

**THE IMPACT OF THE CHOICE OF PERFORMANCE EVALUATION SYSTEM ON  
THE MAGNITUDE OF THE OUTCOME EFFECT**

Lasse Mertins

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy  
in  
General Business  
with a major in Accounting

Debra A. Salvador (Chair)  
Sudip Bhattacharjee  
Deborah F. Cook  
J. Gregory Jenkins  
Larry N. Killough

March 23, 2009  
Blacksburg, Virginia

Keywords: Outcome Effect, Outcome Bias, Performance Evaluation System, Measure Controllability

Copyright 2009, Lasse Mertins

# **THE IMPACT OF THE CHOICE OF PERFORMANCE EVALUATION SYSTEM ON THE MAGNITUDE OF THE OUTCOME EFFECT**

Lasse Mertins

## **ABSTRACT**

This dissertation examines whether the magnitude of the outcome effect is impacted by the type of performance evaluation system (subjective versus formula-based). The outcome effect is a phenomenon that occurs when an evaluator overemphasizes the outcome of a decision and ignores essential information that is available to the evaluator (e.g., market information, information about the decision-making process). This outcome focus leads to a more positive (negative) performance evaluation when the outcome exceeds (fails to meet) expectations. Prior studies have not examined whether the type of evaluation system (formula-based versus subjective) has an impact on the magnitude of the outcome effect. In a formula-based evaluation system, outcome measures are pre-weighted and an overall variance measure is easily calculated. Conversely, there are no predefined weights or overall variance measures in a subjective system. Instead, evaluators weight the importance of outcome information themselves. For this dissertation, I conducted an experiment in which 99 business professionals enrolled in a MBA program evaluated the performance of a retail store manager. Their evaluation was based on information that they received about the manager's decision, along with situational factors that may have impacted the decision outcome. The results demonstrate that although the magnitude of the outcome effect was larger when a formula-based system was employed relative to a subjective system, this difference was not statistically significant. Nonetheless, this study provides initial evidence that managers using formula-based evaluation systems should be particularly aware of the outcome effect when conducting performance appraisals. In addition, this study documents the perceived controllability of four financial and four non-financial

measures that are commonly employed to evaluate performance in the retail industry. As hypothesized, the non-financial measures were perceived to be more controllable than the financial measures. This suggests that non-financial measures should be included in the mix of performance measures used in a performance appraisal system.

## **DEDICATION**

This dissertation is dedicated to my wonderful wife Rachel, my parents Barbara and Jörg, and my siblings Björn and Imi. Without their support and encouragement throughout my life, this dissertation would have never been completed. Thank you!

## **ACKNOWLEDGMENTS**

First and foremost, I would like to acknowledge the tremendous support I received from my dissertation chair, Debra Salbador, throughout each stage of this process. She invested a significant amount of time advising me on every aspect of this project. I would also like to thank both the current and past members of my committee: Sudip Bhattacharjee, Deborah Cook, Greg Jenkins, Larry Killough and Edward Wolfe. Each of them provided valuable feedback. Furthermore, I would like to express my appreciation to Sudip Bhattacharjee, James Littlefield, and Patty Lobingier, who assisted me with the data collection. I am also grateful to Bob Paty for providing his advice during the development of the instrument.

Moreover, I would like to thank James Long, not only for the editorial and intellectual input that he provided on many occasions, but also for his incredible friendship over the last four years. In addition, I wish to thank two of my other fellow Ph.D. program friends, Meghann Cefaratti and Robert Crossler, for their advice, encouragement and friendship.

Furthermore, I am indebted to my “American” family, Susan and Joseph Urban, for their kindness and support, which made the transition to a foreign culture much easier. Additionally, I thank my best friends, Sven, Dennis and Holgi, for making the breaks from the Ph.D. program as enjoyable as possible. I am also grateful for the opportunity that Oliver Weiss and the Virginia Tech’s Men’s Soccer program gave me to pursue my Master’s degree while playing competitive sports. Without that experience, I would not have entered the accounting Ph.D. program.

Finally, I would like to acknowledge the financial support that I received from the ACIS Department over the last four years, as well as the advice I received from the director of the doctoral program, Jack Maher. I would also like to thank Phyllis Neece, Arnita Perfater, Anna Hawthorne and Kathy Caldwell for their support and assistance.

## TABLE OF CONTENTS

<b>CHAPTER 1 INTRODUCTION .....</b>	<b>1</b>
<b>CHAPTER 2 LITERATURE REVIEW AND HYPOTHESES .....</b>	<b>6</b>
2.1. Literature Review .....	6
2.1.1. Managerial Performance Evaluations.....	6
2.1.2. Performance Evaluations in Subjective and Formula-Based Evaluation Systems.....	8
2.1.3. Research on the Outcome Effect .....	10
2.1.4. Financial and Non-Financial Measures as Performance Evaluation Measures.....	16
2.2. Hypothesis Development .....	19
2.2.1. Hypothesis 1 .....	19
2.2.2. Hypothesis 2 .....	20
2.2.3. Hypothesis 3 .....	22
<b>CHAPTER 3 RESEARCH METHODOLOGY.....</b>	<b>24</b>
3.1. Sample.....	24
3.2. Variables.....	25
3.3. The Instrument and Pilot Testing .....	26
3.3.1. The Instrument .....	26
3.3.2. Pilot Testing .....	27
3.4. The Experiment .....	28
<b>CHAPTER 4 RESULTS.....</b>	<b>31</b>
4.1. Instrument Reliability Test and Principal Component Analysis .....	31
4.1.1. Performance Evaluation Instrument Reliability (Cronbach's Alpha) .....	31
4.1.2. Principal Component Analysis.....	31
4.2. ANOVA Assumption Testing .....	32
4.3. Results for Hypothesis 1 .....	32
4.4. Results for Hypothesis 2 .....	33
4.5. Results for Hypothesis 3 .....	34
4.6. Additional Analyses .....	36
4.6.1. Hypothesis Testing Excluding Part-time MBA Students .....	36

<b>CHAPTER 5 DISCUSSION, IMPLICATIONS AND LIMITATIONS .....</b>	<b>40</b>
5.1. Discussion and Implications.....	40
5.2. Limitations .....	43
<b>REFERENCES.....</b>	<b>44</b>
<b>FIGURES.....</b>	<b>50</b>
<b>TABLES.....</b>	<b>54</b>
<b>APPENDIX A: THE INSTRUMENT .....</b>	<b>65</b>

## LIST OF FIGURES

FIGURE 1: Experimental Design .....	51
FIGURE 2: Measurement of the Outcome Effect.....	52
FIGURE 3: Results for Hypothesis 2.....	53

## LIST OF TABLES

TABLE 1: Demographic Data .....	55
TABLE 2: Comparison of Demographic Data among Treatment Groups .....	56
TABLE 3: Final Sample .....	57
TABLE 4: ANOVA Results .....	58
TABLE 5: Perceived Controllability of Financial and Non-Financial Measures.....	59
TABLE 6: Perceived Controllability: Ranking of Measures and Description of Homogeneous Subsets .....	60
TABLE 7: Multiple Comparisons of Perceived Controllability Means for all Eight Measures....	61
TABLE 8: ANOVA Results for PMBA and EMBA Students .....	62
TABLE 9: Perceived Controllability of Financial versus Non-Financial Measures for PMBA and EMBA Students .....	63
TABLE 10: Perceived Controllability: Ranking of Measures and Description of Homogeneous Subsets for PMBA and EMBA Students .....	64

## **CHAPTER 1**

### **INTRODUCTION**

Performance evaluations play a critical role in the success or failure of a manager's career. Managerial compensation, promotion, and motivation are significantly impacted by performance evaluations. Managerial accounting research has extensively examined several factors that may influence the objectivity of these appraisals. In the past, performance evaluations were based solely on financial accounting measures such as return on investment and profits (Balkcom, Ittner, & Larcker, 1997; Feltham & Xie, 1994; Ittner, Larcker, & Rajan, 1997). However, in recent years, performance measurement systems have begun to incorporate a mixture of financial and non-financial measures (e.g., Kaplan and Norton's balanced scorecard approach (1992)). One of the primary reasons for this development is the notion that financial measures do not adequately capture the long-term impacts of managerial actions and decisions (Feltham & Xie, 1994; Hemmer, 1996). Contemporary research emphasizes the importance of including non-financial measures in evaluations because they are generally linked to a company's critical success factors; therefore, they are indicators of a company's strategic success (DeBusk, Brown, & Killough, 2003; Ittner & Larcker, 1998b; Ittner, et al., 1997).

Although incorporating non-financial measures into an appraisal system provides a more complete picture of a manager's performance relative to financial measures alone, it is unclear whether evaluation systems that incorporate both non-financial and financial measures lead to higher quality performance assessments relative to less complex traditional systems. For example, the addition of non-financial measures increases the diversity among performance measures, leading to increased levels of subjectivity in the evaluation process (Ittner, Larcker, &

Meyer, 2003; Moers, 2005). However, Moers (2005) found that increases in measure diversity and subjectivity are positively correlated with performance evaluation bias.

Prior research has included some discussion about whether a multi-criteria performance evaluation approach should employ well-defined formulas (specifying the weighting of each appraisal measure) or should utilize a more subjective approach, defined by the evaluator.

According to Kaplan and Norton (1996), integrating a multiple measures evaluation system with a formula-based performance evaluation system is problematic because this type of system is vulnerable to gaming, and compensation bonuses are sometimes paid even when certain goals are not achieved. On the other hand, subjective evaluations are susceptible to favoritism, bias, and the risk that evaluatees will perceive appraisals to be unfair (Prendergast & Topel, 1993). In their review article, Ittner and Larcker (1998b) state that “since the net benefits of formulaic vs. subjective performance evaluations are unclear, this issue offers an exciting topic for future research” (p. 228).

The outcome effect occurs when evaluators with outcome knowledge assess an evaluatee more positively (negatively) when the outcome of a decision is better (worse) than expected (Emby, Gelardi, & Lowe, 2002; Ghosh & Ray, 2000). This transpires because the evaluator focuses on the decision outcome when that outcome becomes available, and ignores additional available and relevant information about other factors that may have influenced the actual outcome (e.g., market factors that were uncontrollable for the evaluated manager). For example, if a manager spends money to redesign his firm’s marketing strategy, and the company’s sales volume increases afterward, the marketing manager will be evaluated more positively when the outcome is known. This occurs even if the sales increase was primarily a result of rapid market growth, and not related to the firm’s marketing strategy. Several studies have examined

strategies that may mitigate the outcome effect (Clarkson, Emby, & Watt, 2002; de Villiers, 2002; Fisher & Selling, 1993; Frederickson, Peffer, & Pratt, 1999; Ghosh & Ray, 2000).

However, no prior study has examined how the type of performance evaluation system may influence the occurrence and magnitude of the outcome effect.

The primary goal of this study is to examine the impact of two different types of performance evaluation systems – formula-based and subjective – on the magnitude of the outcome effect. Formula-based evaluation systems clearly define each performance measure's weighting and enable the calculation of an overall performance variance measure. The overall variance measure is calculated by multiplying the weight that is assigned to each outcome measure with its variance from the budget. The products are then summed to compute an overall variance measure ( $\sum_i[\text{weight}_i \times \text{target variance}_i]$ ). This encourages objective performance evaluations within a firm. On the other hand, a subjective performance assessment system allows evaluators to define their own weighting for each performance measure.

Because an aggregated overall variance measure is extremely salient, I expect that performance evaluators, using a formula-based system, would focus primarily on the overall performance variance measure (overall variance from target), and fail to integrate information about environmental and situational factors into their evaluation. Prior salience effect research proposes several different contexts in which salience may play a role (Haynes & Kachelmeier, 1998). The context most relevant to this study involves features that may focus the evaluator's interest on specific aspects of the decision-making environment (Sanbonmatsu, Kardes, Posavac, & Houghton, 1997). In a formula-based performance evaluation system, pre-defined weights allow the evaluator to calculate the overall performance variance, which may be the most salient feature of the decision context. Evaluators may focus on the overall variance measure, while

ignoring subjective information that might be relevant to the evaluation. Consequently, my primary research question is: Does the nature of the performance evaluation system (formula-based vs. subjective) lead to different levels of the outcome effect?

The outcome effect has been broadly examined in prior research. Within the psychology and accounting domains, researchers have studied the effects of outcome knowledge on performance evaluations extensively (Brown & Solomon, 1987, 1993; Frederickson, et al., 1999; Ghosh, 2005; Ghosh & Lusch, 2000; Ghosh & Ray, 2000; Hershey & Baron, 1992). Ghosh (2005) combined research on the outcome effect with research on financial and non-financial measures. He found that retail managers perceived certain non-financial measures (customer and employee satisfaction) to be more controllable than certain financial measures (return on investment and sales per square foot). The outcome effect was greater when performance evaluations were based on two non-financial measures relative to performance evaluations that used financial measures. These differences were attributed to the differences in perceived controllability between financial and non-financial measures. However, the participants only assessed the controllability of a single measure (four treatment groups with one measure each). In this dissertation, each participant evaluated the perceived controllability of eight commonly used performance measures.

This study makes several contributions to the existing accounting literature and practice. First, it provides an initial examination of the impact of different performance evaluation systems on the outcome effect. This study may alert managers to the vulnerability of their type of performance evaluation system with respect to outcome effects. Managers may use the results of this study to decide whether they need to implement mitigating strategies in their appraisal systems to reduce evaluator outcome focus. Secondly, it demonstrates that the outcome effect

also occurs when subjects are explicitly told that certain uncontrollable events led to the outcome. Finally, this paper contributes to the performance evaluation measure literature by providing evidence about the perceived controllability of eight financial and non-financial measures. Because performance measures should be controllable by the evaluatee, the findings of this analysis may provide evaluators with a better understanding on which measures are most suitable for their performance evaluation instrument.

I conducted an experiment with 99 business professionals enrolled in Virginia Tech's EMBA, PMBA and part-time MBA programs. The participants were asked to read a case about a retail store manager who made an investment decision in the prior year and to then evaluate the manager's performance based on information provided in the case (including the outcome of eight financial and non-financial measures). After making their evaluation, each participant was asked to assess the perceived controllability of the eight measures. The results showed that the magnitude of the outcome effect in the formula-based performance evaluation system was larger than the outcome effect in the subjective system. However, this difference was not statistically significant.

The remainder of this paper is structured as follows: Chapter Two reviews the literature for the key elements of this study, which include performance evaluation, the type of evaluation system, the outcome effect, and performance measure controllability. Additionally, it develops three hypotheses. Chapter Three explains the research methodology. Chapter Four presents the results and analyses. Finally, Chapter Five discusses the findings, implications, and limitations of the study.

## CHAPTER 2

### LITERATURE REVIEW AND HYPOTHESES

#### 2.1. LITERATURE REVIEW

##### **2.1.1. Managerial Performance Evaluations**

Organizational performance evaluation systems are multifaceted, and include political, social, affective (emotional) and cognitive components (Ferris, Judge, Rowland, & Fitzgibbons, 1994; Ferris, Munyon, Basik, & Buckley, 2008). Performance evaluations are an integral part of both the developmental and compensation processes for business managers and employees.

According to Prendergast and Topel (1993), performance evaluations are generally intended to compensate employees according to their contributed effort and to verify their actual abilities. Hence, one primary performance evaluation objective is to distinguish between highly and less skilled managers (Moers, 2005). Furthermore, performance evaluations can also significantly affect an evaluatee's behavior, productivity, and motivation (Feltham & Xie, 1994; Wier, Stone, & Hunton, 2002). Therefore, performance evaluations are also structured to encourage employees to align their behavior with the strategic objectives of the company.

The results of performance evaluations are commonly used for promotion and compensation decisions. They often represent the only tool an organization uses to assess a manager's accomplishments. Positive evaluations provide access to career opportunities. Therefore, a performance evaluation is critical for a manager's career success and motivation and organizations should focus on implementing a performance evaluation system that accurately and objectively assesses its managers' performance. It is unclear which performance evaluation system (formula-based or subjective) best achieves these goals. While a formula-based system appears to result in a more objective performance assessment, one could argue that by including

all relevant perspectives in the performance evaluation, a subjective system could potentially lead to a more accurate assessment. A formula-based system is unidimensional in the sense that the evaluator only processes a few cues in the form of highly aggregated measures to make a performance judgment (in the case of pure formula-based systems, the evaluator only processes one cue). In contrast, a subjective system represents a multi dimensional performance system because the evaluator is required to include several perspectives on the employee's performance because no highly aggregated data is available.

Performance evaluations were traditionally based on aggregate financial accounting measures (Bushman, Indjejikian, & Smith, 1996; Ittner, et al., 1997). However, this perspective changed in the early 1990s when Kaplan and Norton introduced the balanced scorecard (Kaplan & Norton, 1992). Their principal critique of the traditional evaluation system was that financial measures are subject to a time lag and therefore only reflect outcomes from past events (Kaplan & Norton, 2001a). Based on Fisher's (1995) and Brancato's (1995) case studies, Ittner and Larcker (1998b) provided three primary reasons why organizations should integrate non-financial performance evaluation measures: 1) perceived limitations of traditional accounting-based measures (e.g., too historical, focus on short-term success factors), 2) competitive pressures and 3) consistency with other initiatives (e.g., total quality management [TQM], just in time [JIT] production) (Banker, Potter, & Schroeder, 1993). Furthermore, financial measures often fail to adequately capture the long-term effects of activities that might impact a company's critical success factors, including innovation or productivity improvements (Kelly, 2007).

## **2.1.2. Performance Evaluations in Subjective and Formula-Based Evaluation Systems**

To date, few studies have examined how the type of performance evaluation system influences the occurrence and magnitude of the outcome effect. The circumstances under which explicit weights should be applied to performance measures are generally unknown, nor is it clear whether the evaluator should be responsible for defining the weights for each measure (Ittner & Larcker, 1998b). In a multiple measure performance assessment system, a company can generally choose between a formula-based and a subjective system. In a formula-based evaluation system, specific weights are assigned to each evaluation measure. In contrast, the evaluator implicitly assigns weights to each measure in a subjectively-oriented system (Ittner & Larcker, 1998b). Each approach has distinct advantages and disadvantages, and it is unclear which system best assesses a manager's skills. The major advantage of the formula-based system is that it generally provides a higher level of objectivity. However, Kaplan and Norton (1996) note that there are several inherent difficulties when a multiple performance measure approach is integrated with a formula-based performance evaluation system. First, it may be difficult to assign appropriate weights to estimate the relative importance of each performance measure. Secondly, formula-based systems provide an opportunity for individuals to "game" the system. Finally, many companies pay bonuses when some goals are exceeded, even when other goals are not met.

The alternative is to employ a subjective performance evaluation system. The primary advantage of a subjective performance evaluation system is that the evaluator may include every available piece of information in the assessment decision (Ittner & Larcker, 1998b). Therefore, the evaluation would also include environmental and situational factors that may not be captured

by the performance measures. However, there are some disadvantages associated with this method. According to Prendergast and Topel (1993), some non-financial measures are difficult to quantify. Additionally, an evaluator may favor an evaluatee based on personal preferences instead of objective measures (Prendergast & Topel, 1996). Finally, evaluatees may perceive the performance system to be unfair if they do not understand the weighting of the evaluation measures.

Ittner et al. (2003) employed a case study approach to examine how firms actually weight financial and non-financial performance measures in a subjective balanced scorecard system used for compensation bonus decisions. They investigated this issue from both an economic and a psychological perspective. They established that evaluators do not weight each balanced scorecard measure equally. The subjectivity inherent in the system allowed evaluators to weight financial measures significantly higher than non-financial measures, a method that conflicted with the stated intention of the balanced scorecard. Evaluators also changed the weights on certain measures from period to period. The authors concluded that psychologically oriented approaches to investigating performance measurement procedures may be more useful than economically oriented approaches. While there was no significant correlation between assigned weights to non-financial measures and their ability to forecast upcoming financial outcomes (which would support economic theories), they discovered that “most of the weight was placed on quantitative, outcome-oriented financial measures that were used in earlier plans, results that are consistent with psychology-based predictions.” (Ittner, Larcker, & Randall, 2003, p. 754)

This provides additional motivation to conduct research in this area from a psychological perspective.

Haywood (2001) examined the impact of financial and non-financial measures on the outcome effect and framing biases by including specific weights for specific balanced scorecard categories. However, she did not focus on the performance rating or on the presence and magnitude of the outcome effect. Instead, she examined whether the evaluator employed more compensatory strategies or non-compensatory strategies in their decision-making process under varying conditions. While Haywood (2001) did not test for the outcome effect directly, she concluded that the implementation of specific weights may not be an appropriate tool to mitigate outcome effects or framing biases. Although she expected that explicit weights would increase the amount of information an evaluator includes in the performance evaluation, she found the opposite to be the case. This finding supports my expectation that evaluators decrease their search for additional information in a formula-based system in which specific weights are assigned to all outcome measures.

### **2.1.3. Research on the Outcome Effect**

Outcome knowledge typically leads evaluators towards a higher (lower) decision rating when the outcome exceeds (does not meet) expectations (Ghosh & Ray, 2000). Research on the effects of outcome knowledge is rooted in psychology literature. Fischhoff (1975) found that people estimated a higher probability that a certain event would occur after they knew about the actual outcome relative to individuals who did not receive such information. This phenomenon was labeled “hindsight bias.” Hawkins and Hastie (1990) defined hindsight bias as “the tendency for people with outcome knowledge to believe falsely that they would have predicted the reported outcome of an event” (p. 311). Once an outcome is known, evaluators perceive that their ability to predict that outcome is higher than it likely was at the time the decision was made.

This belief will then affect their assessment of the actual outcome probability. On the other hand, the outcome effect deals with the effects of outcome knowledge on the evaluation of a decision. The outcome effect is not concerned with the ability to predict the outcome but rather with the actual effect that outcome has on a decision evaluation. Several psychology studies have shown that outcome knowledge frequently affects the evaluator's judgment about decision quality (Lipshitz, 1989; Mitchell & Kalb, 1981). The outcome effect has been observed even when outcome information did not contribute to the evaluation of the quality of a decision (Brown & Solomon, 1993).

Prior research referred to the outcome effect as an error or a bias (Ghosh, 2005; Lipshitz, 1989). However, in contrast to hindsight bias, outcome effects do not always lead to bias (Brown & Solomon, 1987; Hershey & Baron, 1992). For example, an evaluator may have difficulty observing and assessing another individual's decision process, and the outcome may provide cues about the decision process. This may allow an evaluator to better assess a decision's quality (Brown & Solomon, 1987). Tan and Lipe (1997) agreed, stating that "judgment by outcomes may not be dysfunctional when the evaluator does not know how the decision maker chose his or her action" (p. 315). However, they warned that it is not always advantageous to infer decision quality from outcomes. They argued that information about the decision maker's decision process (if available), and not outcome information, should be primarily used to assess decision quality.

Hershey and Baron (1995) developed a simple model to evaluate the circumstances under which the outcome effect does not result in undesirable bias in the evaluation process. This model requires that two conditions be met. First, the probability of a successful outcome (S) as the result of a good decision (G) must be greater than the probability of a successful outcome

resulting from a bad decision (B). Secondly, the probability of a failed outcome (F) after a bad decision must be greater than the probability of a failed outcome after a good decision:

$$P(S/G) > P(S/B) \text{ and } P(F/B) > P(F/G)$$

When these conditions are met, outcome information may be used to evaluate decision quality.

While outcome effects do not necessarily lead to bias, prior findings indicate that outcome effects often influence evaluation decisions. Therefore, contemporary accounting research attempts to provide a better understanding of the causes of the outcome effect (motivational versus cognitive) (Brown & Solomon, 1993; Clarkson, et al., 2002; Lipe, 1993; Tan & Lipe, 1997), and to identify potential mitigating factors (Clarkson, et al., 2002; de Villiers, 2002; Fisher & Selling, 1993; Frederickson, et al., 1999; Ghosh & Ray, 2000; Jones & Chen, 2005; Kadous, 2001).

Brown and Solomon (1987) were the first to investigate the effects of outcome knowledge on performance evaluations within the area of managerial accounting. They found that managerial decision evaluations can be significantly influenced by outcome information. They also examined two factors that could mitigate the effects of outcome knowledge: the evaluator's prior involvement with the decision process and the extent of the decision-maker's responsibility for anticipating the outcome. When an evaluator was involved in the decision process, the outcome effect was significantly mitigated. The outcome effect was also found to be increasing relative to the level of decision-maker responsibility of anticipating the outcome.

Lipe (1993) tested the effects of variance investigation outcomes on managerial performance evaluations and examined cognitive explanations for the occurrence of the outcome effect. She found that superiors evaluated their managers more positively when the managers

found failures in the system during the variance investigation. She attributed this finding to the subjects classifying investigation expenditures as costs when they thought that the expenditure benefitted the company and as losses when they did not think that there was any benefit in spending these expenditures. Subjects that based their decision evaluation on the cost frame gave higher evaluation ratings than subjects utilizing the loss frame.

Ghosh and Ray (2000) examined additional ways in which to mitigate the outcome effect. They based their research on the information asymmetry that exists between the evaluator and the decision-maker/manager, i.e., an evaluator cannot be informed about all facets of a manager's performance. They tried to identify methods to decrease this information asymmetry, which in turn may mitigate the outcome effect. They considered three approaches: 1) increased evaluator participation in the manager's decision-making process (Brown & Solomon, 1987, 1993; Fisher & Selling, 1993), 2) more observation of the decision process by the evaluator (Fisher & Selling, 1993), and, 3) framing (Lipe, 1993).

The authors described several limitations for each approach and presented an alternative method that might mitigate the outcome effect. They found that if the evaluator receives sufficient information about the nature and extent of uncertainty presented as a range of probabilities, rather than as a precise estimate, the outcome effect can be mitigated. The authors proposed that the outcome effect can also be mitigated by training evaluators to be aware of potential outcome biases and through better communication about the decision-making process.

Tan and Lipe (1997) investigated whether perceived outcome controllability and providing the evaluator with specific information about the quality of the decision process had an influence on the outcome effect. They found that while a higher degree of controllability increased the level of the outcome effect for inexperienced subjects, a higher level of

controllability only affected the experienced subject group in the negative condition (performance ratings were lower when outcome was controllable). In addition, the outcome effect was reduced (but not eliminated) in two circumstances: when specific information about the quality of the decision process was available to the evaluator and, for the less experienced group, when the performance evaluation was based on uncontrollable measures.

The Tan and Lipe (1997) study motivated a research stream that examined the effects of the controllability of measures on performance ratings and the outcome effect. Ghosh and Lusch (2000) conducted a field study that investigated whether the outcome effect was present in an organizational setting. They found that the outcome effect existed in this subjective performance evaluation setting. They also found that performance evaluation ratings increased (decreased) when managers had control over a specific positive (negative) outcome determinant (inventory, advertising, customer service capacity, and pricing). Evaluation ratings did not increase (decrease) when a positive (negative) outcome was caused by uncontrollable environmental factors (e.g., store saturation, economic health of the trade area). In addition, one unexpected finding was that positive (negative) outcomes due to central management determinants significantly increased (decreased) the store managers' appraisals. Central management determinants (such as total square footage of the store) are typically not controllable for evaluatees. These results demonstrated that the outcome effect was present in an organizational setting and that positive (negative) outcomes attributable to uncontrollable factors can increase (decrease) a manager's performance rating.

De Villiers (2002) explored whether outcome controllability influences performance evaluations in a budgeting setting (above/below budget). In addition, he tested whether the evaluator's knowledge of the reasons that caused the budget variance would impact the outcome

effect. In the above (below) target condition, he demonstrated that performance evaluations were higher (lower) when performance was controllable. In addition, he found that in the above (below) target condition, lower (higher) evaluations resulted from uncontrollable variance causes. When the cause of the variance was known to the evaluators in advance, they provided a more positive performance appraisal than when the causes were unknown.

Rose (2004) examined whether unexpected situational information influenced the outcome effect in a budget setting. The participants, who took on the role of evaluators, were made aware of situational factors that were generally uncontrollable for the evaluated manager. However, the participants were informed that the evaluatee was still responsible for all financial outcomes. In addition, cognitive loads were manipulated in order to examine the potential causes of the outcome effect in a performance evaluation setting.

He found that when financial measures were lower than expected, evaluators rated performance higher when they considered situational factors to be partly responsible. In contrast, in the positive outcome condition, the outcome effect was not decreased by situational factors. In an additional analysis, Rose (2004) found that participants in all conditions processed situational information, but that a financial outcome focus interfered with situational factor processing.

Ghosh (2005) investigated how different types of financial and non-financial performance appraisal measures (return on investment, sales per square foot, customer satisfaction, and employee satisfaction) increased or decreased the outcome effect at various levels of controllability. He found that the outcome effect increased when a manager exerted a high level of control over a measure. In addition, the results revealed that the outcome effect was larger for non-financial measures relative to financial measures. This may be due to the

higher degree of perceived controllability for the non-financial measures used in this case study.

Finally, he reported that when evaluators assessed the controllability of measures prior to the performance evaluation, the outcome effect was significantly mitigated.

#### **2.1.4. Financial and Non-Financial Measures as Performance Evaluation Measures**

Modern performance evaluation systems, such as the balanced scorecard (Kaplan & Norton, 1992, 1996, 2001a, 2001b), include non-financial measures to mitigate the shortcomings of traditional measures and to complement the financial measure-based performance evaluation system. These non-financial measures often drive future financial performance. Therefore, one of the principal objectives of the non-financial measures literature is to determine whether certain non-financial measures (e.g., customer satisfaction) are appropriate indicators of future firm performance. Numerous research studies in this area find a positive relation between strategically-oriented non-financial performance measures and long-term financial performance (Anderson, Fornell, & Lehmann, 1994; Anderson, Fornell, & Rust, 1997; Banker, et al., 1993; Banker, Potter, & Srinivasan, 2000; Ittner & Larcker, 1998a; Smith & Wright, 2004). These results indicate that non-financial measures are often important indicators of future firm performance. This finding is significant because non-financial measures are only useful additions to traditional financial measures when they are reliable predictors of future firm success and add relevant supplementary information about the manager's effort (Ittner & Larcker, 1998b).

However, there is no general rule about whether certain non-financial measures are appropriate or inappropriate performance measures. Although numerous studies support the assumption that non-financial measures can be reliable predictors of future financial performance

(as discussed above), multiple industry or firm specific factors may influence the predictive capability of non-financial measures. For example, Anderson et al. (1994; 1997) found that customer satisfaction was positively correlated with return on investment for Swedish manufacturing firms, but negatively associated for Swedish service companies. In addition, Ittner and Larcker (1998a) found that customer satisfaction was a leading indicator for future revenue growth; however, this result only applied to certain industries at the firm-level. In some industries, this relationship was found to be negative or not statistically significant. This provides evidence that non-financial measures may not be an appropriate tool for performance measurement for all industries and at all levels.

Additionally, prior research has argued that financial and non-financial measures differ in terms of controllability (Ghosh, 2005; Ghosh & Lusch, 2000). Controllability in this context refers to the degree to which a manager's actions can directly influence the outcome of a measure. It is important to realize that the controllability of certain financial measures may be perceived differently, depending on each manager's level within the company. Mid-level managers (such as the participants in Ghosh's (2005) study) perceive financial measures (such as return on investment and sales per square foot) to be less controllable than non-financial measures (such as customer satisfaction and employee satisfaction). High-level managers may be more likely to perceive return on investment to be more controllable (relative to lower-level managers) because they manage and control many factors that have an effect on return on investment. Practitioner oriented literature emphasizes that measures within a performance evaluation system are only useful when they are controllable by the manager (at least to a certain degree). Otherwise, they do not provide additional information about the manager's performance (Merchant, 2007). Therefore, a manager should only be held responsible for

outcomes caused or influenced by her/his actions (Kelley & Michela 1980; Tan & Lipe, 1997) , and should only be rewarded for the same.

Tan and Lipe (1997) examined whether outcome controllability has an impact on the outcome effect. They based their expectations on attribution theory, hypothesizing those positive outcomes which are perceived to be more controllable will lead to an even higher performance rating than positive outcomes that are perceived to be less controllable. They found that controllability affected the magnitude of the outcome effect, but not its presence. Ghosh (2005) documented the perceived controllability of two non-financial and two financial measures. He found that the two non-financial measures had a higher level of perceived controllability; this led to an increase in the outcome effect. These findings, taken in combination, indicate that the inclusion of non-financial measures in performance evaluation systems may increase the magnitude of the outcome effect.

Several factors influence a measure's controllability. First, there is a negative relation between the number of people who can influence a measure and its controllability. Secondly, the degree of controllability is also dependent on the measure's level of aggregation. Aggregate measures provide partial information across many actions, while explicit measures supply information about more specific actions (Prendergast, 2002). Financial measures are often more aggregated, particularly comprehensive financial measures such as net income and return on assets (Moers, 2006), and are often more difficult to control in the short term because several variables may affect them. These variables may be influenced by external factors that management cannot control (e.g., unexpected economic or market changes). On the other hand, “non-financial measures are ‘specific’ performance measures, in the sense that they provide (some) information about a specific subset of actions” (Moers, 2006, p. 901), making them more

controllable for an individual manager. A specific measure is more suitable for bonus compensation purposes because it is more “sensitive” to individual managerial actions than more aggregated measures (Banker & Datar, 1989; Feltham & Xie, 1994; Holmstrom, 1979; Moers, 2006).

## 2.2. HYPOTHESES DEVELOPMENT

### 2.2.1. Hypothesis 1

The first hypothesis focuses on the occurrence of the outcome effect when specific situational information is available to partially explain the positive or negative outcome. The presence of the outcome effect has been found to be robust under various conditions in which additional information is available to the evaluator. Tan and Lipe (1997) provided evaluators with explicit information about the quality of the evaluatee’s decision making process. They found that it reduced (but did not eliminate) the outcome effect. Rose (2004) tested whether outcome inhibiting or facilitating situational factors impacted the outcome effect at various levels of cognitive load. He used unexpected events which impacted the financial outcomes to proxy for situational information. He found that providing the evaluators with information about the situational factors did not effectively mitigate the outcome effect. He concluded that “directing evaluators to attend to situational information in all outcome conditions will not reduce outcome effects” (p. 60).

Rose’s (2004) case materials described several unexpected events that impacted evaluatee performance. Although these events were uncontrollable, the participants were informed that the evaluatee was to be held completely responsible for the financial outcomes. By priming the evaluator to hold the evaluatees responsible for the financial outcomes, evaluator focus on

measure controllability may have been reduced. In this study, I allow the evaluator to determine whether the evaluatee should be held responsible for uncontrollable outcomes. It is unclear whether the outcome effect will still occur when the evaluator receives specific information about uncontrollable situational factors and their impact on financial and non-financial outcomes without being explicitly held responsible for those outcomes. Unlike Rose's (2004) study, participants in this study were not informed that they were completely responsible for all outcome measures and must therefore form their own judgment on the controllability of the outcome measures. Recognizing the uncontrollability of the situational factors and their impact on the outcomes should reduce the outcome effect.

However, because prior research suggests that the actual outcome dominates the performance evaluation decision, and that the processing of situational factors is reduced when information about favorable or unfavorable outcome becomes available to the evaluator (Rose, 2004), I expect that an outcome effect will continue to exist. Therefore, I hypothesize:

H1: Although situational factors are explicitly provided to partially explain the positive/negative variance from the budgeted target, the outcome effect will be present in both the formula-based and the subjective-oriented evaluation systems.

### **2.2.2. Hypothesis 2**

The second hypothesis deals with the mitigating impact that evaluation system type may have on the outcome effect. Prior research has identified several potential mitigating strategies, such as involving the evaluator in the evaluatee's decision making process, providing information about the quality of the evaluatee's decision making process, and including different types of debiasing instructions (Brown & Solomon, 1987; Clarkson, et al., 2002; Tan & Lipe, 1997). Fisher and Selling (1993) found that the outcome effect can be mitigated when the

evaluator observes the decision-making process. This occurs because the evaluator includes a considerable amount of information about the decision making-process in the final performance evaluation, reducing outcome focus and mitigating the outcome effect.

Two types of performance evaluation systems are frequently employed in organizations: formula-based and subjective. In a formula-based evaluation system, the company puts explicit weights on each performance measure beforehand so that evaluatees know which measures are important to the organization's success. An overall target variance measure is computed based on the outcomes and weights.

In a subjective performance evaluation system, the evaluator assigns implicit or explicit weights to the performance measures. Therefore, the evaluator must expend more cognitive effort in this system than in a formula-based system to make an optimal performance evaluation. The absence of predefined weights on measures could lead evaluators to include additional information that might affect the decision-making process of the evaluatee. Attending to this additional information could reduce the evaluator's outcome focus.

In contrast, evaluators in the formula-based condition are expected to focus on the aggregated measure (overall variance measure) when making the performance judgment to decrease the risk of potential information overload, even though this behavior may decrease the quality of their appraisal because some relevant information may not be included in their decision. In this case, evaluators select heuristic strategies in which they leave out certain parts of available data and utilize a lower level of cognitive effort (Johnson & Payne, 1985; Speier, 1998).

In addition, their focus will be directed in this manner because of the salience of the aggregated measure. According to Haynes and Kachelmeier (1998), "holding constant the

number and clarity of informational cues, contextual variations in emphasis can direct the decision maker's attention to certain cues over others" (p. 107). In this context, I would expect that evaluators using a formula-based system will shift their focus from the mix of measures to this single source of information. I would not expect this shift to occur in a subjective performance system because there is no aggregate performance measure to overshadow the mix of individual measures and the additional available information. Based on the differences between the two types of performance systems, I expect that evaluators who use a formula-based system will focus on the overall performance variance measure, which is calculated based on the explicit weights of the performance measures. Therefore, I hypothesize that the outcome effect will be greater for evaluations using formula-based systems than those using subjectively-oriented evaluation systems:

H2: The outcome effect will be greater for evaluations in formula-based evaluation systems than in subjective-oriented evaluation systems.

### **2.2.3. Hypothesis 3**

The third hypothesis examines whether non-financial measures (used to evaluate performance at a large American home improvement chain) are perceived to be more controllable than financial measures. Prior literature has focused on the perceived controllability of measures and the impact on the outcome effect. Ghosh (2005) found that non-financial measures are perceived to be more controllable than financial measures. He also noted that a higher level of perceived controllability led to an increase in the outcome effect. However, he only included two financial and two non-financial measures in his research design. This will be the first study to examine the degree of perceived controllability for several different financial and non-financial measures within a set of measures in a performance evaluation context. While

Ghosh's (2005) study utilized a between-subjects design (each participant assessed the perceived controllability of one single measure), this study employs a within-subject design wherein each participant assesses the controllability of eight different measures. Performance evaluations are usually based on a mix of measures and performance measures should be controllable by the evaluatee. Therefore, the results of this analysis may provide evaluators with better insight on the measures that are most suitable for a performance evaluation instrument. Managers could also use the results to create an "optimal mix" of performance measures for their company.

One reason why non-financial measures are often perceived to be more controllable than financial measures are their lower level of aggregation (Moers, 2006). This makes it easier to define an individual manager's impact on a specific non-financial measure (Baker, Gibbons, & Murphy, 1994). In contrast, several forces may impact aggregated measures. These forces may include managerial decisions from multiple managers, as well as market forces. These factors make it more difficult to control the outcome of these highly aggregated financial measures. Further, non-financial measures better isolate specific actions compared to financial measures. Therefore, my third hypothesis is as follows:

H3: Managers perceive non-financial measures to be more controllable than financial measures.

## **CHAPTER 3**

### **RESEARCH METHODOLOGY**

#### **3.1. SAMPLE**

The sample consists of 99 MBA students from a large public state university. Sixty-five students are professional MBAs (PMBA), 12 are executive MBAs (EMBA), and 22 are part-time MBAs (PartMBA). Ashton and Kramer (1980) suggest that students can proxy for real-world decision-makers in accounting studies that examine human information processing and decision-making because they process information similarly and are susceptible to the same biases. Using student surrogates may only be problematic when factors such as age or experience play a significant role in the decision-making process. I will minimize this risk by limiting participation to MBA students who have work experience. This approach is consistent with prior literature (Banker, Chang, & Pizzini, 2004; DeBusk, et al., 2003; Lipe & Salterio, 2000, 2002; Roberts, Albright, & Hibbets, 2004).

On average, the participants had 9.41 years (S.D. = 5.51) of work experience, 2.34 years of management experience (S.D. = 3.40), 2.19 years (S.D. = 3.20) of evaluator experience, and 7.30 years of evaluatee experience (S.D. = 5.20) (Table 1). Evaluator experience captures the experience that participants have with conducting performance evaluations. Evaluatee experience refers to their experience with having their performance evaluated.

There are no significant differences among the treatment groups with respect to their work experience, management experience, evaluator experience, evaluatee experience, age, education, or gender ( $p > 0.05$ ) (Table 2). I also conducted a series of ANCOVAs in which I included each demographic variable as a covariate individually. The analysis demonstrated that none of these demographic characteristics had a significant effect on the performance ratings.

Initially, 106 business professionals enrolled in a MBA program participated in this study. However, seven subjects were removed from the final sample. Five of these subjects did not recognize that two of the four performance evaluation rating scales were reverse-coded, one was a full-time student who is not employed, and one participant skipped several questions (Table 3).

### 3.2. VARIABLES

I employed a 2 x 2 experimental design. The independent variables are the type of evaluation system (SYSTEM) and the outcome (OUTCOME). The type of evaluation SYSTEM and the OUTCOME are both between-subjects manipulations. Each independent variable has two levels: the type of evaluation SYSTEM is either subjective (weights for evaluation measures are not pre-defined) or formula-based (weights for evaluation measures are specifically assigned). In the positive OUTCOME condition, actual performance exceeded performance targets, while in the negative outcome condition, the performance targets were not met. The dependent variable used to test hypotheses 1 and 2 is the performance rating (RATING) that the evaluator assigns to the manager's performance. The dependent variable used to test hypothesis 3 is the mean rating score (CONTROLLABILITY) for each performance measure (MEASURE - financial/non-financial).

Each participant completed a case using either a formula-based or a subjective performance evaluation system and having either a positive or negative outcome (above or below target). Each case included a set of four financial and four non-financial measures. The experimental design is described in Figure 1.

Each of the participants rated the manager's performance in the case on four scales ranging from 1 (worst) to 10 (best). A performance rating of "1" indicates a "very poor" performance. In contrast, a score of a "10" would signify a "very good" performance evaluation. The scales have been used in a prior outcome effect study (Ghosh, 2005).

Perceived CONTROLLABILITY was measured using a 7-point Likert scale<sup>1</sup>. Each participant rated eight common retail manager performance measures with respect to their perceived controllability, ranging from 1 "strongly disagree" to 7 "strongly agree". Seven of the eight measures (all except "customer satisfaction") are utilized by a large American home improvement chain for evaluating their store managers. Utilizing industry specific measures was intended to increase the applicability of the study's results to practice. The financial measures include "Sales versus Budget", "Sales versus Last Year", "Gross Profit versus Budget", and "Gross Profit versus Last Year". The set of non-financial measures consists of "Customer Satisfaction", "Employee Satisfaction", "Percentage of Stock-outs", and "Average Tenure of Sales Personnel."

### 3.3. THE INSTRUMENT AND PILOT TESTING

#### 3.3.1 The Instrument

The basic structure of the instrument was based on an experiment conducted by Ghosh (2005) that used a case from the retail industry. Other prior studies in this research area have also utilized a retail store case (Banker, et al., 2004; Lipe & Salterio, 2000, 2002). The retail industry is appropriate because it is easier to define how the activities of retail store managers affect the store's outcomes relative to manufacturing or service firm managers (Ghosh & Lusch,

---

<sup>1</sup> I chose a 7-point Likert scale (instead of a 10-point scale as in the former task) because this task requires an "agree/disagree" judgment. This judgment does not require a more detailed differentiation between answer choices.

2000). The case included eight performance measures on which the manager was evaluated in each of the four treatment groups. In addition, I included a paragraph in the case to partially explain some uncontrollable situational circumstances that could have contributed to the positive/negative outcome of the store performance. If the participants reduce their outcome focus, they should include that information in their performance evaluation.

To measure the performance rating, I utilized four questions that were employed previously by Ghosh (2005) and Tan and Lipe (1997). The first question asks each participant to “evaluate John Miller’s decision to make the changes in the store.” Therefore, it assesses the quality of the manager’s decision. The second question allocates credit or blame to the manager by asking “to what extent is this a case of success or failure for John Miller?” The remaining two questions, “to what extent is John Miller a good manager?” and “do you think that John Miller’s actions were consistent with what is expected from a good manager?” were included to consider rewards and penalties for the manager’s performance and to acquire information about the evaluator’s impressions of the manager. As in prior studies, the second and fourth questions were reverse-coded (Ghosh, 2005; Tan & Lipe, 1997). The four scores were then added to determine the overall evaluation score for the manager. This resulted in a dependent variable RATING range from 4 to 40.

### **3.3.2. Pilot Testing**

Before conducting the experiment with the MBA students, I ran three different pilot tests. The first pilot test was conducted using 56 advanced undergraduate students from an intermediate accounting class to examine the clarity of the instrumental materials and tasks. Each student completed a paper version of the case. They were not restricted with respect to

time. I was able to use the feedback and results of this initial pilot test to clarify several aspects of the instructions.

The second pilot test was conducted using an online survey tool. The primary purpose of this pilot test was to determine whether the experiment could be distributed and completed electronically, allowing me to sample from a broader participant population. This test was conducted using 139 additional advanced undergraduate students from four intermediate accounting classes, none of whom participated in the first round of pilot testing. Student comments regarding this experimental format indicated that the experiment was best delivered as a paper-based instrument because students had difficulty locating all available information about the manager's performance when conducting the performance judgment. In addition, the non-financial measure "Mystery Shopper Audit Rating" was substituted with the non-financial measure "Customer Satisfaction" because the students were not familiar with this audit rating measure. "Customer Satisfaction" has been used in prior studies (e.g., (Ghosh, 2005)) and its meaning is well understood by the participants.

Three accounting Ph.D. students participated in the third and final round of pilot testing. These three students were not familiar with the study, and they critically evaluated the experimental materials. Based on their comments, several adjustments were made to further improve the clarity of the instrument.

### 3.4. THE EXPERIMENT

The experiment took place in MBA classes during regularly scheduled class time. Each student was randomly assigned to one of the four versions of the case. The students received a paper-based package that included instructions and a set of typical retail industry performance

measures. In addition, they also received information about an investment decision that the store manager (John Miller) made during the year. Finally, each participant was provided with supplementary information about situational factors that influenced the actual outcomes of the performance measures. The supplementary information included descriptions of events that occurred during the evaluation period but were not controllable by the retail manager. In the “Above Target” condition, the evaluators were informed that some uncontrollable positive events increased the overall store performance. In the “Below Target” condition, they were informed that uncontrollable negative events decreased overall store performance (Appendix A). If the evaluators factored the supplementary information into their performance rating, then the rating should reflect a more neutral performance evaluation (toward a score of “5.5” on each scale), indicating that the outcome effect has been mitigated.

The participants in the formula-based condition received information about the outcomes of the eight pre-weighted financial and non-financial measures, as well as an overall performance variance measure ( $\sum_i [weight_i \times target\ variance_i]$ ). The participants in the subjective condition did not receive any information about the weighting of the measures, nor did they receive information about an overall performance variance measure. Based on this information, each participant evaluated the performance of the manager on four scales from 1 to 10. The participants were informed that the outcome of the performance evaluation would be used to decide whether the manager should receive a compensation bonus or not. Each participant completed one case and recorded a performance evaluation judgment. They were then asked to assess the controllability of each performance measure. Finally, each participant filled out a short demographic questionnaire. To assure comparability between the financial and non-financial measure conditions, the magnitude of positive and negative variances was held constant

for both treatments. Participants were able to take as much time as they needed to complete the experiment, but they usually took between 10 to 15 minutes.

## **CHAPTER 4**

### **RESULTS**

#### **4.1. INSTRUMENT RELIABILITY TEST AND PRINCIPAL COMPONENT ANALYSIS**

##### **4.1.1. Performance Evaluation Instrument Reliability (Cronbach's Alpha)**

To determine the appropriateness of the survey questions used for the manager's performance evaluation (the four performance evaluation questions), I calculated Cronbach's alpha to assure the reliability (internal consistency) of the survey. Cronbach's alpha is a common measure used to determine the internal consistency of items in a scale. It shows how well the items in a survey measure a latent construct. The construct that was examined in this study is the manager's performance. The participants answered four questions about that construct. Cronbach's alpha was 0.771. There is no commonly accepted threshold value for Cronbach's alpha, but Hair et al. (1998), George and Mallory (2003) and Robinson et al. (1991) suggest that a lower acceptable boundary is at least 0.7. This threshold has also been used in recent accounting studies (Deumes & Knechel, 2008; Smith & Hall, 2008). Therefore, Cronbach's alpha of 0.771 shows that the instrument has an acceptable level of reliability.

##### **4.1.2. Principal Component Analysis**

I also conducted a principal component analysis to validate the scales. The principal component analysis extracted a single component (eigenvalue > 1) and the factor loadings were high (0.752, 0.838, 0.789, and 0.724). These results suggest that the scales measure the same latent construct.

## 4.2. ANOVA ASSUMPTION TESTING

Before I tested the hypotheses, I made sure that the three assumptions required to use ANOVA were met. These assumptions are (1) independent observations (random sample), (2) normal distribution, and (3) equal (homogeneous) standard deviations among the subject groups (Ott & Longnecker, 2001). The first assumption, random samples, was achieved by randomly assigning the participants to the conditions.

To test the assumption of normality I performed Shapiro-Wilk tests and observed a normal distribution among the normal probability plots. The Shapiro-Wilk test examines whether the sample is comprised of normally distributed data (Shapiro & Wilk, 1965). Based on the Shapiro-Wilk tests (all  $p > 0.05$ ) and the normal probability plots, I was able to assume normality. Finally, Levene's test was utilized to determine the homogeneity of the standard deviations among the groups. If Levene's test is statistically significant, then the assumption of equal variance is not supported. The Levene statistic for this sample ( $p = 0.972$ ) indicated that the variances of the groups are homogenous. Based on the results of the assumption testing, the ANOVA is an appropriate tool to calculate the significance of mean differences among the four experimental groups.

## 4.3. RESULTS FOR HYPOTHESIS 1

H1 predicts that there will be an outcome effect in both the formula-based and the subjective system, even though situational factors were explicitly provided to partially explain the positive/negative variance from the budgeted target. To investigate this expectation, I employed a 2 (SYSTEM) x 2 (OUTCOME) analysis of variance (ANOVA) model using a priori contrasts. The ANOVA is an appropriate tool to compare differences in the dependent variable

that are caused by the independent variables (Whitley, 2002). The dependent variable was performance rating. The magnitude of the outcome effect was measured by calculating the mean difference between the performance ratings in the negative and positive outcome conditions (Figure 2).

The ANOVA model was statistically significant ( $p < 0.01$ ). Further, the ANOVA showed a statistically significant main effect for the OUTCOME variable ( $F = 29.51, p < 0.001$ ) (Table 4, Panel A). This indicates that the outcome effect exists independently from the type of system. However, to determine whether the hypothesis can be supported, I also need to verify that there is an outcome effect for each of the two performance systems. As the statistical analysis in Table 4, Panel B shows, there is a significant effect for both the subjective system ( $F = 10.39, p = 0.002$ ) and the formula-based system ( $F = 19.81, p < 0.001$ ). Based on this result, it appears that the participants mainly focused on the outcome information which was presented in the form of eight financial and non-financial measures. Therefore, Hypothesis 1 is supported.

#### 4.4. RESULTS FOR HYPOTHESIS 2

To test the second hypothesis, I employed the same ANOVA model used to test the first hypothesis (Table 4). I expected an outcome effect that was statistically significantly larger for participants in the formula-based evaluation system condition relative to the subjectively-oriented condition. Therefore, I predicted the average performance rating to be as follows (Figure 3): I expected the highest average performance rating in the “Formula-based System/Above Target” (FA) condition and the lowest average rating in the “Formula-based System/Below Target” (FB) condition because the participants would primarily focus on the overall variance measure and ignore additional environmental information. The actual average

scores were 27.95 (S.D. = 4.27) in the FA and 22.08 (S.D. = 5.18) in the FB condition. Additionally, I expected the “Subjective System/Above Target” (SA) condition to be the second highest average and the “Subjective System/Below Target” (SB) condition to be the third highest average rating because the participants would not have an overall measure to focus on. Therefore, they would take at least some additional situational information into consideration when making a performance judgment. The actual average scores in these two conditions were 26.73 (S.D. = 4.30) and 22.53 (S.D. = 4.57) respectively. The outcome effect score for each system was calculated by subtracting the average score for the negative outcome condition from the average score for the positive outcome condition (Figure 2). This resulted in an outcome effect score for the formula-based system of 5.87; the subjective system had an outcome effect score of 4.20. However, this difference is not statistically significant ( $F = 0.81, p = 0.370$ ).

To support this hypothesis, I expected an interaction effect between the OUTCOME and the type of SYSTEM (Figure 3, Table 4). This result would demonstrate that the magnitude of the outcome effect differs between the two performance evaluation systems. Based on the ANOVA results in Table 4, I can conclude that there is no statistically significant interaction effect between the SYSTEM and the OUTCOME ( $F = 0.81, p = 0.370$ ). Therefore, Hypothesis 2 is not supported.

#### 4.5. RESULTS FOR HYPOTHESIS 3

I employed a survey that examined the perceived controllability of eight (four non-financial and four financial) widely used retail store manager performance evaluation measures. The instrument contained one controllability question about each measure. The question about perceived measure controllability came from Ghosh's (2005) study.

I employed an ANOVA approach to compare the means of the two groups (financial versus non-financial measures) to examine whether perceived controllability differs between them. I expected that mean perceived controllability would be significantly higher for the set of non-financial measures when compared to the set of financial measures (H3). The set of financial measures had a mean of 4.46 (S.D. = 1.115) while the mean score of the set non-financial measures was 5.69 (S.D. = 1.121) (Table 5, Panel B). This difference in means is statistically significant ( $F = 238.16$ ,  $p < 0.01$ ) (Table 5, Panel A).

In addition, I ranked the individual measures based on their perceived controllability scores to determine whether the majority of non-financial measures were perceived to be more controllable than the financial measures and formed homogenous subsets using Scheffe's multiple comparison test (Table 6). A homogeneous subset consists of measures whose average controllability score did not significantly differ from each other ( $p < 0.05$ ). The four non-financial measures were all perceived to be more controllable than each of the financial measures. In addition, the four financial measures formed a subset of similar perceived measures. The four non-financial measures formed two subgroups. The participants perceived the controllability of "Customer Satisfaction" to be similar to the controllability of "Employee Satisfaction", and the controllability levels of "Percentage of Stock-outs" and "Average Tenure of Sales Personnel" were also almost identical.

Finally, a multiple comparison ("Scheffe's multiple comparison test") of the perceived controllability scores of the individual measures was conducted (Table 7). In this matrix, the means of the eight measures were compared to each other. The table shows that each of the four financial measures (Sales vs. Budget, Sales vs. Last Year, Gross Profit vs. Budget, and Gross Profit vs. Last Year) were perceived to be significantly less controllable than the non-financial

measures (Customer Satisfaction, Employee Satisfaction, Percentage of Stock-outs, and Average Tenure of Sales Personnel) ( $p < 0.05$ ). Based on the statistical results, Hypothesis 3 is supported.

The results in this multi-measure environment are consistent with Ghosh's (2005) findings in a single measure setting that common non-financial retail manager performance measures are perceived to be more controllable than more highly aggregated financial performance measures. However, one finding was inconsistent with Ghosh's (2005) study: the insignificant difference between the perceived controllability of customer satisfaction and employee satisfaction. In this study, the average perceived controllability scores of the two measures, 6.04 and 6.15 ( $p = 1.00$ ), were almost identical. One plausible explanation for this contradictory finding relates to the experimental design of the two studies. In Ghosh's (2005) experiment, a participant only evaluated the controllability of either customer satisfaction or employee satisfaction but not both (between-subject design). In this study, participants assessed the controllability of both measures one after the other (within-subject design). Therefore, some participants may have failed to conceptually differentiate between the two measures and assigned a similar level of controllability.

## 4.6. ADDITIONAL ANALYSES

### **4.6.1. Hypothesis Testing Excluding Part-time MBA Students**

I conducted additional analyses to examine why the results for Hypothesis 2 were not statistically significant. First, the performance evaluation rating variance was large (S.D. = 4.61). The large variance in the dependent variable is a major factor in the low power of this study (0.223). This may have contributed to the non-significant findings for H2. Although the power seems to be extremely low, prior research has shown that behavioral studies often have

low power. For example, Sedlmeier and Gigerenzer (1989) found that studies that were published in the *Journal of Abnormal Psychology* in 1984 had a median power of 0.37, assuming a medium effect size. In addition, a study by Borkowski et al. (2001) shows that behavioral accounting studies that were published in three accounting journals (*Issues in Accounting Education, Behavioral Research in Accounting*, and *Journal of Management Accounting Research*) between 1993 and 1997 had statistical power of 0.71 when a medium effect size was assumed, but only 0.23 when low effect size was assumed. If the power is significantly lower than 0.8, the sample size should be increased to increase the statistical power (Cohen, 1992).

Secondly, the five part-time MBA students in the “Subjective System/Below Target” treatment group provided significantly lower performance ratings than the EMBA ( $p = 0.018$ ) and PMBA ( $p < 0.01$ ) students in the same treatment group, which could have influenced my results. Although Grubbs’ test revealed that the individual scores did not meet the criteria to be classified as outliers, this can primarily be attributed to the significant amount of variance within the treatment group ( $S.D. = 4.57$ ). Therefore, I conducted an additional analysis that excluded all part-time MBA students from the sample and then tested each of the three hypotheses with the reduced sample<sup>2</sup>. To examine the effect that inclusion of the five part-time MBA students from the “Subjective System/Below Target” group had on the results for H2, I conducted a secondary analysis that excluded those five part-time MBA students from the sample.

The first analysis (wherein the entire part-time MBA group was excluded from the sample) yielded a statistically significant OUTCOME variable ( $F = 19.83, p < 0.01$ ). This indicates that the outcome effect is present when both types of system are combined (Table 8, Panel A). Although there is still a significant main effect for the formula-based evaluation

---

<sup>2</sup> The remaining sample consisted of 77 EMBA and PMBA students.

system ( $F = 21.13$ ,  $p < 0.01$ ), the main effect for the subjective system is only marginally significant ( $F = 2.95$ ,  $p = 0.09$ ) (Table 8, Panel B). Nonetheless, Hypothesis 1 is supported.

The second hypothesis proposed that the outcome effect would be larger in a formula-based evaluation system than in a subjective system. The reanalysis demonstrated that the interaction of OUTCOME x SYSTEM was statistically significant ( $F = 4.04$ ,  $p = 0.048$ ). This means that the magnitude of the outcome effect was significantly larger in the formula-based system than in the subjective system, supporting Hypothesis 2.

In the third hypothesis, I stated that non-financial measures would be perceived to be more controllable than the financial measures. The results of the PMBA and EMBA sample are very similar to the results of the full sample. There is a statistically significant difference in the means of the financial and non-financial measures ( $F = 169.59$ ,  $p < 0.01$ ) where the mean of the financial measures (4.546) is lower than the mean of the non-financial measures (5.696) (Table 9). In addition, the ranking of the measures is the same as the ranking with the full sample, in which the participants perceived the four non-financial measures to be more controllable than the financial measures (Table 10). The only slight difference with respect to the results of the full sample is that a fourth homogeneous subset group was formed that includes two financial measures (Sales versus Budget and Sales versus Last Year) and one non-financial measure (Stock-outs). However, the p-value of 0.079 indicates that the two financial measures and the one non-financial are not necessarily similar. Instead, because the cut-off value of non-homogenous groups was 0.05, they were treated as a homogenous group (Table 10). Based on these results, Hypothesis 3 is supported when using only the PMBA and EMBA student sample.

For the second supplemental analysis (excluding the five part-time MBA student ratings in the “Subjective System/Below Target” treatment group), the interaction of OUTCOME X

SYSTEM was marginally significant ( $p = 0.061$ ). This demonstrates that these five scores do have a substantial impact on the difference of the magnitude of the outcome effect between the two types of systems. In this case, the magnitude of the outcome effect in the formula-based system is significantly higher than the outcome effect in the subjective system. The analysis using the full sample did not reveal a statistically significant difference ( $p = 0.370$ ).

## CHAPTER 5

### DISCUSSION, IMPLICATIONS AND LIMITATIONS

#### 5.1. DISCUSSION AND IMPLICATIONS

The purpose of this study was to examine whether the type of performance evaluation system influences the occurrence and magnitude of the outcome effect. The research design included two performance evaluation systems that are commonly used, a subjective evaluation system and a formula-based system. The first hypothesis examined whether both systems are vulnerable to the outcome effect. The results showed that the outcome effect was present in both evaluation systems. The second hypothesis examined which of the two systems is more prone to the outcome effect, i.e., in which system is the outcome effect larger. The results revealed that the outcome effect was larger in the formula-based system relative to the subjective system, but this difference was not statistically significant. However, when the part-time MBA students were eliminated from the sample, the difference was statistically significant. Therefore, a major cause of the insignificant results for the overall sample were the low performance rating scores<sup>3</sup> that the part-time MBA students assigned to the store manager's performance in the "Subjective System/Below Target" condition. For the other three treatment groups, the part-time MBA students provided performance ratings that were not statistically different from the PMBA and EMBA groups.

I also conducted a power analysis which revealed that the statistical power is low (0.223). The main cause of the low power was the large standard deviation in performance ratings. This may have contributed to the non-significant findings for H2. Acquisition and analysis of additional data should reduce the large standard deviations and increase the power of the

---

<sup>3</sup> The performance evaluation ratings provided by the part-time MBA students were significantly ( $p < 0.05$ ) lower than the performance evaluation ratings provided by both the PMBA and the EMBA students.

statistical analysis. This would allow me to better determine whether there is a statistically significant difference in the magnitude of the outcome effect between the two types of performance evaluation systems.

Finally, this study analyzes the perceived controllability of several financial and non-financial performance measures from the retail industry. The results show that non-financial measures are perceived to be more controllable than financial measures when several different financial and non-financial measures are available to the evaluator. Since most performance evaluation systems seek to include measures that are highly controllable, companies should consider these findings when they select performance evaluation measures.

This study makes several contributions to the accounting literature. First, it extends and combines prior research on performance evaluation systems and the outcome effect. Ittner and Larcker (1998b) emphasized that the type of evaluation system employed may significantly impact performance evaluation. This study is the first to examine the behavioral impact of performance evaluation systems (formula-based vs. subjective) on the outcome effect. Additionally, this study extends Ghosh's (2005) study by examining the perceived controllability of multiple financial and non-financial measures using a within-subject design wherein each participant assesses the controllability of eight different measures. The results document that evaluators acknowledge the higher perceived controllability of non-financial measures in a multi-measure environment. Furthermore, this study informs the balanced scorecard literature. Future research could investigate different combinations of financial and non-financial measures in a balanced scorecard setting and determine their impact on the outcome effect.

Higher perceived measure controllability leads to an increase in the outcome effect, which may weaken evaluation objectivity. On the other hand, performance evaluation system measures should be highly controllable so that managerial performance can be evaluated appropriately. Future outcome effect research could focus on this outcome effect “dilemma” and could investigate ways to mitigate the outcome effect when evaluations are based on highly controllable measures.

This study also has practical implications. The results may assist managers in evaluating potential risks and biases for both types of performance evaluation systems. The purpose of this study is not to provide evidence about which performance evaluation system is superior. Instead, this study investigates the vulnerability to the outcome effect of two commonly used evaluation systems. Executives should consider these findings when designing and implementing a new evaluation system. They may use the findings from this study to train their managers about the risks that outcome focus may pose to the quality of performance evaluations, especially when their company employs a formula-based appraisal system. Since performance measures are supposed to be controllable, the results may provide an argument for managers to include non-financial measures (in addition to financial measures) in an evaluation system. The inclusion of highly controllable performance measures may provide a better representation of an evaluatee’s performance and effort. In addition, this study’s findings regarding the controllability of some performance measures may help managers to create their own “optimal mix” of financial and non-financial measures.

## 5.2. LIMITATIONS

Although I provide several contributions to the existing literature, this study is also subject to several limitations. First, the utilization of MBA students instead of real-world retail store managers limits the internal validity of the study. MBA students come from different industries and have diverse backgrounds. They also have different levels of expertise with respect to the experimental setting. Secondly, the case utilizes a retail store setting and may not apply to manufacturing and service firms, limiting the external validity of the study. Finally, the participants only received limited information about the decision-making process of the evaluatee. Perhaps real world evaluators would include more information in their assessment if they had more information available. However, in an experimental setting, one can only provide a limited amount of information because a participant is limited with respect to the number of cues s/he can process without experiencing information overload.

## References

- Anderson, E. W., Fornell, W., & Lehmann, D. R. (1994). Customer Satisfaction, Market Share, and Profitability: Findings from Sweden. *Journal of Marketing*, 58(3), 53-66.
- Anderson, E. W., Fornell, W., & Rust, R. (1997). Customer Satisfaction, productivity, and profitability: Differences between goods and services. *Marketing Science*, 16, 129-145.
- Ashton, R. H., & Kramer, S. S. (1980). Students As Surrogates in Behavioral Accounting Research: Some Evidence. *Journal of Accounting Research*, 18(1), 1-15.
- Baker, G., Gibbons, R., & Murphy, K. J. (1994). Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics*, 109, 1125-1156.
- Balkcom, J. E., Ittner, C. D., & Larcker, D. F. (1997). Strategic performance measurement: Lessons learned and future directions. *Journal of Strategic Performance Measurement* 1(2), 22-32.
- Banker, R. D., Chang, H., & Pizzini, M. J. (2004). The Balanced Scorecard: Judgmental Effects of Performance Measures Linked to Strategy. *The Accounting Review*, 79(1), 1-23.
- Banker, R. D., & Datar, S. M. (1989). Sensitivity, Precision, and Linear Aggregation of Signals for Performance Evaluation. *Journal of Accounting Research*, 27(1), 21-39.
- Banker, R. D., Potter, G., & Schroeder, R. G. (1993). Reporting manufacturing performance measures to workers: An empirical study. *Journal of Management Accounting Research*, 5, 33-55.
- Banker, R. D., Potter, G., & Srinivasan, D. (2000). An empirical investigation of an incentive plan that includes nonfinancial performance measures. *The Accounting Review*, 75(1), 65-92.
- Borkowski, S. C., Welsh, M. J., & Zhang, Q. M. (2001). An Analysis of Statistical Power in Behavioral Accounting Research. *Behavioral Research in Accounting*, 13, 63-84.
- Brancato, C. K. (1995). *New performance measures - A research report*. New York, NY: The Conference Board.
- Brown, C. E., & Solomon, I. (1987). Effects of Outcome Information on Evaluations of Managerial Decisions. *The Accounting Review*, 62(3), 564-577.
- Brown, C. E., & Solomon, I. (1993). An Experimental Investigation of Explanations for Outcome Effects on Appraisals of Capital-Budgeting Decisions. *Contemporary Accounting Research*, 10(1), 83-111.
- Bushman, R. M., Indjejikian, R. J., & Smith, A. (1996). The role of individual performance evaluation. *Journal of Accounting and Economics*, 21, 161-193.

- Clarkson, P., Emby, C., & Watt, V. (2002). Debiasing the Outcome Effect: The Role of Instructions in an Audit Litigation Setting. *Auditing*, 21(2), 7-20.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112(1), 155-159.
- de Villiers, C. J. (2002). The effect of attribution on perceptions of managers' performance. *Meditari Accountancy Research*, 10, 53-70.
- DeBusk, G. K., Brown, R. M., & Killough, L. N. (2003). Components and relative weights in utilization of dashboard measurement systems like the Balanced Scorecard. *The British Accounting Review* 35, 215-231.
- Deumes, R., & Knechel, W. R. (2008). Economic Incentives for Voluntary Reporting on Internal Risk Management and Control Systems. *Auditing*, 27(1), 35-66.
- Emby, C., Gelardi, A. M. G., & Lowe, D. J. (2002). A research note on the influence of outcome knowledge on audit partners' judgments. *Behavioral Research in Accounting*, 14, 87-103.
- Feltham, G. A., & Xie, J. (1994). Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relationship. *The Accounting Review*, 69(3), 429-453.
- Ferris, G. R., Judge, T. A., Rowland, K. M., & Fitzgibbons, D. E. (1994). Subordinate influence and the performance evaluation process: Test of a model. *Organizational Behavior and Human Decision Processes*, 58, 101-135.
- Ferris, G. R., Munyon, T. P., Basik, K., & Buckley, M. R. (2008). The performance evaluation context: Social, emotional, cognitive, political, and relationship components *Human Resource Management Review*, 18, 146-163.
- Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288-299.
- Fisher, J. (1995). Use of nonfinancial performance measures. In S. M. Young (Ed.), *Readings in Management Accounting* (pp. 329-335). Englewood Cliffs, NJ: Prentice Hall.
- Fisher, J., & Selling, T. I. (1993). The Outcome Effect in Performance Evaluation: Decision Process Observability and Consensus. *Behavioral Research in Accounting*, 5, 58-77.
- Frederickson, J. R., Peffer, S. A., & Pratt, J. (1999). Performance Evaluation Judgments: Effects of Prior Experience Under Different Performance Evaluation Schemes and Feedback Frequencies. *Journal of Accounting Research*, 37(1), 151-165.
- George, D., & Mallory, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Ghosh, D. (2005). Alternative measures of managers' performance, controllability, and the outcome effect. *Behavioral Research in Accounting*, 17, 55-70.

- Ghosh, D., & Lusch, R. (2000). Outcome effect, controllability, and performance evaluation on managers: some field evidence from multi-outlet businesses. *Accounting, Organizations and Society*, 25, 411-425.
- Ghosh, D., & Ray, M. (2000). Evaluating Managerial Performance: Mitigating the "Outcome Effect". *Journal of Managerial Issues*, 2, 247-260.
- Hair, J. E., Anderson, R. E., Tatham, R. L. and Black, W. C. (1998). *Multivariate Data Analysis* (5th ed.). Upper Saddle River, New Jersey: Prentice Hall.
- Hawkins, S., & Hastie, R. (1990). Hindsight: Biased judgment of past events after the outcomes are known. *Psychological Bulletin*, 107(3), 311-327.
- Haynes, C. M., & Kachelmeier, S. J. (1998). The effects of accounting contexts on accounting decisions: A synthesis of cognitive and economic perspectives in accounting experimentation. *Journal of Accounting Literature*, 17, 97-136.
- Haywood, M. E. (2001). *The balanced scorecard and decision strategies: What roles do outcome effects, framing biases and explicit weights play in the balancing act?* , The University of Georgia, Athens, GA.
- Hemmer, T. (1996). On the design and choice of "modern" management accounting measures. *Journal of Management Accounting Research*, 8, 87-116.
- Hershey, J. C., & Baron, J. (1992). Judgment by outcomes: When is it justified? *Organizational Behavior and Human Decision Processes*, 53, 89-93.
- Hershey, J. C., & Baron, J. (1995). REJOINDER Judgment by outcomes: When is it warranted? *Organizational Behavior and Human Decision Processes*, 62(1), 127.
- Holmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics*, 10(1), 74-91.
- Ittner, C. D., & Larcker, D. F. (1998a). Are Nonfinancial Measures Leading Indicators of Financial Performance? An Analysis of Customer Satisfaction. *Journal of Accounting Research*, 36(Supplement), 1-35.
- Ittner, C. D., & Larcker, D. F. (1998b). Innovations in Performance Measurement: Trends and Research Implications. *Journal of Management Accounting Research* 10, 205-238.
- Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the Weighting of Performance Measures: Evidence from a Balanced Scorecard. *The Accounting Review*, 78(3), 725-758.
- Ittner, C. D., Larcker, D. F., & Rajan, M. V. (1997). The choice of performance measures in annual bonus contracts. *The Accounting Review*, 72(2), 231-255.

- Ittner, C. D., Larcker, D. F., & Randall, T. (2003). Performance implications of strategic performance measurement in financial services firms. *Accounting, Organizations and Society*, 28, 715–741.
- Johnson, E. J., & Payne, J. W. (1985). Effort and Accuracy in Choice. *Management Science*, 31(4), 395-414.
- Jones, K. T., & Chen, C. C. (2005). The effect of audit outcomes on evaluators' perceptions. *Managerial Auditing Journal*, 20(1), 5-18.
- Kadous, K. K. (2001). Improving juror evaluation of auditors in negligence cases. *Contemporary Accounting Research*, 18(3), 425-444.
- Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard - Measures that drive performance. *Harvard Business Review*, 70(1), 71-79.
- Kaplan, R. S., & Norton, D. P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press.
- Kaplan, R. S., & Norton, D. P. (2001a). Transforming the Balanced Scorecard from Performance Measurement to Strategic Management: Part 1. *Accounting Horizons*, 15(1), 87-104.
- Kaplan, R. S., & Norton, D. P. (2001b). Transforming the Balanced Scorecard from Performance Measurement to Strategic Management: Part 2. *Accounting Horizons*, 15(2), 147-160.
- Kelley, H. H., & Michela , J. L. (1980). Attribution Theory and Research. *Annual Review of Psychology*, 31(January), 457-501.
- Kelly, K. O. (2007). Feedback and Incentives on Nonfinancial Value Drivers: Effects on Managerial Decision Making. *Contemporary Accounting Research*, 24(2), 523–556.
- Lipe, M. (1993). Analyzing the variance investigation decision: The effects of outcomes, mental accounting, and framing. *The Accounting Review*, 68(4), 748-764.
- Lipe, M. G., & Salterio, S. (2000). The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review*, 75(3), 283-298.
- Lipe, M. G., & Salterio, S. (2002). A note on the judgmental effects of the balanced scorecard's information organization. *Accounting, Organizations and Society* 27, 531-540.
- Lipshitz, R. (1989). "Either a Medal or a Corporal": The Effects of Success and failure on the Evaluation of Decision Making and Decision Makers. *Organizational Behavior and Human Decision Processes*, 44, 380-395.
- Merchant, K. A. (2007). Evaluating General Managers' Performances. *Strategic Finance*, 88(11), 12, 14, 59-61.

- Mitchell, T., & Kalb, L. (1981). Effects of outcome knowledge and outcome valence on supervisors' evaluations. *Journal of Applied Psychology*, 66(5), 604-612.
- Moers, F. (2005). Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations and Society*, 30, 67-80.
- Moers, F. (2006). Performance Measure Properties and Delegation. *The Accounting Review*, 81(4), 897-924.
- Ott, R. L., & Longnecker, M. (2001). *An Introduction to Statistical Methods and Data Analysis* (5th ed.). Pacific Grove, CA: Duxbury.
- Prendergast, C. (2002). The tenuous trade-off between risk and incentives. *Journal of Political Economy*, 110, 1071-1102.
- Prendergast, C., & Topel, R. (1993). Discretion and bias in performance evaluation. *European Economic Review*, 37, 355-365.
- Prendergast, C., & Topel, R. H. (1996). Favoritism in Organizations. *The Journal of Political Economy*, 104(5), 958-978.
- Roberts, M. L., Albright, T. L., & Hibbets, E. R. (2004). Debiasing Balanced Scorecard Evaluations. *Behavioral Research in Accounting*, 16, 75-88.
- Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). Criteria for scale selection and evaluation *Measures of personality and social psychological attitudes* (pp. 1-16). San Diego, CA: Academic Press.
- Rose, J. M. (2004). Performance Evaluations Based on Financial Information: How Do Managers use Situational Information? *Managerial Finance*, 30(6), 46-65.
- Sanbonmatsu, D. M., Kardes, F. R., Posavac, S. S., & Houghton, D. C. (1997). Contextual Influences on Judgment Based on Limited Information. *Organizational Behavior and Human Decision Processes*, 69(3), 251–264.
- Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of the studies. *Psychological Bulletin*, 105, 309-316.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611.
- Smith, D., & Hall, M. (2008). An Empirical Examination of a Three-Component Model of Professional Commitment among Public Accountants. *Behavioral Research in Accounting*, 20(1), 75-92.
- Smith, R. E., & Wright, W. F. (2004). Determinants of Customer Loyalty and Financial Performance. *Journal of Management Accounting Research*, 16, 183-205.

- Speier, C. (1998). *Using Aggregated Data Under Time Pressure: A Mechanism for Coping with Information Overload*. Paper presented at the Thirty-First Annual Hawaii International Conference on System Sciences.
- Tan, H. T., & Lipe, M. G. (1997). Outcome Effects: The Impact of Decision Process and Outcome Controllability. *Journal of Behavioral Decision Making*, 10, 315-325.
- Whitley, B. E. (2002). *Principles of Research in Behavioral Science* (2nd ed.). New York, NY: McGraw-Hill.
- Wier, B., Stone, D. N., & Hunton, J. E. (2002). Promotion and Performance Evaluation of Managerial Accountants. *Journal of Management Accounting Research*, 14, 189-208.

## **FIGURES**

---

**FIGURE 1**  
**Experimental Design**

---

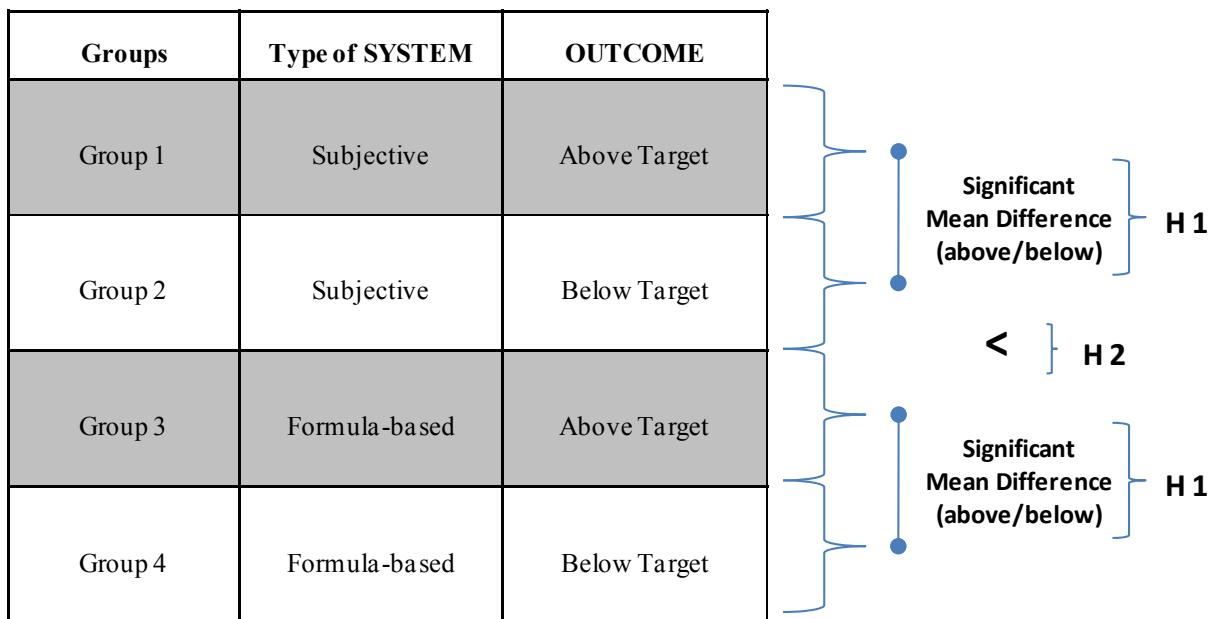
Groups	Type of SYSTEM	OUTCOME
Group 1	Subjective	Above Target
Group 2	Subjective	Below Target
Group 3	Formula-based	Above Target
Group 4	Formula-based	Below Target

---

---

**FIGURE 2**  
**Measurement of the Outcome Effect**

---

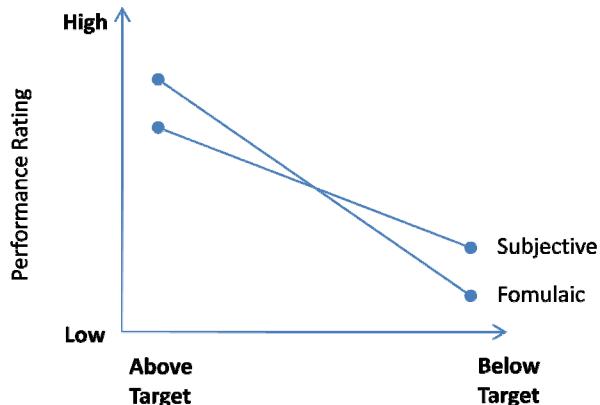


---

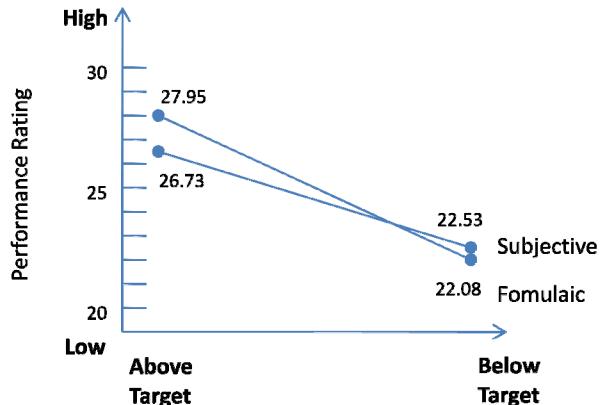
**FIGURE 3**  
Results for Hypothesis 2

---

Expected Results Hypothesis 2



Actual Results Hypothesis 2<sup>\*</sup>



---

\* The interaction is not statistically significant ( $p = 0.37$ ).

## **TABLES**

---

**TABLE 1**  
**Demographic Data**

---



---

Panel A: Continuous Variables (in Years)

---

	<u>Mean</u>	<u>Standard Deviation</u>
<b>Work Experience*</b>	9.41	5.51
<b>Management Experience**</b>	2.34	3.40
<b>Evaluator Experience**</b>	2.19	3.20
<b>Evaluatee Experience*</b>	7.30	5.20

---

Panel B: Discrete Variables

---

	<u>N</u>	<u>%</u>
<b>Age</b>		
<b>23-30</b>	53	54
<b>31-40</b>	37	37
<b>41-50</b>	9	9
<b>Education*</b>		
<b>4-year</b>	68	69
<b>Master</b>	29	30
<b>Doctorate</b>	1	1
<b>Gender</b>		
<b>Female</b>	30	30
<b>Male</b>	69	70

---

\* One participant did not provide information on this attribute.

\*\* Two participants did not provide information on this attribute.

---

**Table 2**  
**Comparison of Demographic Data among Treatment Groups<sup>1</sup>**

---

Panel A: Means (in Years) per Treatment Group (Continuous Variables)				
	Subjective System/ Above Target	Subjective System/ Below Target	Formula-based System/ Above Target	Formula-based System/ Below Target
Work Experience	8.28	8.42	9.07	11.69
Management Experience	1.29	2.78	2.80	2.50
Evaluator Experience	1.58	1.96	2.17	3.00
Evaluatee Experience	6.67	6.16	6.89	9.33

Panel B: Frequency per Treatment Group (Categorical Variables)				
	Subjective System/ Above Target	Subjective System/ Below Target	Formula-based System/ Above Target	Formula-based System/ Below Target
Age				
23-30	14	15	14	10
31-40	9	9	6	13
41-50	2	1	3	3
Education				
4-year	14	16	19	20
Master	10	8	5	6
Doctorate	0	1	0	0
Gender				
Female	7	8	5	10
Male	18	17	18	16

---

<sup>1</sup>There are no statistically significant differences among any of the treatment groups for any demographic variable (at a level of  $p < 0.05$ ) based on Scheffe's Multiple Comparison Test.

---

**TABLE 3**  
**Final Sample**

---

PMBA Students	67
EMBA Students	13
PartMBA Students	26
<b>Initial Sample – MBA Students</b>	<b>106</b>
- Participants that did not recognize reverse-coded scales	5
- Full Time Student	1
- Incomplete Survey	1
<b>Final Sample</b>	<b>99</b>

---

---

**Table 4**  
**ANOVA Results**

---



---

**Panel A: ANOVA for the Outcome Effect**

---

<i>Source of Variation</i>	<i>df</i>	<i>Sum of Squares</i>	<i>F</i>	<i>p</i>
SYSTEM	1	3.591	0.17	0.682
OUTCOME	1	626.031	29.51	<0.001
SYSTEM * OUTCOME	1	17.196	0.81	0.370
ERROR	95	2015.621		

---

**Panel B: Contrasts to Determine the Impact of the Two Types of Systems on the Outcome Effect**

---

<i>Contrast</i>	<i>df</i>	<i>Sum of Squares</i>	<i>F</i>	<i>p</i>
SUBJECTIVE SYSTEM	1	220.500	10.39	0.002
FORMULA SYSTEM	1	420.333	19.81	<0.001
SYSTEM * OUTCOME	1	17.196	0.81	0.370

---

**Panel C: Performance Evaluation Ratings**

---

<u>Treatment Groups</u>	<u>Mean</u>	<u>Std. Deviation</u>	<u>N</u>
Subjective System/Above Target	26.73	4.30	25
Subjective System/Below Target	22.53	4.57	25
Formula-based System/Above Target	27.95	4.27	23
Formula-based System/Below Target	22.08	5.18	26

---



---

---

**Table 5**  
**Perceived Controllability of Financial and Non-Financial Measures**

---

**Panel A: ANOVA Model for Perceived Controllability of Financial and Non-Financial Measures**

---

<i>Source of Variation</i>	<i>df</i>	<i>Sum of Squares</i>	<i>Mean Square</i>	<i>F</i>	<i>p</i>
Between Groups	1	297.606	297.606	238.16	<0.001
Within Groups	787	983.453	1.250		
Total	788	1281.060			

---

**Panel B: Perceived Controllability Scores**

---

<i>Type of Measure</i>	<i>Mean</i>	<i>S.D.</i>	<i>N</i>
Financial	4.460	1.115	395*
Non-Financial	5.688	1.121	394**

---

\* One participant did not evaluate the perceived controllability of one financial measure.

\*\* Two participants did not evaluate the perceived controllability of one non-financial measure.

**Table 6**  
**Perceived Controllability: Ranking of Measures and Description of Homogeneous Subsets**  
**(Subset for Alpha = 0.05)<sup>1</sup>**

<b>Rank</b>	<b>Measure</b>	<b>N</b>	<b>Subset 1</b>	<b>Subset 2</b>	<b>Subset 3</b>	<b>Type of Measure</b>
1)	Employee Satisfaction	99	6.15			Non-Financial
2)	Customer Satisfaction	98*	6.04			Non-Financial
3)	Tenure	98**		5.30		Non-Financial
4)	Stock-outs	99		5.26		Non-Financial
5)	Sales vs. Last Year	98***			4.60	Financial
6)	Sales vs. Budget	99			4.58	Financial
7)	Gross Profit vs. Budget	99			4.35	Financial
8)	Gross Profit vs. Last Year	99			4.31	Financial
<i>Significance</i>			1.000	1.000	0.811	

<sup>1</sup>Based on Scheffe's Multiple Comparison Test.

\* One participant did not evaluate the controllability of the measure "Customer Satisfaction."

\*\* One participant did not evaluate the controllability of the measure "Average Tenure of Sales Personnel."

\*\*\* One participant did not evaluate the controllability of the measure "Sales versus Last Year."

**Table 7**  
**Multiple Comparisons of Perceived Controllability Means for all Eight Measures<sup>1</sup>**

Type of Measure (Mean)	Compared Measure	Mean	p-value
Employee Satisfaction (6.15)	Sales vs. Budget	4.58	<0.001
	Sales vs. Last Year	4.60	<0.001
	Customer Satisfaction	6.04	1.000
	GP vs. Budget	4.35	<0.001
	GP vs. Last Year	4.31	<0.001
	Stock-outs	5.26	<0.001
	Tenure	5.30	<0.001
Customer Satisfaction (6.04)	Sales vs. Budget	4.58	<0.001
	Sales vs. Last Year	4.60	<0.001
	Employee Satisfaction	6.15	1.000
	GP vs. Budget	4.35	<0.001
	GP vs. Last Year	4.31	<0.001
	Stock-outs	5.26	0.001
	Tenure	5.30	0.002
Tenure (5.30)	Sales vs. Budget	4.58	0.003
	Sales vs. Last Year	4.60	0.005
	Customer Satisfaction	6.04	0.002
	Employee Satisfaction	6.15	<0.001
	GP vs. Budget	4.35	<0.001
	GP vs. Last Year	4.35	<0.001
	Stock-outs	5.26	1.000
Stock-outs (5.26)	Sales vs. Budget	4.58	0.007
	Sales vs. Last Year	4.60	0.012
	Customer Satisfaction	6.04	0.001
	Employee Satisfaction	6.15	<0.001
	GP vs. Budget	4.35	<0.001
	GP vs. Last Year	4.35	<0.001
	Tenure	5.30	1.000
Sales vs. Last Year (4.60)	Sales vs. Budget	4.58	1.000
	Customer Satisfaction	6.04	<0.001
	Employee Satisfaction	6.15	<0.001
	GP vs. Budget	4.35	0.914
	GP vs. Last Year	4.31	0.812
	Stock-outs	5.26	0.012
	Tenure	5.30	0.005
Sales vs. Budget (4.58)	Sales vs. Last Year	4.60	1.000
	Customer Satisfaction	6.04	<0.001
	Employee Satisfaction	6.15	<0.001
	GP vs. Budget	4.35	0.945
	GP vs. Last Year	4.31	0.864
	Stock-outs	5.26	0.007
	Tenure	5.30	0.003
GP vs. Budget (4.35)	Sales vs. Budget	4.58	0.945
	Sales vs. Last Year	4.60	0.914
	Customer Satisfaction	6.04	<0.001
	Employee Satisfaction	6.15	<0.001
	GP vs. Last Year	4.31	1.000
	Stock-outs	5.26	<0.001
	Tenure	5.30	<0.001
GP vs. Last Year (4.31)	Sales vs. Budget	4.58	0.864
	Sales vs. Last Year	4.60	0.812
	Customer Satisfaction	6.04	<0.001
	Employee Satisfaction	6.15	<0.001
	GP vs. Budget	4.35	1.000
	Stock-outs	5.26	<0.001
	Tenure	5.30	<0.001

<sup>1</sup>Based on Scheffe's Multiple Comparison Test.

---

**Table 8**  
**ANOVA Results for PMBA and EMBA Students**

---

**Panel A: ANOVA for the Outcome Effect**

---

<i>Source of Variation</i>	<i>df</i>	<i>Sum of Squares</i>	<i>F</i>	<i>p</i>
SYSTEM	1	4.101	0.22	0.642
OUTCOME	1	372.429	19.83	<0.001
SYSTEM * OUTCOME	1	75.918	4.04	0.048
ERROR	73	1370.929		

---

**Panel B: Contrasts to Determine the Impact of the Two Types of Systems on the Outcome Effect**

---

<i>Contrast</i>	<i>df</i>	<i>Sum of Squares</i>	<i>F</i>	<i>p</i>
SUBJECTIVE SYSTEM	1	55.393	2.95	0.090
FORMULA SYSTEM	1	396.849	21.13	<0.001
SYSTEM * OUTCOME	1	75.918	4.04	0.048

---

**Panel C: Performance Evaluation Ratings**

---

<u>Treatment Groups</u>	<u>Mean</u>	<u>Std. Deviation</u>	<u>N</u>
Subjective System/Above Target	26.68	3.76	18
Subjective System/Below Target	24.26	3.01	20
Formula-based System/Above Target	28.21	4.49	18
Formula-based System/Below Target	21.81	5.55	21

---

---

**Table 9**  
**Perceived Controllability of Financial versus Non-Financial Measures**  
**for PMBA and EMBA Students**

---



---

Panel A: ANOVA Model for Perceived Controllability of Financial and Non-Financial Measures

---

<i>Source of Variation</i>	<i>df</i>	<i>Sum of Squares</i>	<i>Mean Square</i>	<i>F</i>	<i>p</i>
Between Groups	1	203.222	203.222	169.59	<0.001
Within Groups	612	733.349	1.198		
Total	613	936.571			

---



---

Panel B: Perceived Controllability Scores

---

<i>Type of Measure</i>	<i>Mean</i>	<i>S.D.</i>	<i>N</i>
Financial	4.546	1.094	308
Non-Financial	5.696	1.096	306*

---



---

\* Two participants did not evaluate the perceived controllability of one non-financial measure.

**Table 10**  
**Perceived Controllability: Ranking of Measures and Description of Homogeneous Subsets**  
**for PMBA and EMBA Students<sup>1</sup>**  
**(Subset for Alpha = 0.05)**

<b>Rank</b>	<b>Measure</b>	<b>N</b>	<b>Subset 1</b>	<b>Subset 2</b>	<b>Subset 3</b>	<b>Subset 4</b>	<b>Type of Measure</b>
1)	Employee Satisfaction	77	6.14				Non-Financial
2)	Customer Satisfaction	76*	6.03				Non-Financial
3)	Tenure	76**		5.36			Non-Financial
4)	Stock-outs	77		5.26	5.26		Non-Financial
5)	Sales vs. Last Year	77			4.65	4.65	Financial
6)	Sales vs. Budget	77			4.65	4.65	Financial
7)	Gross Profit vs. Budget	77				4.47	Financial
8)	Gross Profit vs. Last Year	77				4.41	Financial
<i>Significance</i>			1.000	1.000	0.079	0.959	

<sup>1</sup>Based on Scheffe's Multiple Comparison Test.

\* One participant did not evaluate the controllability of the measure "Customer Satisfaction."

\*\* One participant did not evaluate the controllability of the measure "Average Tenure of Sales Personnel."

**APPENDIX A:**  
**The Instrument**

## **ALL CONDITIONS**

### **INSTRUCTIONS**

**This study examines individuals' decisions. There are no right or wrong answers to any of the questions. We are just interested in your opinion. Please turn the page and follow the directions provided in the booklet. This study should take approximately 15 minutes to complete.**

**Your participation is completely voluntary, and you may cease to participate at any time. Your name will not appear on any of the experimental materials and will not be traceable to you. Your answers will only be used for research purposes.**

**THANK YOU FOR PARTICIPATING IN THIS PROJECT.**

## SUBJECTIVE/ABOVE TARGET CONDITION

### The Company

ABC Shoes operates 250 retail shoe stores in the U.S. They offer a wide range of shoes. Each store manager is evaluated annually by the regional manager. You are an ABC Shoes regional manager. Therefore, it is your responsibility to evaluate a store manager's performance. The yearly performance evaluation provides the basis for the store manager's compensation, promotions and bonus. Below is the information that you should consider when completing this evaluation.

### The Store Manager

John Miller is the store manager at one of ABC Shoes' retail stores. The store has performed below the company's average in recent years. Miller began to investigate potential reasons and identified a number of plausible explanations. Miller had some ideas to make some changes/modifications in the store to improve the functionality of the store layout and overall store appearance. Miller wrote an assessment report to the regional headquarters requesting a one-time \$20,000 fund to make the changes. This was a significant amount because Miller's total yearly budget for investments is \$25,000. Miller's request was granted and he made the proposed changes. Miller was able to exceed store targets, which were defined at the beginning of the year.

### Additional Information

Several events occurred during the year that helped the store to achieve its goals. One of the major brands that ABC Stores sells introduced a new shoe design that doubled sales from this brand. Further, a corporation wide TV advertisement was launched that increased sales. Finally, the market for ABC Shoes' merchandise grew at a faster rate than in previous years.

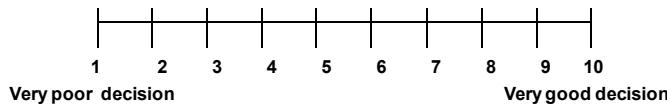
It is the end of the fiscal year and you receive the following information about John Miller's performance measures before you conduct his performance evaluation:

### Performance Measures

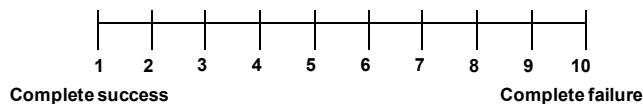
Measure	Target	Actual	Above/Below Target
Sales versus Budget (in '000)	12,200	12,902	+5.8%
Sales versus Last Year (in '000)	11,932	12,902	+8.1%
Customer Satisfaction*	84.00%	87.09%	+3.7%
Employee Satisfaction**	82.00%	84.74%	+3.3%
Gross Profit Margin versus Budget	27.10%	28.71%	+5.9%
Gross Profit Margin versus Last Year	27.00%	28.71%	+6.3%
Percentage of Stock-outs***	2.50%	2.41%	+3.5%
Average Tenure of Sales Personnel****	2.30	2.38	+3.5%

**Indicate your initial performance evaluation for John Miller below by placing an X on the scale at the point that corresponds to your evaluation.**

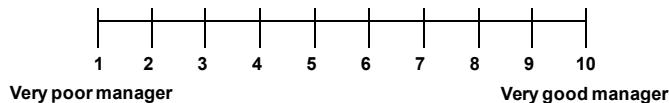
1. Evaluate John Miller's decision to make the changes in the store.



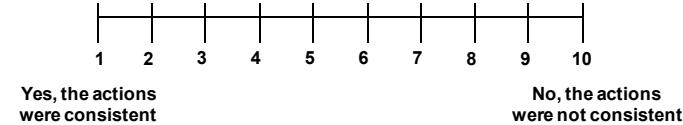
2. To what extent is this a case of success or failure for John Miller?



3. To what extent is John Miller a good manager?



4. Do you think that John Miller's actions were consistent with what is expected from a good manager?



\* - based on a customer survey from 0% (worst) to 100% (best)

\*\*- based on an employee survey from 0% (worst) to 100% (best)

\*\*\* - stock-out = running out of inventory/ failure to restock shelves

\*\*\*\* - average years that sales personnel stays in the firm

*Please continue to the next page*

## **SUBJECTIVE SYSTEM/ BELOW TARGET CONDITION**

### **The Company**

ABC Shoes operates 250 retail shoe stores in the U.S. They offer a wide range of shoes. Each store manager is evaluated annually by the regional manager. You are an ABC Shoes regional manager. Therefore, it is your responsibility to evaluate a store manager's performance. The yearly performance evaluation provides the basis for the store manager's compensation, promotions and bonus. Below is the information that you should consider when completing this evaluation.

### **The Store Manager**

John Miller is the store manager at one of ABC Shoes' retail stores. The store has performed below the company's average in recent years. Miller began to investigate potential reasons and identified a number of plausible explanations. Miller had some ideas to make some changes/modifications in the store to improve the functionality of the store layout and overall store appearance. Miller wrote an assessment report to the regional headquarters requesting a one-time \$20,000 fund to make the changes. This was a significant amount because Miller's total yearly budget for investments is \$25,000. Miller's request was granted and he made the proposed changes. Miller was not able to achieve store targets, which were defined at the beginning of the year.

### **Additional Information**

Several events occurred during the year that prevented the store from achieving its goals. One of the major brands that ABC Stores sells introduced a new shoe design that did not sell as well as had been forecasted. Further, a corporation wide TV advertisement was launched that customers did not like. Finally, the market for ABC Shoes' merchandise grew at a lower rate than in previous years.

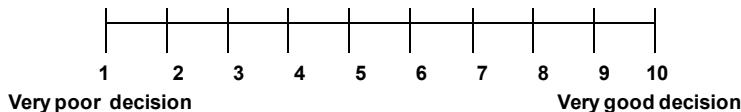
It is the end of the fiscal year and you receive the following information about John Miller's performance measures before you conduct his performance evaluation:

### Performance Measures

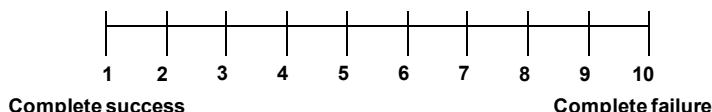
Measure	Target	Actual	Above/Below Target
Sales versus Budget (in '000)	12,200	11,498	-5.8%
Sales versus Last Year (in '000)	12,516	11,498	-8.1%
Customer Satisfaction*	84.00%	80.91%	-3.7%
Employee Satisfaction**	82.00%	79.26%	-3.3%
Gross Profit Margin versus Budget	27.10%	25.49%	-5.9%
Gross Profit Margin versus Last Year	27.21%	25.49%	-6.3%
Percentage of Stock-outs ***	2.50%	2.59%	-3.5%
Average Tenure of Sales Personnel****	2.30	2.22	-3.5%

**Indicate your initial performance evaluation for John Miller below by placing an X on the scale at the point that corresponds to your evaluation.**

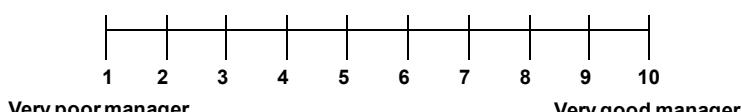
1. Evaluate John Miller's decision to make the changes in the store.



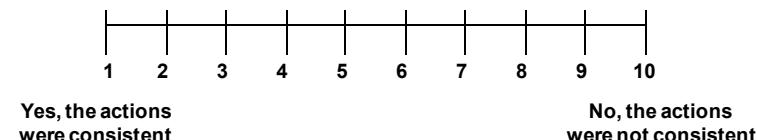
2. To what extent is this a case of success or failure for John Miller?



3. To what extent is John Miller a good manager?



4. Do you think that John Miller's actions were consistent with what is expected from a good manager?



\* - based on a customer survey from 0% (worst) to 100% (best)

\*\* - based on an employee survey from 0% (worst) to 100% (best)

\*\*\* - stock-out = running out of inventory/ failure to restock shelves

\*\*\*\* - average years that sales personnel stays in the firm

*Please continue to the next page*

## **FORMULA BASED-SYSTEM/ ABOVE TARGET CONDITION**

### **The Company**

ABC Shoes operates 250 retail shoe stores in the U.S. They offer a wide range of shoes. Each store manager is evaluated annually by the regional manager. You are an ABC Shoes regional manager. Therefore, it is your responsibility to evaluate a store manager's performance. The yearly performance evaluation provides the basis for the store manager's compensation, promotions and bonus. Below is the information that you should consider when completing this evaluation.

### **The Store Manager**

John Miller is the store manager at one of ABC Shoes' retail stores. The store has performed below the company's average in recent years. Miller began to investigate potential reasons and identified a number of plausible explanations. Miller had some ideas to make some changes/modifications in the store to improve the functionality of the store layout and overall store appearance. Miller wrote an assessment report to the regional headquarters requesting a one-time \$20,000 fund to make the changes. This was a significant amount because Miller's total yearly budget for investments is \$25,000. Miller's request was granted and he made the proposed changes. Miller was able to exceed store targets, which were defined at the beginning of the year.

### **Additional Information**

Several events occurred during the year that helped the store to achieve its goals. One of the major brands that ABC Stores sells introduced a new shoe design that doubled sales from this brand. Further, a corporation wide TV advertisement was launched that increased sales. Finally, the market for ABC Shoes' merchandise grew at a faster rate than in previous years.

It is the end of the fiscal year and you receive the following information about John Miller's performance measures before you conduct his performance evaluation:

## Performance Measures

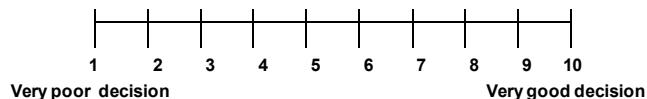
Measure	Target	Actual	Weight	Above/Below Target	Calculation
Sales versus Budget ('000)	12,200	12,902	20%	+5.8%	+1.15%
Sales versus Last Year ('000)	11,932	12,902	15%	+8.1%	+1.22%
Customer Satisfaction*	84.00%	87.09%	15%	+3.7%	+0.55%
Employee Satisfaction**	82.00%	84.74%	20%	+3.3%	+0.67%
Gross Profit Margin versus Budget	27.10%	28.71%	10%	+5.9%	+0.59%
Gross Profit Margin versus Last Year	27.00%	28.71%	5%	+6.3%	+0.32%
Percentage of Stock-outs***	2.50%	2.41%	10%	+3.5%	+0.35%
Average Tenure of Sales Personnel****	2.30	2.38	5%	+3.5%	+0.17%

**Overall Variance (above/below target): + 5.02%**

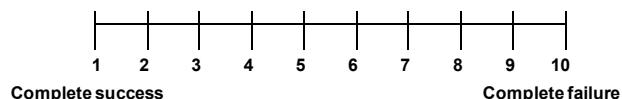
According to company policies the overall above/below target variance ( $\sum_i[\text{weight}_i \times \text{target variance}_i]$ ) was calculated before you make your final performance evaluation.

**Indicate your initial performance evaluation for John Miller below by placing an X on the scale at the point that corresponds to your evaluation.**

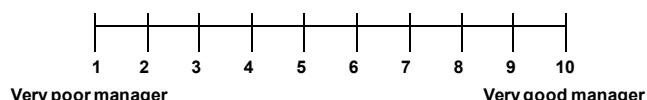
1. Evaluate John Miller's decision to make the changes in the store.



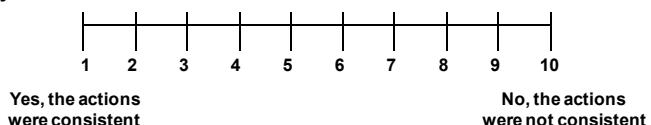
2. To what extent is this a case of success or failure for John Miller?



3. To what extent is John Miller a good manager?



4. Do you think that John Miller's actions were consistent with what is expected from a good manager?



\* - based on a customer survey from 0% (worst) to 100% (best)

\*\* - based on an employee survey from 0% (worst) to 100% (best)

\*\*\* - stock-out = running out of inventory/ failure to restock shelves

\*\*\*\* - average years that sales personnel stays in the firm

*Please continue to the next page*

## **FORMULA-BASED SYSTEM/BELOW TARGET CONDITION**

### **The Company**

ABC Shoes operates 250 retail shoe stores in the U.S. They offer a wide range of shoes. Each store manager is evaluated annually by the regional manager. You are an ABC Shoes regional manager. Therefore, it is your responsibility to evaluate a store manager's performance. The yearly performance evaluation provides the basis for the store manager's compensation, promotions and bonus. Below is the information that you should consider when completing this evaluation.

### **The Store Manager**

John Miller is the store manager at one of ABC Shoes' retail stores. The store has performed below the company's average in recent years. Miller began to investigate potential reasons and identified a number of plausible explanations. Miller had some ideas to make some changes/modifications in the store to improve the functionality of the store layout and overall store appearance. Miller wrote an assessment report to the regional headquarters requesting a one-time \$20,000 fund to make the changes. This was a significant amount because Miller's total yearly budget for investments is \$25,000. Miller's request was granted and he made the proposed changes. Miller was not able to achieve store targets, which were defined at the beginning of the year.

### **Additional Information**

Several events occurred during the year that prevented the store from achieving its goals. One of the major brands that ABC Stores sells introduced a new shoe design that did not sell as well as had been forecasted. Further, a corporation wide TV advertisement was launched that customers did not like. Finally, the market for ABC Shoes' merchandise grew at a lower rate than in previous years.

It is the end of the fiscal year and you receive the following information about John Miller's performance measures before you conduct his performance evaluation:

### Performance Measures

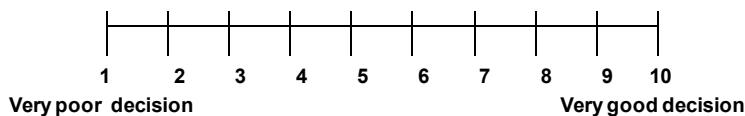
Measure	Target	Actual	Weight	Above/Below Target	Calculation
Sales versus Budget ('000)	12,200	11,498	20%	-5.8%	-1.15%
Sales versus Last Year ('000)	12,516	11,498	15%	-8.1%	-1.22%
Customer Satisfaction*	84.00%	80.91%	15%	-3.7%	-0.55%
Employee Satisfaction**	82.00%	79.26%	20%	-3.3%	-0.67%
Gross Profit Margin versus Budget	27.10%	25.49%	10%	-5.9%	-0.59%
Gross Profit Margin versus Last Year	27.21%	25.49%	5%	-6.3%	-0.32%
Percentage of Stock-outs***	2.50%	2.59%	10%	-3.5%	-0.35%
Average Tenure of Sales Personnel****	2.30	2.22	5%	-3.5%	-0.17%

**Overall Variance (above/below target):** - 5.02%

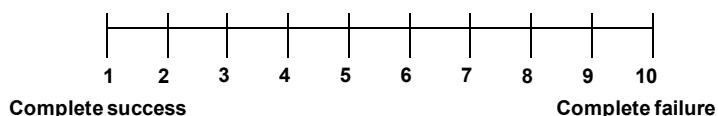
According to company policies the overall above/below target variance ( $\sum_i [weight_i \times target variance_i]$ ) was calculated before you make your final performance evaluation.

**Indicate your initial performance evaluation for John Miller below by placing an X on the scale at the point that corresponds to your evaluation.**

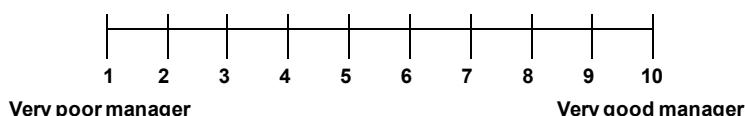
1. Evaluate John Miller's decision to make the changes in the store.



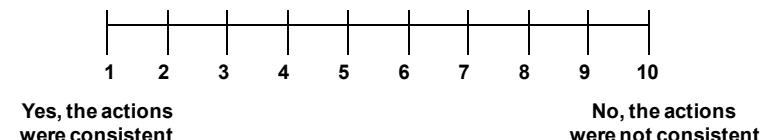
2. To what extent is this a case of success or failure for John Miller?



3. To what extent is John Miller a good manager?



4. Do you think that John Miller's actions were consistent with what is expected from a good manager?



\* - based on a customer survey from 0% (worst) to 100% (best)

\*\* - based on an employee survey from 0% (worst) to 100% (best)

\*\*\* - stock-out = running out of inventory/ failure to restock shelves

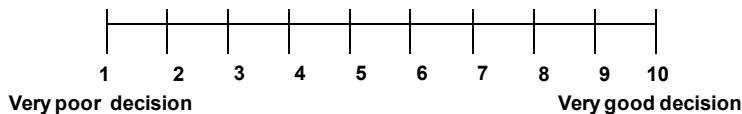
\*\*\*\* - average years that sales personnel stays in the firm

*Please continue to the next page*

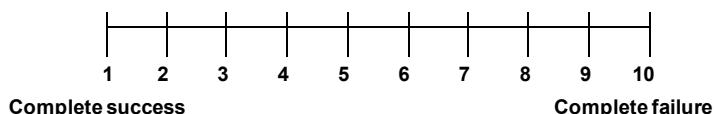
## ALL CONDITIONS

**Performance evaluation scales (following each version of the case):**

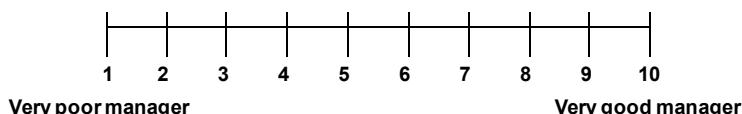
- 1. Evaluate John Miller's decision to make the changes in the store.**



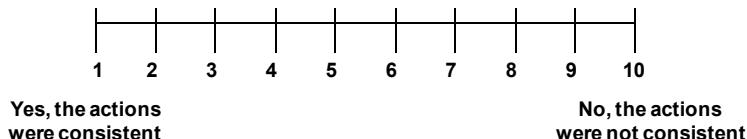
- 2. To what extent is this a case of success or failure for John Miller?**



- 3. To what extent is John Miller a good manager?**



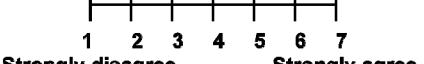
- 4. Do you think that John Miller's actions were consistent with what is expected from a good manager?**

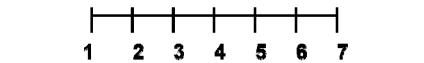
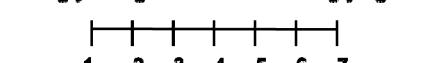
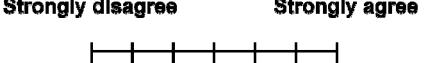
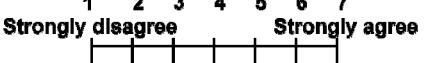
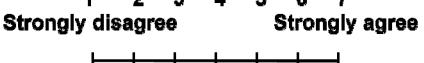


**ALL CONDITIONS**

**Please do not change your assessment on the previous page when answering the following questions.**

**Please answer the following questions about the performance measures that are used in the performance evaluation system in your company.**

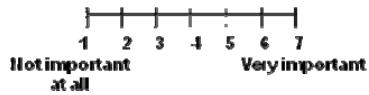
- 1) In your opinion, to what extent does John Miller have control on the outcome of the following measures for his store?
- a) Sales versus Budget (in '000)  

  - b) Sales versus Last Year (in '000)  

  - c) Customer Satisfaction  

  - d) Employee Satisfaction  

  - e) Gross Profit Margin versus Budget  

  - f) Gross Profit Margin versus Last Year  

  - g) Percentage of Stock-outs  

  - h) Average Tenure of Sales Personnel  

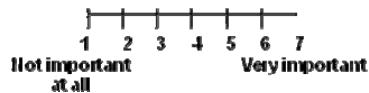

## ALL CONDITIONS

2) When you made your performance evaluation judgment, how important were the following pieces of information for your assessment:

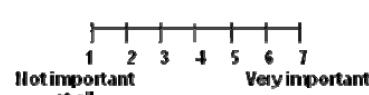
a) Company information



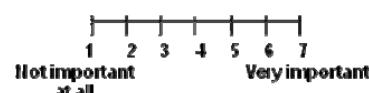
b) Store manager information



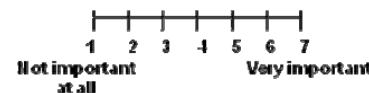
c) Additional information



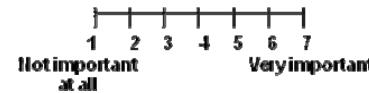
d) Sales versus budget



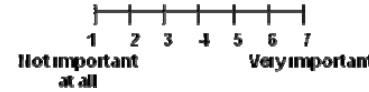
e) Sales versus Last Year



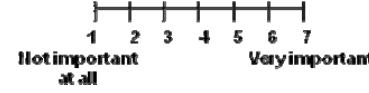
f) Customer Satisfaction



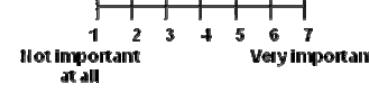
g) Employee Satisfaction



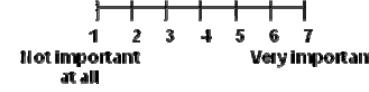
h) Gross Profit Margin versus Budget



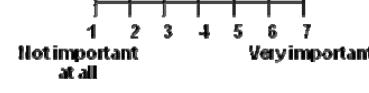
i) Gross Profit Margin versus Last Year



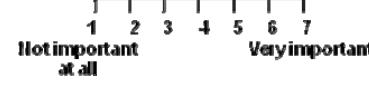
j) Percentage of Stock-outs



k) Average Tenure of Sales Personnel



l) Overall variance measure\*



\*Overall variance measure only in the two Formula-based conditions.

## ALL CONDITIONS

### **Demographic Survey**

Please answer the following questions that address demographic information that we will use for research purposes only. This information will be kept strictly confidential and will be reported only in aggregate form.

1. Gender: (Please check one.)       MALE       FEMALE
2. Age: (Please check one.)  
 0-22 years     23-30 years     31-40 years  
 41-50 years     51-60 years     61+ years
3. What industry do you work in?  
\_\_\_\_\_
4. What is your job title within the company?  
\_\_\_\_\_
5. How many years of work experience do you have?  
\_\_\_\_\_ years
6. How many years have you worked in your current job?  
\_\_\_\_\_ years
7. If your job is a management position, how many years have you worked in management?  
\_\_\_\_\_ I do not have a management position      \_\_\_\_\_ years
8. For how many years have you completed top-down performance evaluations as an evaluator?  
\_\_\_\_\_ Never      \_\_\_\_\_ years
9. For how many years has your performance been evaluated in the form of regular (e.g., annual, semi-annual) performance evaluations?  
\_\_\_\_\_ Never      \_\_\_\_\_ years
10. Have you ever performed or received an evaluation that was based on a balanced scorecard approach (i.e., based on a mix of financial and non-financial measures)?  
\_\_\_\_\_ yes      \_\_\_\_\_ no
11. What is your highest level of education completed: (Please check one.)  
 4-year undergraduate college degree (bachelor's degree)  
 master's degree  
 doctoral degree

THANK YOU FOR YOUR PARTICIPATION IN THIS PROJECT