be included in a BSC.

References

- Banker, R. D., G. Potter, and D. Srinivasan. 2000. An empirical investigation of an incentive plan that includes non-financial performance measures. *The Accounting Review* Vol 75 (1): 65-92.
- Banker, R.D., H. Chang, M.J. Pizzini. 2004. The balanced scorecard: Judgmental effects of performance measures linked to strategy. *The Accounting Review* Vol 79 (1): 1-23.
- Banker, R.D., H. Chang, M.J. Pizzini. 2006. Strategy mapping in a balanced scorecard. Working paper.
- Baumeister, R.F., E. Bratslavsky, C. Finkenauer, and K.D. Vohs. 2001. Bad is stronger than good. *Review of General Psychology* 5(4): 323-370.
- Bryant, L., D.A. Jones, S.K. Widener. 2004. Managing value creation within the firm: An examination of multiple performance measures. *Journal of Management Accounting Research* 16: 107-131.
- Cacioppo, J.T., W.L. Gardner, and G.G. Berntson. 1997. Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review* 1(1): 3-25.
- Einhorn, H.J., and R.M. Hogarth. 1981. Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology* 32: 53-88.
- Ganzach, Y. 1995. Negativity (and positivity) in performance evaluation: Three field studies. *Journal of Applied Psychology* Vol 80 (4): 491-499.
- Glenberg, A., and W. Langston. 1992. Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language* 31: 129-151.
- Haywood, M.E., and N. V. Stuart. 2006. Mental accounting and the balanced scorecard. Working Paper.
- Ito, T.A., J.T. Larson, N.K. Smith and J.T. Cacioppo. 1998. Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology* Vol 75 (4): 887-900.
- Ittner, C.D., and D.F. Larcker. 2003. Coming up short on nonfinancial performance measurement. *Harvard Business Review* 81(11): 88-95.
- Ittner, C.D., and D.F. Larcker. 1998. Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research* 36: 1-35.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decisions under risk. *Econometrica* 47(3): 263-291.

Kanouse, D.E., and L.R. Hanson. 1972. *Negativity in evaluation*. New York: General Learning Press.

Kaplan, R.S., and D.P. Norton. 2001. Transforming the balanced scorecard from performance measurement to strategic management: Part I. *Accounting Horizons* 15(1): 87-104.

Kaplan, R.S. and D.P. Norton. 2000. Having trouble with your strategy? Then map it. *Harvard Business Review* 78 (5): 167-176.

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

Kaplan, R.S., and D.P. Norton. 1992. The balanced scorecard – Measures that drive performance. *Harvard Business Review* 70 (1): 71-79.

Lipe, M., and S. Salterio. 2002. A note on the judgmental effects of the balanced scorecard's information organization. *Accounting, Organizations and Society* 27(6): 531-540.

Oliva, T.A., D.L. Day, and W.S. DeSarbo. 1987. Selecting competitive tactics: Try a strategy map. *Sloan Management Review* 28: 5-15.

Peeters, G., and J. Czapinski. 1990. Positive-Negative Asymmetry in Evaluations: The distinction between affective and informational negativity effect. *European Review of Social Psychology* 1:33-60.

Pratto, F., and O.P. John. 1991. Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology* 61(3): 380-391.

Reisinger, H., K.S. Cravens, and N. Tell. 2003. Prioritizing performance measures within the Balanced Scorecard Framework. *Management International Review* 43(4): 429-437.

Rozin, P., and E.B. Royzman. 2001. Negativity bias, negativity dominance and contagion. *Personality and Social Psychology Review*

Schepanski, A., and T. Shearer. 1997. A prospect theory account of the income tax withholding phenomenon. *Organizational Behavior and Human Decision Processing* 63(2): 174-186.

Silk, S. 1998. Automating the balanced scorecard. *Management Accounting* (May): 38-44.

Skowronski, J.J., and D.E. Carlston. 1989. Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin* 105: 131-142.

Smith, N.K., J.T. Larsen, T.L. Chartrand, J.T. Cacioppo, H.A. Katsfiasz, and K.E. Moran. 2006. Being bad isn't always good: Affective context moderates the attention bias toward negative information. *Journal of Personality and Social Psychology* 90(2): 210-220.

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.

Yates, J.F. 1990. *Judgment and Decision Making*. Upper Saddle River, NJ: Prentice-Hall, Inc.

TABLE 1
Descriptive Statistics of Manager A and Manager B Evaluations

Panel A: Manager A Evaluation^a

Negativity ^c	Dominance ^d		Overall
	SL Dominant	NSL Dominant	
Absent	10.150 ^e (0.988)	9.273 (1.386)	9.690 (1.278)
Present	7.833 (1.581)	6.524 (1.721)	7.128 (1.765)
Overall	9.052 (1.739)	7.930 (2.075)	8.457 (1.994)

Panel B: Manager B Evaluation^a

Negativity ^c	Dominance ^d		Overall
	SL Dominant	NSL Dominant	
Absent	9.000 (1.257)	10.000 (1.447)	9.524 (1.435)
Present	9.111 (1.183)	9.381 (1.687)	9.256 (1.464)
Overall	9.052 (1.207)	9.698 (1.582)	9.395 (1.446)

Panel C: Evaluation Difference^b

Negativity ^c	Dominance ^d		Overall
	SL Dominant	NSL Dominant	
Absent	1.15 ^e (1.37)	-0.73 (1.49)	0.17 (1.62)
Present	-1.28 (1.64)	-2.86 (2.06)	-2.13 (2.01)
Overall	0.00 (1.85)	-1.77 (2.07)	-0.94 (2.15)

^a Manager A and Manager B were each evaluated on a 13-point scale with 0 labeled 'Reassign' and 12 labeled 'Excellent'.

- ^b Evaluation difference is a within participant measure and is calculated as the evaluation of Manager A less the evaluation of Manager B. Therefore, positive numbers indicate that Manager A was evaluated more favorably than Manager B and negative numbers indicate that Manager A was evaluated less favorably than Manager B.
- ^c Negativity was manipulated between participants at two levels: Negativity absent (managers A and B outperformed on all targets) and Negativity present (Manager A outperformed on some measures and underperformed on other measures while Manager B outperformed on all measures)
- ^d Dominance was manipulated between participants at two levels: Non-Strategically Linked Dominant (Manager A dominated on NSL measures and Manager B dominated on SL measures) and Strategically Linked Dominant (Manager A dominated on SL measures and Manager B dominated on NSL measures).
- ^e Panel values are the means. Standard deviations are shown in parentheses below the means.

TABLE 2
Analysis of Bonus

Panel A: Descriptive statistics of Bonus^a

Negativity ^b	Dominance ^c		Overall
	SL Dominant	NSL Dominant	
Absent	11,925 ^d (2,319)	8,868 (1,938)	10,324 (2,608)
Present	9,139 (3,138)	6,762 (2,327)	7,859 (2,949)
Overall	10,605 (3,045)	7,840 (2,365)	9,137 (3,025)

^a Participants allocated a \$20,000 bonus between Manager A and Manager B. Manager A Bonus is a within participant measure and is the amount of bonus allocated to Manager A. Therefore, amounts greater than \$10,000 indicate that Manager A was allocated more bonus than Manager B and amounts less than \$10,000 indicate that Manager A was allocated less bonus than Manager B.

^b Negativity was manipulated between participants at two levels: Negativity absent (managers A and B outperformed on all targets) and Negativity present (Manager A outperformed on some measures and underperformed on other measures while Manager B outperformed on all measures)

^c Dominance was manipulated between participants at two levels: Non-Strategically Linked Dominant (Manager A dominated on NSL measures and Manager B dominated on SL measures) and Strategically Linked Dominant (Manager A dominated on SL measures and Manager B dominated on NSL measures).

^d Panel values are the means. Standard deviations are shown in parentheses below the means.

TABLE 2
Analysis of Promotion Decision

Panel A: Descriptive statistics of Promotion of Manager A^a

Negativity ^b	Dominance ^c		Overall
	SL Dominant	NSL Dominant	
Absent	83.3% ^d	22.7%	50.0%
Present	27.8%	4.8%	15.4%
Overall	55.6%	14.0%	32.9%

^a Participants were told that an important vice-president job was available. They were asked to choose either Manager A *or* Manager B for a promotion.

^b Negativity was manipulated between participants at two levels: Negativity absent (managers A and B outperformed on all targets) and Negativity present (Manager A outperformed on some measures and underperformed on other measures while Manager B outperformed on all measures)

^c Dominance was manipulated between participants at two levels: Non-Strategically Linked Dominant (Manager A dominated on NSL measures and Manager B dominated on SL measures) and Strategically Linked Dominant (Manager A dominated on SL measures and Manager B dominated on NSL measures).

^d The table presents the percentage of participants that selected Manager A for promotion (a result of 50% would indicate that ½ of the participants selected Manager A for promotion and the remaining 50% of participants selected Manager B for promotion).

APPENDIX 1
BALANCED SCORECARD USED IN EXPERIMENT

Measure	Target	Chris Peters, Region A		Taylor Graham, Region B	
		Actual ²	% Over / (Under) Target	Actual ³	% Over / (Under) Target
Financial:					
1. Average price markups of merchandise ¹	60%	58.02%	(3.30%)	60.83%	1.38%
2. Sales growth from existing stores ¹	15%	14.78%	(1.47%)	16.39%	9.27%
3. Inventory turnover	6.00	6.86	14.33%	6.22	3.60%
4. Return on assets	20%	22.24%	11.20%	21.31%	6.54%
Customer:					
1. Price relative to competitors' price	7%	7.99%	14.14%	7.52%	7.43%
2. Customer satisfaction rating ¹	80%	78.32%	(2.10%)	83.62%	4.53%
3. Percentage of sales to new customers ¹	5.0%	4.8%	(4.00%)	5.20%	4.00%
4. Number of Smithson's credit card customers per store	8,000	9,113	13.91%	8,479	5.99%
Internal Process:					
1. Percentage of sales from 'market leaders' ¹	80%	78.30%	(2.13%)	83.04%	3.80%
2. Advertised merchandise in stock	78%	89.70%	15.00%	85.08%	9.08%
3. In-Store Experience - "Mystery Shopper" audit rating ¹	85%	82.66%	(2.75%)	93.07%	9.49%
4. Average number of items per sale	2.00	2.275	13.75%	2.03	1.50%
Learning and Growth:					
1. Employee satisfaction rating ¹	80%	77.86%	(2.68%)	81.37%	1.71%
2. Employee suggestions per year	2.50	2.81	12.40%	2.57	2.80%
3. Store computerization	60%	68.52%	14.20%	66.02%	10.03%
4. Hours of training invested in brand managers each year ¹	80.0	78.8	(1.50%)	86.3	7.88%

¹ These measures were identified as "strategically linked" measures in the case materials.

² Manager A's actual performance changed based upon the negativity condition. The example shown here is "SL Negative".

³ Manager B's actual performance was identical in all conditions.