Relative Performance Evaluation in Management Control

Explanations and Evidence for RPE Use and Effectiveness

Hilco J. van Elten

Relative Performance Evaluation in Management Control

Explanations and Evidence for RPE Use and Effectiveness

ISBN 978-90-8980-038-1

Typesetting in LaTeX

Printed by HAVEKA, Alblasserdam

NYENRODE BUSINESS UNIVERSITEIT

Relative Performance Evaluation in Management Control

Explanations and Evidence for RPE Use and Effectiveness

Thesis

with regard to the doctorate/PhD degree
at Nyenrode Business Universiteit
on authority of the
Rector Magnificus, prof. dr. M. van Rooijen
in accordance with the Doctorate Committee.

The public defence takes place on
Wednesday 30 May 2012
in the afternoon at exactly four o'clock

by

Hilco Jeroen van Elten

born on 9 August 1979
in Harderwijk

Examination Committee

Supervisor:	Prof. dr. R.F. Speklé
Other members:	Prof. dr. H.C. Dekker (*VU University Amsterdam*)
	Prof. dr. H.P.A.J. Langendijk
	Prof. dr. A.C.W. Sneller RC
	Prof. dr. ir. P.M.G. van Veen-Dirks (*University of Groningen*)

Table of Contents

Preface

Why do people make things hard on themselves? Is it because we find it satisfying to improve ourselves in certain areas? Is it because we derive some kind of self-worth from overcoming obstacles? Or do we just enjoy the challenge in itself? These are questions I have often found myself asking.

In any case, writing my PhD thesis has proven to be just such a challenge. It was a challenge that I certainly enjoyed, but also a challenge that brought frustration at times. Looking back on the past few years, I am first and foremost grateful to have finished this piece of work. Its completion creates the opportunity to focus on new challenges. But I believe I will also miss working on my thesis now that it is done.

Although writing a dissertation is a solitary process in some respects, it is not something that one does alone. Many people contributed to this thesis. Firstly, I acknowledge the support and advice of my thesis supervisor, Roland Speklé. I could not have asked for a more dedicated supervisor. His support and commitment, both to the content of the thesis and to my personal development as a researcher, have been unstinting. I would also like to thank my colleagues at Nyenrode's Centre for Management Accounting & Control. Together, we have created the constructive and open atmosphere that I find a necessary condition for research. Anne-Marie, Sally, Hans, Ivo and Rens, you have made it difficult for me to distinguish between colleagues and friends. I consider this invaluable.

Not only the people in the workplace deserve acknowledgement: my family and friends have also been very supportive. They listened to my stories, supported me whenever there were setbacks, and - most importantly - celebrated every victory as well. I especially want to thank my wife Maria. While I was writing my dissertation, she truly supported me at every step of the way. I started working on my dissertation when we had just met. Now that I am done, I know that I can rely on her in every situation in life.

Even after having completed this challenging project, one question still remains unanswered. This is the question that started this preface: why do people make things hard on themselves? I guess my answer would be: "I still have no idea..."

Hilco van Elten
March, 2012.

Chapter 1

General Introduction

1.1 The Problem of Goal Alignment

"Organizations are social units deliberately constructed to seek specific goals."
(Etzioni 1961)

Many scholars have sought to answer the question of what an organization is. Although a definite and undisputed definition of organizations may never be found, we can at least acknowledge certain characteristics. As posited by Etzioni (1961), one defining characteristic of organizations is that they are social entities seeking the realization of specific goals. These goals are often captured by an organization's owners and/or top management in the organization's vision, mission and strategy. For an organization to successfully pursue its goals, it is important that all individual members of the organization collaborate on achieving these goals. However, realizing collaboration is not an easy task. One reason for this difficulty is that organizations consist of many individuals who have their own goals as well. This is challenging because the organizational goals often differ from the goals of its individual members. In other words, collective and individual goals are not aligned automatically. More importantly, the goals of the individual members of the organization often conflict directly with organizational goals. For example, it is likely in the best interest of the organization that its members maximize their efforts to contribute to the organization's long-term value. However, an individual employee can have different goals. He may prefer to reduce his efforts by procrastinating. Additionally, it is possible that the self-interest seeking employee does not reduce his efforts per se, but repeatedly strays from organizational goals by, for example, adopting pet projects. Both cases have similar effects; long-term firm value is not optimized. Such behaviours are undesirable to the organization's owner, because they directly hamper the realization of organizational goals.

If any organization wants to successfully pursue its goals, the goals of the owner and the individual members of the organization need to be aligned. If goals are aligned, the members of the organization have the same interests as the owner. This is a necessary condition for successfully pursuing the achievement of organizational goals.

The academic field of Management Control addresses this problem of goal alignment. In general, Management Control studies how to motivate the members of the organization to act congruently with organizational goals. Management Control often relies on economics-based theorizing to address goal congruency issues. Economic reasoning argues that it is possible to rely on the provision of incentives to align the goals of the organization and its individual members. Such an incentive would be to reward individuals for goal-congruent behaviour. A typical reward would be a bonus for good performance. However, it is also possible that the reward is non-financial in nature. Examples of non-financial rewards include increased career prospects, or an acknowledgement from higher management, for example, as 'teacher of the year' or 'salesman of the month'. Both financial and non-financial rewards can be considered as a form of pay for performance.

The general idea is that pay for performance aligns the goals of individual employees with the owner's goals. After all, providing effort congruent with organizational goals is now in the employee's self-interest. The employee receives a reward when he performs well or, in other words, when he attains his performance target.

Aligning the goals of the organization with those of its individual members by providing incentives is a delicate task that requires careful consideration of all of the elements of such an 'incentive system'. Several design aspects must be determined before an incentive system can function properly, including: 1) the choice of the performance metrics, 2) the strength of the pay-performance relationship, and 3) the determination of the performance target level. This dissertation addresses this last aspect of incentive structure design, as will be explained in more detail in the remainder of this chapter. First, the three design aspects are discussed briefly.

The first design aspect concerns the choice of performance metrics. Performance measurement is an essential part of the design of any incentive system (Baker 2002). The basic question is how to measure performance in a reasonably undistorted manner. A performance measure is congruent or undistorted to the extent that it is systematically aligned with the organization's true objective (Bouwens & Speklé 2007). A wide variety of financial and non-financial performance measures exists. Examples include: profits, revenues, market share and customer satisfaction. However, it is also possible to select measures at the personal level, such as the development of professional skills and participation in permanent education programmes. To create a congruent performance measurement system, one must select those metrics that accurately measure what the owner desires of the

individual employee. Despite the variety of performance metrics, it is often very difficult to determine a performance measurement system that is highly congruent with the owner's goals.

The second incentive system design difficulty is determining the strength of the pay-performance relationship. This aspect relates to the question of how large of a reward an employee should receive if his performance metric increases. The economics literature shows that providing 'high-powered incentives' (or large rewards) increases the risk of incongruent behaviour if the performance metrics are not perfectly undistorted (which is almost always the case). For example, the possibility to obtain a large reward can provoke fraud with the performance metrics or other misrepresentations of performance. To avoid the risk of rewarding the wrong behaviour, the economics literature generally concludes that organizations should delimit the strength of the pay-for-performance relationship (Baker 2002).

The third and final design element is the determination of the performance target level. To achieve goal alignment by using an incentive system, it is essential for organizations to make sensible choices regarding the difficulty of the performance target. In business practice, this seems to be a challenging task. Conventional wisdom suggests that, for optimum motivation, targets should be set sufficiently (but not excessively) high, so that they are more often missed than not (Merchant & Manzoni 1989:554). However, Merchant and Manzoni find empirically that, on average, organizations set their targets too low, at 'relatively easy' levels where they are very likely to be achieved. Setting performance targets at suboptimal levels harms the realization of organizational objectives. Relatively easy targets do not optimally motivate employees, because they do not require employees to maximize their efforts. Even with lower effort the targets can be achieved. On the other hand, overly challenging targets also do not optimally motivate the employee. Then, it becomes highly unlikely that the employee will achieve his target and receive his reward, which reduces his incentive to provide effort. This too unnecessarily limits the realization of organizational goals.

Whereas the choice of performance metrics and the pay-performance sensitivity aspects of the incentive structure design are widely studied in both the economics and accounting literatures, the target-setting element is currently understudied (Ittner & Larcker 2001, Murphy 2001, Dekker *et al.* 2012). Consequently, we know very little about how organizations determine the difficulty of their performance targets. This is unfortunate because the performance standard choice is a consequential dimension of the incentive structure. This doctoral dissertation fills this gap by focussing on the determination of the performance standard level.

According to Murphy (2001), two main ways for performance target determination can be distinguished. First, targets can be internally determined. The term 'internally determined' refers here to standards that are directly affected by employee actions in the current

or prior year (Murphy 2001:245). For example, next year's revenue target can be based on last year's sales. Another example is internal negotiations between higher and middle management about the target's difficulty. Second, targets can be externally determined, referring to standards that are less easily affected by actions from employees who are involved in realizing the target. Externally determined targets include: timeless performance standards, companywide cost of capital, and the use of a peer group benchmark.

My research project focuses on one method of 'external' target determination, namely Relative Performance Evaluation (or RPE). RPE is a means to determine the performance standard by using a peer group benchmark: the performance of the employee (or: manager) is compared to the performance of a reference group. This reference group typically consists of individuals facing similar tasks and circumstances. For example, a manager's performance can be compared to the five best performing managers within the organization or to a group of important competitors operating on the same market. The idea behind this comparison is that the performance of the reference group is informative for the performance potential of the employee. The performance potential of the employee functions as an (explicit or implicit) performance target against which his actual performance is evaluated.

This thesis' overall research topic is Relative Performance Evaluation. I study the use and effectiveness of RPE at the organizational level of middle managers. Based on prior literature, I argue that RPE has a sound theoretical basis, but lacks overall conclusive empirical support. My research topic is further motivated in the remainder of this introductory chapter. First, I review the literature on RPE in section 1.2. Then, in section 1.3, I introduce the research questions of this project and discuss how these research questions contribute to the literature. Finally, section 1.4 describes the structure of the entire thesis.

1.2 Literature Review

This section addresses the current theoretical and empirical literature on RPE. In section 1.2.1, I discuss two theoretical perspectives on RPE that explain why organizations might use it. Subsequently, in subsection 1.2.2, I present what the literature suggests about RPE's empirical validity and to what extent we find empirical support for the presented theoretical explanations.

1.2.1 Theoretical Explanations for RPE Use

The extant RPE literature argues that the adoption of RPE practices can be a useful endeavour for several reasons. Prior studies argue for the use and effectiveness of RPE

along two lines. First, RPE can provide the owner/top management with more complete information about the performance of the manager. This makes the performance evaluation less 'noisy'. Second, RPE can reduce the room that the employee has to behave opportunistically. In this section, the mechanics of both the noise-reduction and opportunism-mitigating properties of RPE are introduced.

Relative Performance Evaluation has been recognized in the literature primarily for its noise-reduction properties (Holmstrom 1982). Noise in the performance evaluation is caused by factors beyond the employee's control that affect his measured performance. With noisy performance measures, factors beyond the control of the employee influence whether he reaches his performance target. This external influence reduces the quality of the performance evaluation, which aims to reward the employee for his provision of effort, not for uncontrollable events. Reasoning from an analytical agency theoretical tradition, Holmstrom (1982) proves that RPE reduces noise in the evaluation by incorporating information about peer performance into the performance contract between the owner/top management (in agency terminology, the principal) and the employee/middle manager (designated as the agent in this relationship). The performance of the reference group is informative about the quality of the agent's performance, due to shared exposure to the uncontrollable events. Comparing the performance of the agent to the reference group partially filters out the effects of external uncontrollable factors. As a result of this noise reduction, RPE improves the quality of performance evaluation.

Example Consider the hypothetical situation of a sales manager at the Japanese guitar manufacturing company 'Ibanez Guitars' in the early 1990s. Due to the increasing popularity of electronic dance music, playing the guitar became less popular. As a result, guitar sales dropped, and Ibanez's previously healthy company profits slowly vanished. The external event (the changing musical preference) affected the performance evaluation of Ibanez's sales manager, who was rewarded based on his sales revenue. The changing music scene did not only affect Ibanez's performance. Other guitar companies suffered from dropping revenues and profits too. The performance evaluation of the manager, which compared his actual revenue to his sales budget, was noisy. This noise was due to the impact of external events over which the manager had no control. This situation was not motivating for the sales manager because his performance evaluation indicated decreasing performance every year, no matter how much effort he put in. However, if the superior of Ibanez's sales manager had taken peer performance into account when evaluating the manager, the performance evaluation would have given a more accurate impression of the sales manager's efforts. By judging the manager's relative performance (*i.e.*, his performance compared to his peers), the evaluation would be insulated from the effects of society's eroding musical taste.

The second explanation for RPE is the opportunism-mitigation perspective. The opportunism-mitigation perspective relates to the way in which performance standards are determined. Murphy (2001) distinguishes between two main ways in which standards can be set: 1) they can be determined in an internal, administrative process; or 2) they can be externally determined. The difference between these alternatives lies in the extent to which employees can influence the target level (Murphy 2001). Internally determined standards include standards based on prior-year performance and standards derived from company plans or budgets. Such standards are affected by managers' actions, and can have dysfunctional effects. For example, when standards are based on prior-year performance, managers have an incentive to avoid unusually positive outcomes, because good performance in the current period is penalized through an increased standard in the next period (Murphy 2001). In a similar vein, budget-based standards provide incentives to negotiate easy standards (Fisher, Frederickson & Peffer 2002) and disincentives to beat the budget, especially in a regime of incremental budgeting (Murphy 2001). In contrast, externally determined standards are less affected by managers' actions because the difficulty of the standard is based on something outside the sphere of influence of the manager (*e.g.*, it is based on market conditions or peer performance).

RPE is an example of such an externally determined performance standard, because it incorporates information about the performance of an external reference group of agents. The performance of an industry peer group lies well outside the sphere of influence of the manager, yet it determines his performance target. RPE-based targets are not subject to the prior year's performance, to negotiations between the evaluated manager and the top management of the firm, or to any other factor that the evaluated manager can easily manipulate. These considerations suggest that the room for managerial opportunistic behaviour can be delimited by incorporating the performance of a reference group of agents into the compensation plan and that RPE can aid in the standard-setting process. Throughout the thesis, I motivate and test the use of RPE with the two perspectives presented above.

1.2.2 Empirical Results on RPE

Despite a promising theoretical basis, the empirical literature, focusing mainly on executive compensation praxis, provides inconclusive evidence on the empirical relevance of RPE theory. Both Jensen & Murphy (1990) and Aggarwal & Samwick (1999) find that RPE is not a prominent feature of executive compensation contracts. Antle & Smith (1986) find weak support for RPE use in the executive compensation contracts of 16 firms in a sample of 39 firms with longitudinal data from 1947 to 1977. Also Liu & Stark (2008) find mixed support for RPE amongst 169 UK non-financial listed companies. They provide evidence for the presence of RPE in the design of board compensation based on stock market returns, but they find no relation between cash compensation of the board and peer group

market returns. This leaves the cash compensation of boards of directors partially exposed to industry uncertainties.

Gibbons & Murphy (1990) report strong support for RPE use, but the interpretation of their results has been criticized quite severely by Janakiraman *et al.* (1992). Janakiraman *et al.* (1992) argue that Gibbons & Murphy's (1990) analysis of 7.757 cash compensation observations of 1.049 firms over a 12-year period only supports a 'weak' form of RPE. This weak form of RPE refers to situations where peer performance is not entirely, but only partially filtered out of the compensation contract. Janakiraman *et al.*'s (1992) own empirical results are generally negative regarding RPE's presence.

Other, more recent studies by Bertrand & Mullainathan (2001), Garvey & Milbourn (2006), and Himmelberg & Hubbard (2000) report no support for RPE. However, Albuquerque (2009) argues that this lack of support is due to misspecification of the analyses. According to Albuquerque (2009), most RPE studies suffer from a poor specification of the peer group, which blurs the results. After respecification of the analyses, Albuquerque (2009) finds partial support for RPE use amongst executives.

Overall, the empirical evidence remains mixed at best, suggesting that RPE has limited empirical relevance, at least among executives.

1.3 Research Questions and Contributions

From the literature review presented in the previous section, we see that RPE has a sound theoretical basis, but lacks overall conclusive empirical support. The objective of this research project is to fill this gap partially by: 1) seeking further empirical documentation about RPE; 2) explaining why organizations do or do not rely on RPE practices for the evaluation of business unit managers; and 3) studying whether RPE use increases the effectiveness of the Management Control System of its adopters. This results in the following research questions and accompanying contributions:

1. *To what extent is RPE used for the performance evaluation of business unit managers?*

   The first research question directly relates to the empirical puzzle presented in the previous section of this chapter: RPE seems theoretically sensible for business practice, but it lacks empirical support. However, the prior literature has several characteristics that may reduce the chance of finding support for RPE's empirical existence.

First, the literature focuses almost exclusively on the executive level. However, RPE might be not very efficient if deployed at the highest organizational level, as Garvey & Milbourn (2003) argue. It is entirely possible that relatively wealthy executives can reduce noise in their incentive contracts more efficiently themselves by adjusting their own private investment portfolios. The option of managerial hedging may decrease the attractiveness of RPE, reducing its adoption amongst top managers. However, private hedging is costly. As a result, it is probably not an efficient option at the level of (less wealthy) business unit managers. In support of this argument, Garvey & Milbourn found that, although RPE is not an important feature in the compensation contract of the average manager, it is in fact present for younger and less wealthy executives (*i.e.*, those managers for whom personal hedging is likely to be prohibitively costly (Garvey & Milbourn 2003:1557)). If personal hedging is indeed too costly for less wealthy executives, it is probably also too costly for managers at lower echelons. This fact increases the attractiveness of RPE at these levels. My thesis contributes to the literature by studying RPE at the level of business unit managers, where it may be more prevalent than it is at the executive level.

The second characteristic of the extant RPE literature concerns the type of data that is used. Most currently available empirical studies rely on archival data from before 2006. Using archival data has several benefits (*e.g.*, data availability, objectivity of the data, and sample size). However, prior to the SEC's 2006 executive compensation disclosure rules, firms were not required to disclose detailed information about their compensation plans. As a result, pre-2006 archival data have the drawback of requiring indirect instead of direct analyses. Indirect analyses approach RPE *as if* performance was evaluated relative to competitors' performance, as opposed to explicitly studying whether the performance evaluation system incorporates peer performance information. This approach is inevitable when using archival data to investigate a phenomenon on which firms do not specifically report, such as RPE use. Although this is likely to change in the near future, studies that incorporate the newly available disclosure information are still scarce. Currently, the prevalent archival approach provides mixed support for RPE's empirical validity.

As opposed to the prior literature described here, this thesis uses data from a survey. This survey instrument is specifically designed for the current study. This should add more detail and richness to the analyses, resulting in a better description of RPE in practice.

The first research question is addressed in chapter 2 of this dissertation.

2. *Can RPE use amongst business unit managers be explained from a noise-reduction and/or opportunism-reduction perspective?*

   The second research question refers to understanding RPE use. In chapter 2, I present a model that aims to explain RPE use at the business unit level. To do so, the model consists of antecedents derived from the noise- and opportunism-mitigation perspectives, as presented in section 1.2.1 of the current chapter.

   Here lies the main theoretical contribution of my thesis. Whereas most studies adopt a noise-reduction perspective on RPE, I study RPE use with more than one explanation. I rely simultaneously on both the noise perspective and on Murphy's opportunism mitigation standpoint on RPE. To my knowledge, the latter perspective has not gained prior attention in the RPE literature. Also, using two explanations simultaneously provides a clearer picture of the effects of the individual explanations, than using one explanation in isolation. Together, these two perspectives provide us with a richer framework to understand organizations' reliance on RPE.

3. *Is RPE effective at reducing noise and opportunism?*

   Third, this study addresses the effectiveness of RPE use in reducing noise and opportunism. To an economist, studying whether RPE is effective may seem obsolete after studying why and to what extent organizations use RPE. The notion of economic Darwinism posits that organizations generally make optimal choices in order to survive, and that organizations that make suboptimal choices are eliminated via natural selection in the marketplace. This argument implies that organizational behaviour (such as the adoption of RPE) that is observed in practice must be -at least in general- effective[1]. This implication means that spending time on additional effectiveness studies is not a very efficient use of valuable research time.

   However, I argue that additional tests are both useful and necessary as additional proof for explanations about how RPE works in practice. Especially in the case of the noise-reduction argument, prior literature suggests alternative explanations that could explain the findings in chapter 2, potentially just as well as noise-reduction considerations. These alternative explanations (such as promoting internal competition within the organization, or stimulating organizational learning through benchmarking) may cause RPE to be effective, and therefore survive in the marketplace. The alternative explanations are addressed in more detail in the introduction section of

[1] A further discussion of economic Darwinism and whether the effectiveness claim can or even should be relaxed, lies beyond the scope of this thesis.

chapter 3 (see page 61). As additional support of the noise-reduction explanation for RPE use, chapter 3 presents more direct analyses of RPE's effectiveness in reducing noise in the performance evaluation.

Although alternative explanations may also exist for opportunism-mitigation, the primary motivation for the analysis of RPE's opportunism-mitigating effectiveness conducted in chapter 4 is of a statistical rather than a theoretical nature. In chapter 2, I find mixed support for RPE's effectiveness in mitigating managerial opportunism. Chapter 4 provides additional analyses as a further examination of the relation between RPE and managerial opportunism. The mixed findings in chapter 2 may result from the limited statistical power of its analyses. If limited statistical power is the cause of the weak support for the opportunism explanation, relying exclusively on the findings of chapter 2 leads to the false rejection of the research hypothesis. However, chapter 4 conducts statistically more powerful tests of the opportunism-explanation. Whereas chapter 2 studies RPE use from both noise- and opportunism-reduction perspectives, chapter 4 focuses solely on opportunism-mitigation. This reduces the theoretical 'fullness' sought in chapter 2, but allows for analysing direct effects instead of statistically more demanding interaction effects.

Therefore, I argue that it is not only sensible but also necessary to conduct additional effectiveness studies to augment the study in chapter 2. These effectiveness studies examine whether RPE use indeed yields lower noise and opportunism. These studies are documented in chapters 3 and 4.

By studying the performance target determination of middle-managers, this project contributes more broadly to the target setting literature in general (*e.g.*, Merchant & Manzoni 1989; Murphy 2001; Indjejikian & Nanda 2002; Leone & Rock 2002). Research on the methods that firms use to determine their performance targets is currently scarce. Following Ittner & Larcker (2001) and Murphy (2001), Dekker *et al.* (2012) argue that target level determination has received relatively little research attention, although it is a key issue of the use of performance measures and targets. My thesis contributes to our understanding of how organizations determine the difficulty of their performance target levels.

1.4 Structure of the thesis and relations between the individual studies

This dissertation presents three more or less independent studies on RPE in three consecutive chapters (chapters 2, 3, and 4). After the general introductory chapter, chapter 2

presents the first study of this thesis. This study addresses the use of RPE for the performance evaluation and the target-setting process of business unit managers. It describes the prevalence of RPE at the level of business unit managers and antecedents to RPE use. This chapter studies to what extent RPE use is associated with high levels of common uncertainty (as a noise-inducing factor) and information asymmetry (as a driver for managerial opportunism). Summarizing the findings, the first study shows that RPE use is indeed associated with drivers of noise and opportunism, albeit with mixed results for the latter.

Chapter 2 tests theoretical relations that build on considerations of noise-reduction and opportunism-mitigation to motivate the use of RPE. Adding to the analyses of chapter 2, the third and fourth chapter empirically study these underlying considerations in a more direct manner. In chapter 3, the second study investigates whether RPE is effective in reducing noise that stems from the external environment. The data show that when RPE is used specifically to determine the performance target and more generally to mark the performance evaluation of the agent to market, noise in the performance evaluation is reduced. The third study of this thesis, documented in chapter 4, focuses on RPE's effectiveness in mitigating the room for managerial opportunism. Contrary to my expectations, the study does not support the opportunism-mitigation characteristic of RPE.

Although the three studies in this thesis share a common theme and methodology, each research chapter is presented as an independent study. As a result, parts of the chapters overlap, for example, in parts of the theory, but most noticeably in the sample descriptions and method-sections. I deem this overlap necessary to preserve the character of each chapter as an independent research report addressing its own specific research topic.

Finally, chapter 5 summarizes and concludes this dissertation and discusses its limitations.

Chapter 2

The Use of Relative Performance Evaluation at the Business Unit Manager Level

Preamble to Chapter 2

Chapter 2 presents the first study of this thesis. This study addresses research questions 1 and 2. The first research question is: *To what extent is RPE used for the performance evaluation of business unit managers?* This question relates to the empirical puzzle presented in the introduction: although theory reasons that RPE is sensible for business practice at that organizational level, it lacks conclusive empirical support. By researching the prevalence of RPE amongst business unit managers, this chapter studies whether RPE is a concept with practical relevance.

The second research question is formulated as: *Can RPE use amongst business unit managers be explained from a noise-reduction and/or opportunism-reduction perspective?* Given the fact that RPE is used in practice, the second research question aims to explain why some organizations use it and others do not. This chapter explains RPE use in terms of both its noise- and opportunism-mitigating properties. These two explanations are both derived from prior literature, although the latter explanation has received far less theoretical and empirical attention in existing research compared to the former explanation. The current chapter contributes to the literature by extending the empirical evidence on both explanations. Additionally, using two explanations simultaneously provides a clearer picture of the effects of the individual explanations, compared to using one explanation in isolation. Together, these two perspectives provide us with a richer framework in which to understand organizational reliance on RPE.

2.1 Introduction

This chapter studies the use of RPE amongst business unit managers. RPE is a means to determine performance targets relative to a reference group. Performance targets are an important and consequential part of the incentive structure (Murphy 2001) and function as a norm to which the performance of an employee is compared. According to economics-based reasoning, setting targets at a difficult, yet attainable level helps to motivate employees in such a way that they provide optimal and goal-congruent efforts.

However, determining the appropriate level for the performance targets can be difficult. Factors both external and internal to the organization influence the target and its attainability. External factors refer to events that originate from the environment of the organization, such as rising prices or declining demand. If external events affect organizational performance, they also affect whether an employee reaches his target. This external influence makes the performance measures less informative about the employee's efforts and increases noise in the performance measurement. Noise mitigates the motivational effects of an otherwise well functioning incentive system. Ultimately, noise harms the realization of organizational goals. However, analytical research shows that RPE can reduce noise and restore the relevance of the performance target (Holmstrom 1982). RPE compares the employee's actual performance to the performance of a reference group facing partially the same external events. Peer performance helps to determine an explicit or implicit performance target that is informative about the impact of external events. This reduces noise in the performance evaluation, increasing the relevance of the performance target and the quality of the performance evaluation (Holmstrom 1982).

Besides external events, factors that originate from within the organization can also affect the performance target's difficulty because of the way in which performance targets are determined. Murphy (2001) argues that members of the organization can try to lower the performance target if it is within their reach. Examples include negotiating easier targets, or avoiding outperforming the target for this year to prevent more difficult targets next year. Murphy argues that if the target-setting process does not lie within reach of the evaluated employee, the target is better insulated against these kinds of opportunistic behaviour. This is the case with RPE, where the performance target is based on the performance of a peer group, external to the influence of the evaluated employee. Based on this reasoning, Murphy claims that RPE can mitigate the room for opportunistic behaviour.

According to both of these arguments, RPE can be beneficial to the organization by improving the relevance of the performance target. Despite this theoretical basis, the empirical literature provides inconclusive evidence on the empirical relevance of RPE theory. For example, both Jensen & Murphy (1990) and Aggarwal & Samwick (1999) find that RPE

is not a prominent feature of executive compensation contracts. Antle & Smith (1986) and Liu & Stark (2008) find some support for RPE, but they also present results that are inconsistent with RPE. Gibbons & Murphy (1990) report strong support, but the interpretation of their results has been criticized quite severely by Janakiraman *et al.* (1992). Even less in favour of RPE, Janakiraman *et al.*'s own empirical results (1992) are generally negative about RPE's presence. Additionally, Bertrand & Mullainathan (2001), Garvey & Milbourn (2006), and Himmelberg & Hubbard (2000) report no support for RPE. However, Albuquerque (2009) argues that this is due to misspecification of the analyses. After respecification of the analyses, she finds partial support for RPE use amongst executives. Overall, the empirical evidence remains mixed at best, suggesting that RPE has limited empirical relevance[1].

However, the empirical evidence may be limited due to several characteristics that prior RPE studies have in common. A first reason for the prior literature's limited support for RPE may lie in its focus on the executive level. If deployed at the highest organizational level, RPE might be less efficient, as Garvey & Milbourn (2003) argue. It is entirely possible that relatively wealthy executives can reduce noise in their incentive contracts more efficiently themselves by adjusting their own private investment portfolios. The option of managerial hedging may decrease the attractiveness of RPE, reducing its adoption amongst top-managers. However, private hedging is costly. As a result, it is probably not an efficient option at the level of (less wealthy) business unit managers. In support of this argument, Garvey & Milbourn found that although RPE is not an important feature in the compensation contract of the average manager, it is in fact present for younger and less wealthy executives, *i.e.*, those managers for whom personal hedging is likely to be prohibitively costly (Garvey & Milbourn 2003:1557). If personal hedging is indeed too costly for less wealthy executives, it is probably also too costly for managers in lower echelons. This fact may render RPE more attractive at these levels.

Another important characteristic of the extant RPE literature that may have caused the literature's negative findings, concerns the type of data being used. Most empirical studies that are currently available in the literature rely on archival data from before 2006. Using archival data has several benefits (*e.g.*, data availability, objectivity of the data, and sample size). However, prior to the SEC's 2006 executive compensation disclosure rules, firms were not required to disclose detailed information about their compensation plans. As a result, pre-2006 archival data have the drawback of requiring indirect instead of direct analyses. Indirect analyses typically regress executive pay on firm performance and peer performance (*e.g.*, Aggarwall & Samwick 1999, Albuquerque 2009, Janakiraman *et al.* 1992, Liu & Stark 2008). In these analyses, a positive coefficient for firm performance together with a negative coefficient for peer performance provides support for the

[1]The empirical RPE literature is discussed in more detail in the introductory chapter of this thesis in section 1.2.2 on page 6.

reliance on RPE to determine the height of executive pay. However, these analyses make numerous assumptions about, for example, the composition of the peer group, and the performance targets for the CEO. In other words, the indirect analyses on which these papers rely, study RPE *as if* performance is evaluated relative to competitors' performance, as opposed to explicitly studying whether the performance evaluation system incorporates peer performance information. This approach is inevitable when using archival data for a phenomenon on which firms do not specifically report, such as RPE use. Although this is likely to change in the near future, studies that incorporate the newly available disclosure information are still scarce. At present, the prevalent archival approach provides mixed support for RPE's empirical validity.

This chapter studies RPE at a lower organizational level, using more detailed data. Additionally, it uses a broader theoretical lens and studies explanations for RPE use empirically from both a noise- and an opportunism-reduction perspective. The opportunism-reduction explanation for RPE use is currently understudied in the literature, which focuses almost exclusively on noise-reduction. Also, analysing two explanations simultaneously provides a clearer picture of the individual and joint effects of the individual explanations compared to studying one explanation in isolation. It allows for analysing whether one explanation has an effect over and beyond the other explanation. Additionally, the simultaneous analysis of both explanations for RPE may yield greater explanatory power for RPE use in practice. In summary, analysing these two perspectives together in one study provides us with a richer framework to understand organizational reliance on RPE.

Unlike the extant literature, this research studies RPE at the middle-management level. More specifically, this research studies RPE's empirical relevance by describing its prevalence amongst business unit managers in the Netherlands. As argued above, RPE may be more prevalent at the business unit level than it is amongst executives. Also, unlike prior literature, this research uses data from a purpose-developed survey instrument. Instead of relying on archives, this study assesses RPE by using detailed survey data from 325 business unit managers in the Netherlands. This approach should provide the analyses with more detail and richness, resulting in a better description of RPE in practice.

Overall, the research findings suggest that RPE is indeed a relevant aspect of performance evaluation praxis. Approximately 88% of the respondents in my sample use (at least to some extent) peer performance information for determining the performance standards. More than 50% of the respondents claim to use RPE to a great or very great extent. Additionally, I find that both noise- and opportunism-reduction arguments help to explain RPE use in practice. Using ordinary least squares (OLS), Tobit and Logit models, I find overall robust support for the noise-reduction perspective and mixed results concerning the opportunism-mitigation hypothesis.

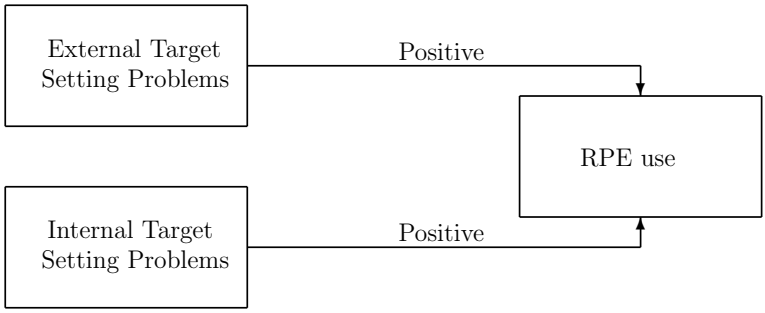
The remainder of this chapter is organized in four sections. Section 2.2 provides the theoretical background of RPE, the theoretical model and the development of the hypotheses. The measurement is described in section 2.3, and the analyses and results in section 2.4. Section 2.5 addresses the findings, and the conclusions.

2.2 Theoretical Background and Development of the Hypotheses

As described in the introduction of this chapter, RPE can reduce two types of target-setting problems: 1) problems that arise from outside the business unit (external problems), and 2) problems that arise within the business unit (internal problems). Based on prior literature, I argue that the presence of both types of problems can induce RPE use. The conceptual relations between target-setting problems and RPE are presented in figure 2.1.

This section describes the nature of the two types of target-setting problems in terms of their relation to noise in the performance evaluation and room for managerial opportunism, and operationalizes these factors as common uncertainty and information asymmetry, respectively. These antecedents of RPE (common uncertainty and information asymmetry) are assumed to increase the noise and opportunism-levels which drive organizational reliance on RPE[2]. This section develops three research hypotheses based on these antecedents and ends with the introduction of control variables.

Figure 2.1: Conceptual Model



[2]In addition to common uncertainty and information asymmetry, other factors exist that influence noise and opportunism, including 'uncommon' or idiosyncratic uncertainty. The current chapter focuses on common uncertainty because, unlike idiosyncratic uncertainty, it can be reduced by RPE use. Chapters 3 and 4 assess a more comprehensive list of noise and opportunism drivers.

2.2.1 RPE and Noise Reduction: Reducing the Effect of External Factors on the Performance Evaluation

This section describes how RPE can reduce target-setting problems that originate from outside of the business unit. These external factors include events that occur in the environment of the business unit. Examples include rising prices and declining demand. External factors affect the performance of the business unit and thus the difficulty of the performance target. Windfall gains can cause the target to be achieved easily. This reduces the motivation of the business unit manager to provide maximum effort; his performance target is achieved due to good luck, and requires less effort. On the other hand, bad luck can make the target impossible to achieve. This also reduces the motivation of the manager, who cannot achieve his target, even when providing his maximum effort.

External events increase the 'noisiness' of the performance metrics. Measured performance is the result not only of effort, but also of uncontrollable factors. This reduces the informativeness of the performance metrics about the agent's performance. Overall, noise reduces the quality of the performance evaluation, which aims to reward the agent for the provision of effort, not for uncontrollable events.

However, RPE can reduce the noise in the performance evaluation, insulating the performance evaluation from the effects of external events. In the literature, noise reduction is the primary explanation for RPE use (see Holmstrom 1982). RPE reduces noise in the evaluation by incorporating information about peer performance into the performance contract between the principal and the agent (Holmstrom 1982). Because the agent and his peers have partially similar exposure to uncontrollable external events, their performance is similarly affected by these events. Comparing the agent's performance to the performance of his peers provides the principal with more information about what performance level is achievable under the current circumstances. Thus, by judging the agent's performance relative to his peers, the principal is informed about the *quality* of the agent's performance under the (often uncontrollable) external circumstances. This reduces the effects of external uncontrollable factors on the performance evaluation. Due to this noise reduction and increased informativeness, RPE can improve the quality of performance evaluation.

2.2.1.1 Common Uncertainty

The noise-reduction perspective on RPE informs my model with the effect of uncertainty on the use of RPE. Uncertainty is a widely researched antecedent for many control instruments (Chenhall 2003). Uncertainty can introduce noise in the performance evaluation if uncertainty is uncontrollable for the evaluated agent, and influences his performance.

Since RPE has been described in the literature (*e.g.*, Holmstrom 1982), the applicability of RPE has been associated with uncertainty. More specifically, this argument applies to uncertainty shared amongst peers because peer performance is only informative to the extent that the uncertainties are common in nature (Holmstrom 1982, Frederickson 1992, Matsumura & Shin 2006). Only common events influence both the agent's performance and peer performance. On the other hand, individual (or idiosyncratic) uncertainty only affects the agent's performance, leaving peer performance unaffected. This reasoning follows Gibbons & Murphy (1990) and Kren (2002), who explain RPE use with its potential benefit of specifically filtering out common (as opposed to idiosyncratic) uncertainty[3].

I hypothesize specifically that RPE can provide insulation for the noise-increasing effects of common uncertainty. I expect more RPE use under higher levels of common uncertainty. This expectation results in the following hypothesis:

H1: Common uncertainty has a positive effect on the use of RPE.

2.2.2 RPE and Opportunism-Mitigation; the Use of Internally vs. Externally Determined Standards

As argued in the introduction of this chapter, it is not only external events that can cause target-setting problems. Problems that originate from within the business unit can also cause problems by affecting the difficulty of the performance target. Specifically, this argument relates to managerial opportunism and the room that managers have to opportunistically affect (*i.e.*, lower) their performance target level. Murphy's (2001) findings imply that managers lower their performance targets if the possibility to do so, presents itself to them. This reduces the motivational properties of the target, which may lead to suboptimalization of managerial efforts.

Murphy (2001) shows, both theoretically and empirically that RPE can reduce the room that managers have to behave opportunistically. This second perspective on RPE relates to *the way in which* performance standards are determined. Murphy distinguishes between two main ways in which standards can be set: 1) they can be determined in some internal, administrative process; or 2) they can be externally determined. The difference between these alternatives lies in the extent to which agents can influence the level (or: difficulty) of the performance standard. Internally determined standards include standards based on prior-year performance and standards derived from plans or budgets. Such standards are

[3]Both Gibbons & Murphy (1990) and Kren (2002) find partial support for this proposition.

affected by managerial actions, and can have dysfunctional effects. For example, when standards are based on prior-year performance, managers have an incentive to avoid unusually positive outcomes, because good performance in the current period is penalized through an increased standard in the next period (Murphy 2001). In a similar vein, budget-based standards provide incentives to negotiate easy standards (Fisher, Frederickson & Peffer 2002) and disincentives to beat the budget, especially in a regime of incremental budgeting (Bouwens & Kroos 2011, Murphy 2001). In contrast, externally determined standards are less affected by managerial actions because the difficulty of the standard is based on something outside the sphere of influence of the manager (*e.g.* it is based on market conditions or peer performance). Examples of externally determined standards include timeless performance standards (plans measuring performance relative to a fixed standard such as a pre-specified return on assets (Murphy 2001:252)), company-wide cost of capital, and the use of peer groups (Murphy 2001).

RPE is an example of such an externally determined performance standard because it incorporates information about the performance of an external reference group of agents[4]. The performance of an industry peer group lies well outside the sphere of influence of the manager, yet it does constitute his performance target. Such an RPE-based performance target is not subject to the business unit's prior-year performance, to negotiations between the business unit manager and the top management of the firm, or to anything else that the business unit manager can manipulate. These considerations suggest that the room for managerial opportunistic behaviour can be delimited by incorporating the performance of a reference group of agents into the performance evaluation and that RPE can aid the standard-setting process.

2.2.2.1 Information Asymmetry

The opportunism-mitigation perspective informs my model with the effect of information asymmetry on the use of RPE. Information asymmetry relates to the information gap between principal and agent, for example, concerning the knowledge about the quality of the business unit performance, the (internal and external) factors that influence the performance, the technical processes and the transformation processes of the unit. Information asymmetry is often argued to be the core of agency problems. It is the combination of information asymmetry and the agent's aversion both to work and risk that steers him away from cooperative behaviour and results in managerial opportunism (Holmstrom 1982, Kunz & Pfaff 2002). Information asymmetry leads to managerial opportunism if the agent,

[4]'External' means that the reference group is external to the business unit. However, it is not necessarily external to the firm. A reference group might consist of other business units inside the firm that face (some of) the same external conditions (*e.g.*, they operate on the same market).

who has superior insights about the mechanisms that drive performance, resorts to misrepresentation of the performance potential towards the principal or building budgetary slack (Fisher, Maines, Peffer & Sprinkle 2002, Dunk 1993, Chow *et al.* 1988). Information asymmetry makes target setting difficult. Setting performance target difficulty at a relevant level can be problematic when a subordinate has better information than a superior about factors that influence performance, especially when the agent's pay depends on his performance vis-à-vis the target (conform: Chow *et al.* 1988).

When information asymmetries concerning an activity exist, the organization can choose to solve them by investing in knowledge about the activity[5]. Organizations can also choose not to solve the information asymmetry, for example, because of the costs involved. The latter strategy might be fully acceptable if the top management is still able to make meaningful claims about the quality of the business unit performance, which is possible via RPE. Through RPE, the principal can let the peer group 'decide' the difficulty of the performance standard, without the need to understand the specific workings of the business unit and the factors that drive its performance. This does not reduce the amount of information asymmetry itself; the principal remains unaware of the specific factors that determine the business unit's performance potential. Instead, the principal retrieves the information that she requires from the market. This way, the principal works around the business unit to assess its manager's effort, instead of obtaining the information required to evaluate the performance of the business unit, which is often costly to obtain or simply unavailable.

This study argues that RPE can mitigate the managerial opportunism that results from information asymmetries, by determining the performance target outside of the sphere of influence of the agent. RPE can reduce the target-setting consequences of the information asymmetries. Even if the principal has limited knowledge of the factors that influence performance, she can observe the effects of these factors through changes in peer performance. *Ergo*, I expect information asymmetry to increase the use of RPE.

However, I only expect this relation under the condition of comparability to a relevant peer group. For RPE to reduce the effects of information asymmetry, the transformation process of the business unit and the factors that influence the performance potential, must be comparable to, for example, other business units. If a business unit produces highly specific products or services or relies heavily on specific processes, its performance cannot be assessed by the performance of its peers.

[5]For example, top management could hire an expert who supervises the activity, or invest in specific knowledge about the activity for members of the top management team. Alternatively, the company could outsource or sell the activity (or in this case: the business unit) to a party who can control it more effectively.

In conclusion, this study predicts a positive effect of information asymmetry on RPE use, but only if the business unit's activities and processes are comparable to peers. I expect a positive effect of comparability on the relation between information asymmetry and RPE use. This expectation is modelled as a moderating effect of comparability of the business unit on the relation between information asymmetry and RPE use. This relation is explicated in hypothesis H2.

H2: Comparability of the business unit has a positive effect on the association between information asymmetry and the use of RPE.

2.2.3 Interaction Effect Between Common Uncertainty and Information Asymmetry

Finally, I hypothesize an interaction effect[6] between common uncertainty and information asymmetry that increases the benefits of RPE. Together, the two variables can rule out potential control alternatives for RPE and thereby increase RPE use. This hypothesis is not built on either one of the perspectives on RPE in particular, but combines the antecedents (uncertainty and information asymmetry) that follow from both perspectives.

To reduce noise and/or opportunism, organizations have a number of target-setting instruments at their disposal, one of which is RPE. Examples of these instruments include: participative budgeting, the use of a timeless performance target (such as a long-term ROI), and RPE. The usefulness of these target-setting instruments depends on the circumstances. This section describes how these target-setting instruments apply to various combinations of uncertainty and information asymmetry. For example, some target-setting instruments may be very useful in situations of high uncertainty and low information asymmetry or vice versa. Furthermore, these instruments may lose their applicability in a situation of combined high uncertainty *and* high information asymmetry. However, I argue that, in situations of combined high uncertainty and high information asymmetry, RPE remains a viable target setting method.

Earlier in this chapter I argued that, in a situation of high uncertainty, forecasting would be difficult due to unforeseen events. RPE can help to reduce noise in the performance evaluation by filtering out the performance effects of uncertainty - to the extent that the uncertainty is common to the peer group. However, if information asymmetry levels are

[6]This interaction effect is modelled as an 'independent-variable interaction', where both independent variables have an amplifying effect on their relation with the dependent variable (Luft & Shields 2003, Gerdin & Greve 2004).

sufficiently low, alternatives exist. The principal could cancel out the performance effects of external events on the standards by *ex-post* determining the appropriate target level together with the business unit manager (*i.e.*, participative budgeting), without 'consulting the market' via RPE.

In the situation of low uncertainty and high information asymmetry, RPE can reduce the risk of managerial opportunism. But if uncertainty is low, the principal also has other target-setting means at her disposal besides RPE that also mitigate opportunism. Murphy (2001) argues that other externally determined standards[7] -such as the use of timeless performance standards or firm-level cost of capital target- can reduce the risk of managerial opportunism.

However, in a situation of combined presence of high[8] uncertainty and high information asymmetry, the presented alternatives (*i.e.*, participative budgeting, timeless performance standards, and firm cost of capital) will not be viable methods for determining the performance target. Given the presence of high information asymmetry, participative budgeting suffers from the risk of managerial opportunism; the business unit manager can exploit his information advantage to negotiate easy targets. Furthermore, the presence of uncertainty introduces noise into the performance evaluation, which reduces the applicability of fixed, yet externally determined performance standards such as timeless performance standards and firm cost of capital. RPE, however, does not lose its applicability under the condition of combined high common uncertainty and high information asymmetry.

In conclusion, although RPE is a viable method for target setting, situations exist where alternative target-setting instruments are available. The question of whether these alternatives are equally efficient as RPE in these situations lies well beyond the scope of this research. However, in a combined setting of high common uncertainty and high information asymmetry, I expect the presented alternatives to lose their applicability, whilst RPE keeps its usefulness. This leads to more use of RPE. I formulate this synergetic effect of the presented antecedents to RPE use in the following hypothesis:

H3: Common uncertainty and information asymmetry have a combined positive effect on the use of RPE.

[7]Murphy (2001) labels timeless performance standards and company cost of capital as externally determined because these targets cannot be influenced by the business unit manager.

[8]The question of how high 'high' is exceeds the level of theoretical articulation of this argument. The hypothesis and analyses assume a linear relation without a predefined threshold for uncertainty and/or information asymmetry.

2.2.4 Control Variables

This subsection introduces the control variables, including: contractibility, emphasis on firm-level measures, and size and sector controls.

The first control variable refers to the *contractibility of performance* (abbreviated: contractibility). Contractibility is a prerequisite for any output-oriented control structure (Speklé 2001). As generally accepted in the literature, contractibility entails that: 1) goals can be specified unambiguously; 2) the organization is able to measure outputs in a reasonably undistorted manner (that is, the output metrics correspond rather well with the actual goals that the unit needs to accomplish); and 3) the responsible individuals (in this case, business unit managers) understand the process of transforming efforts into results (following: Speklé & Verbeeten 2008[9]). To make a meaningful comparison of the output of an agent to a performance standard, the performance needs to be contractible. Because RPE is based on the comparison of outputs (instead of inputs, processes or efforts), contractibility is a necessary condition for RPE use.

Second, the analyses control for the *emphasis on firm-level measures*. Performance information can be obtained with measures at various organizational levels. If performance metrics at the firm-level are used, the amount of noise in the performance evaluation of the business unit manager increases because the manager being evaluated cannot control other parts of the organization affecting this performance measure. Thus, the use of firm-level measures in the performance evaluation makes the manager responsible for additional uncontrollable events. I expect the choice to rely on RPE (for example for noise-reduction purposes) to be influenced by the inclusion of noisy performance measures, such as firm-level measures.

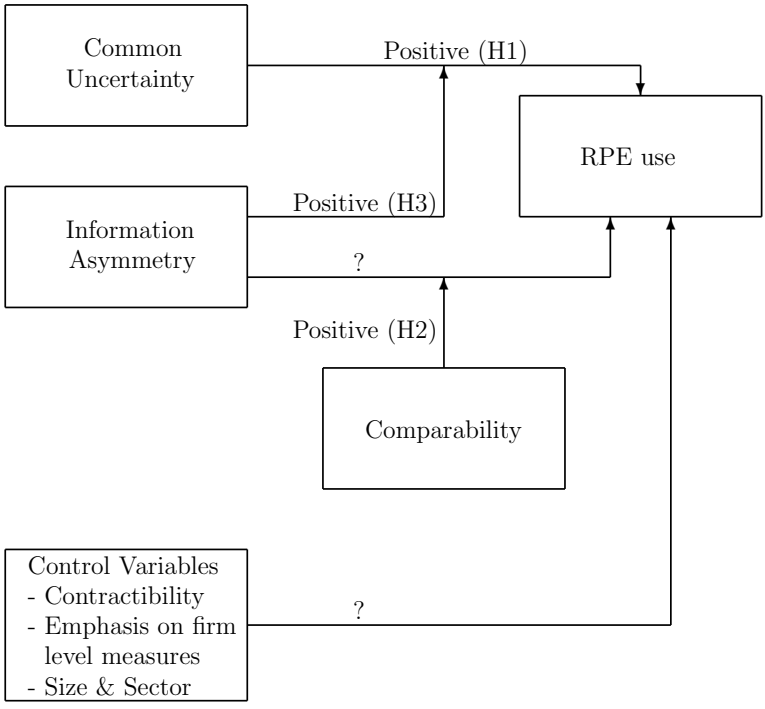
Finally, the analyses control for *size* and *sector* effects. Size is often regarded as a potentially important determinant of performance measurement practice. For example, small organizations might provide a more fertile ground for result-based performance measurement systems (Dewatripont *et al.* 1999). However, larger organizations might be more effective in the use and implementation of such systems (Rogers 1995, De Lancer Julnes & Holzer 2001). The sector control variables distinguish between production, financial services, service organizations, and not-for-profit business units.

[9]Speklé & Verbeeten (2008) derive the list of conditions for contractibility from several seminal papers that apply the conditions (amongst others) in contract oriented control structures, *e.g.* Otley & Berry (1980), Hofstede (1981), and Gibbons (1998).

2.2.5 Overview of the Model

To summarize the theory section, I present the causal model. The model consists of the hypothesized relations, and control variables (concerning contractibility, emphasis on firm-level measures, size and sector). The causal model is presented in figure 2.2.

Figure 2.2: Causal Model



2.3 Sample and Measurement

2.3.1 Sample

This study uses primary data collected from 325 business units with an extensive survey instrument. The questionnaire was filled out by business unit managers contacted through students' professional networks. This approach has several benefits. The first is respondent identification. The students were instructed to identify managers who are responsible

for an autonomous organizational body engaged in multiple activities (*e.g.*, purchasing, production, and sales). Furthermore, the respondent should have a superior within the organization (in other words, the respondent should not be the CEO). The second benefit is the possibility to conduct face-to-face interviews, which improves the respondent's valid interpretation of and careful responding to the questions. During these interviews, the student can explain the survey questions (using a strict protocol provided by the researcher) to the business unit manager who answers the questions in the survey and fills out the questionnaire.

The observations include data from production business units (17%), service BUs (56%), financial services (16%), and BUs operating in the not-for-profit sector (18%)[10]. Similarly diverse responses to the variance of sectors at the BU level were provided on the question about the respondents firm's sector. Table 2.1 summarizes the background information on the respondents' business units and the firm to which each business unit belongs.

Table 2.1: Summary Statistics on Sample

Panel A - size

		N	Mean	Std. dev. of mean	Min.	Max.
BU size	FTE	325	220	439	10	5.000
	Revenue (mio EURO)	292	196	840	1	1.000
Firm size	FTE	321	14.730	41.771	31	470.123
	Revenue (mio EURO)	301	7.775	43.240	4	600.000

Panel B - sector

BU sector		
	Production	54
	Services	181
	Financial	52
	Not-for-profit	57

Firm sector		
	Production	71
	Services	170
	Financial	62
	Not-for-profit	64

[10]These percentages sum to 107% because some business units operate in more than one sector. Multiple answers to the related question are considered valid.

2.3.2 The Questionnaire

The survey instrument was specifically designed for this research, following the Tailored Design Method (Dillman 2000), where applicable to the specific research setting[11]. The survey instrument relies heavily on previously used and validated questions. Some of those measures are well established (*e.g.*, the measure for information asymmetry is largely based on Dunk (1993)). Other questions were specifically constructed for this survey. For example, the measures for the dependent variable that measures the use of RPE are purpose-developed.

To increase the reliability of the instrument, the survey instrument was thoroughly pre-tested using Hak's Three-Step Test Interview method (abbreviated: TSTI) (Hak *et al.* 2008). This method is an observational instrument for pretesting self-completion questionnaires[12]. TSTI works through observing actual instances of interaction between respondents and the instrument. Because this process mainly consists of cognitive processing and is therefore hidden from the observer, 'thinking aloud' is used as a technique for making the thought process observable (Hak *et al.* 2008). Through this process, TSTI helps to identify problems in questionnaires, often leading to modification of the instrument. Three rounds of pre-tests with twelve managers[13] ensured a thorough and correct understanding of the questions and response categories.

2.3.3 Measurement

This section describes the measurement of the variables used for this study. Where references are made to questionnaire items, please see the appendix at the end of the thesis for the exact phrasing, response categories, and sources (where applicable).

[11]The Tailored Design Method is specifically designed for mail and internet surveys. Many of the guidelines provided by Dillman (2000) are applicable to the survey design, but others specifically apply to mail surveys instead of face-to-face survey interviews.

[12]Although the survey was not a self-completion questionnaire in the sense that the respondents did not fill out the survey by themselves, but together with the interviewing student, I still value the contribution of TSTI. Filling out the survey was outside of the direct sphere of influence of the researcher, but was carried out by -carefully briefed- students.

[13]The pre-test participants' organizations show great diversity, ranging from production to consulting, educational and financial services, and from very large to smaller business units, including both for-profit and not-for-profit organizations.

2.3.3.1 Dependent Variable: the Use of RPE

RPE use is measured in two different ways to capture different applications of RPE. The first measure (labelled 'RPE-Use') is broad and unspecific. RPE-Use measures the extent to which peer performance functions as a point of reference for evaluating the quality of the agent's performance. The underlying questionnaire items focus on the *ex-post* nature of performance evaluation. This measure comprises both explicit coupling of the performance target to peer performance and more implicit applications of RPE. Implicit applications of RPE do not require that peer performance affects the performance target explicitly, but they can, for example, play a role in establishing an implicit performance standard or norm for the performance evaluation. This measure asks to what extent the respondent perceives that the performance of his peers is a point of reference for his superior when evaluating the respondent's performance. Additionally, the question differentiates between situations where the evaluated business unit performs substantially better versus substantially worse than its peers. This distinction controls for potential asymmetries in peer comparison, as suggested by Garvey & Milbourn (2006). Throughout the sample, asymmetries do not seem to be of great influence; the three settings (neutral/better/worse) lead to a highly internally consistent scale, as presented in table 2.2[14].

Table 2.2: Items for RPE-Use (Q1)

	Item description	Component loadings
a.	Peer performance point of reference	0.868
b.	Substantially better	0.923
c.	Substantially worse	0.902
	Percentage variance explained	80.6%
	Cronbach's alpha	0.879

The second proxy for RPE measures specifically the explicit use of RPE for target-setting purposes; *i.e.*, how peer performance influences the difficulty of the performance standards against which the manager's performance is evaluated. This measure is labelled 'RPE-based-Targets'. I measure RPE-based-Targets as the impact of peer performance on the performance target. I ask to what extent peer performance influences the difficulty of the target and/or budget for a variety of performance measures (Q2). To interpret the impact of peers on the evaluation as a whole, the various performance metrics are weighted (Q3). I weigh the impact of peers on each of the metrics, by multiplying the scores by their individual importance for the performance evaluation. RPE-based-Targets does not measure solely *ex-post* peer comparison. The influence of peer performance on the agent's target

[14]The presented component loadings result from principal component analysis.

can also take place at the beginning of the period. In this case, the target for the upcoming period is based on the actual performance of peers during the current period. The *ex-ante* use of performance information about the foregoing period can be interpreted based on a broad interpretation of RPE: informing the principal about the quality of the agent's performance with market information. The RPE-based-Targets measure does not differentiate between *ex-ante* and *ex-post* peer comparison, and includes both types of information.

I expect RPE-Use and RPE-based-Targets to correlate, because they both include explicit RPE use and an *ex-post* application of the peer comparison information. Due to the differences between the measures as presented above, the correlation is not necessarily very strong. Correlation analysis confirms this; the two measures for RPE are correlated, but not very strongly (0.498; $p < 0.01$). Both measures will be used in the analyses.

2.3.3.2 Independent Variables

Common Uncertainty To obtain a measure for common uncertainty, the survey measures environmental uncertainty, and then eliminates the idiosyncratic part of the uncertainty from the measure[15]. Environmental uncertainty includes common and idiosyncratic uncertainty. For uncertainty to be common amongst a reference group, its source needs to be environmental. As opposed to uncertain events that occur within the organization, only events in the external environment can affect both the own business unit performance, and the performance of peers who share the same environment. However, environmental uncertainty is a broader concept than common uncertainty because the former includes both common and idiosyncratic uncertainty.

Environmental uncertainty is reflectively measured with a multi-item construct. The construct consists of six items about the impact of external factors, unpredictable events and difficulties in forecasting due to external factors. More specifically, the items refer to the frequency at which external events occur that substantially influence the performance of the BU or the amount of work, the occurrence of exceptions that require different methods or procedures, and how often the same day-to-day situations/problems arise. Factor analysis and reliability analysis produce one, internally consistent measure for environmental uncertainty, as displayed in table 2.3. Principal component analysis yields two factors. However, the first factor has adequate loadings for all individual factors, and scale reliability analysis produces only one reliable scale with a high Cronbach's alpha. Therefore, I

[15]The idiosyncratic part of uncertainty is eliminated through multiplication with a measure for the comparability of the business unit to indicate how similar the business unit is to other business units. This measure is described on page 31.

measure environmental uncertainty with one multi-item construct. The measure for environmental uncertainty is calculated as the mean of the six items[16].

Table 2.3: Items for Environmental Uncertainty (Q4-7)

	Item description	Component loadings (1)	Component loadings (2)
a.	External factors influencing business unit performance	0.544	0.472
b.	Unpredictable variation in the amount of work	0.667	-0.215
c.	Number of exceptions that arise in the unit	0.689	-0.496
d.	Differences in day-to-day situations	0.682	-0.423
e.	Need to react to outside pressure	0.672	0.326
f.	Long-range planning hindered by unpredictable events	0.602	0.495
	Percentage variance explained	41.6%	17.4%
	Cronbach's alpha	0.712	

However, by measuring environmental uncertainty, I measure both common uncertainty and idiosyncratic uncertainty. Therefore, the idiosyncratic part has to be filtered out of the uncertainty measure. To separate common from idiosyncratic uncertainty, I use information about how common the business unit is, compared to its peers. The logic behind this is the following: consider a very unique business unit, with assets that none of its peers has. For example, it has highly specific equipment, a high-profile brand image, and/or a highly diversified client portfolio, non of which its peers have. If an external event occurs that affects performance, it is likely that this event affects only this business unit and does not influence its peers, because its peers use different assets that affect performance in a likely different manner. *Vice versa*, if a highly comparable business unit, using common assets, faces external influences that affect its performance, it is likely that, besides the own performance also the performance of peers (who operate in more-or-less similar ways, using more-or-less similar assets) is affected by the event.

To measure how common the business unit is compared to other business units inside or outside of the firm, I develop a measure labelled 'comparability of the business unit'. Comparability of the business unit is discussed in the next section.

[16]Before calculating the mean score, 7-point scales are recoded to 5-point Likert scales, to prevent overrepresentation of the 7-point scaled items in the uncertainty construct.

Comparability of the Business Unit Comparability of the business unit (abbreviated: comparability) measures the degree of comparability of the business unit's assets to the assets of other business units or organizations.

Comparability is measured with seven items (Q8-14) relating to the comparability of a number of potentially critical assets to the assets of peers. These assets (or: resources) include: equipment, stock, systems, core-business employees, knowledge, brand image, and client portfolio. Because a high score on the individual measures indicates uniqueness as opposed to comparability, these scores are reversed.

Because the relative importance of each of these assets can vary amongst organizations to a large extent, respondents were asked to assign weights to the various assets (Q15). The weights, distributed as points over the various assets, indicate each asset's relative importance to the business unit. To calculate comparability of the business unit, the weighted scores are summed so that a high score on the measure indicates asset comparability.

Comparability of the business unit = (equipment * weight) + (stock * weight) + (systems * weight) + (core-business employees * weight) + (knowledge * weight) + (brand image * weight) + (client portfolio * weight)

Comparability is used to calculate the previously described common uncertainty variable. For this purpose, environmental uncertainty is multiplied by the 'comparability' score. This calculation yields a scale with a boosted uncertainty score if comparability is high and a reduced uncertainty score if comparability is low. The multiplied variables construct the measure for *common uncertainty* (CUNC).

Comparability also serves another role in this study. Comparability of the business unit is also used for hypothesis 3, where it functions as a stand-alone moderator of the effect of information asymmetry on RPE use. This approach does not lead to problems with multicollinearity, as discussed in section 2.4.3.

Information Asymmetry Information asymmetry is measured with a seven-item measure, largely based on Dunk (1993). There are two differences from Dunk's original instrument. First, the wording has been altered to fit better with the specific context of business unit managers. Second, one question was added to differentiate between knowledge about internal versus external factors that might influence the business unit performance (following Kruis 2008). All items load on one factor. Overall, the construct for information asymmetry shows good internal reliability. Information asymmetry was calculated by averaging the scores on all items. Table 2.4 summarizes the construction of the information asymmetry variable.

Table 2.4: Items for Information Asymmetry (Q16)

	Item description	Component loadings
a.	Information about the activities undertaken	0.818
b.	Familiarity with the input/output relations	0.834
c.	Certainty about the performance potential	0.851
d.	Familiarity with the technical aspects of the work	0.711
e.	The impact of internal factors on the managers activities	0.697
f.	The impact of external factors on the managers activities	0.610
g.	Understanding of the achievements of the business unit	0.791
	Percentage variance explained	58.3%
	Cronbach's alpha	0.875

Control Variables The study applies a number of variables for control purposes. The control variables include contractibility, firm-level performance measures, and size and sector controls. This paragraph will discuss all of the control variables subsequently.

Contractibility Contractibility issues are not specifically related to RPE as such, but rather to result oriented control instruments in general. To use result-oriented control instruments effectively, results have to be contractible. Generally, result oriented control instruments are considered to be effective if the following conditions have been satisfied simultaneously: 1) goals can be specified unambiguously in advance; 2) the organization is able to measure output in a reasonably undistorted manner (that is: the output metrics correspond rather well with the actual goals that the unit needs to accomplish); and 3) the responsible individuals (in this case, business unit managers) understand the process of transforming efforts into results (following: Speklé & Verbeeten 2008)[17].

Contractibility consists of a number of dimensions that act as necessary conditions for effective result-oriented control. All three dimensions together constitute contractibility. For

[17]Speklé & Verbeeten (2008) derive the list of conditions for contractibility from several seminal papers that apply the conditions (amongst others) in contract oriented control structures, *e.g.* Otley & Berry (1980), Hofstede (1981), and Gibbons (1998).

this reason, contractibility is a formative construct[18]. Leaving one or more dimensions out (as often happens with reflective items, for example, to boost scale reliability) alters the content and meaning of the construct. Contractibility is measured as the weighted summation of goal clarity, measurability of outputs and knowledge of the transformation process. For the calculation of the weights of the individual measures, I use a MIMIC model. The MIMIC model is presented in appendix A at the end of this chapter on page 52. The remainder of the current subsection describes the contractibility measure primarily from a conceptual instead of a statistical viewpoint.

The measurement of the individual underlying constructs (*goal clarity*, *measurability of outputs* and *knowledge of transformation process*) will be discussed next. Goal clarity is a multi-item construct, evaluated by four questionnaire items. These questions, stemming from Rainey (1983) and Kruis (2008) have been used together to measure goal ambiguity by Kruis (2008). I use these items to measure goal clarity (the exact opposite of ambiguity) by reversing the instrument and averaging the scores. The factor analysis and the Cronbach's alpha results agree with this approach, as shown in table 2.5 - Panel A.

Measurability of outputs is a construct evaluated by four items, including an item asking directly about the measurability of outputs and questions about the goal consistency and completeness of the output measures. The items load onto one factor with good scale reliability. The items and statistics are summarized in table 2.5 - Panel B.

The knowledge of the transformation process variable consists of two questionnaire items, which were averaged to compute the knowledge measure. The items ask about the transparency of the transformation process for the manager and his knowledge about adjustments of the transformation process if necessary. As shown in table 2.5 - Panel C, the two items combined have insufficient scale reliability (Cronbach's alpha 0.58)[19]. However, because the factor analysis shows high component loadings and a high level of variance explained, both items have been included in the knowledge of the transformation process-variable.

Together, *goal clarity*, *measurability of outputs* and *knowledge of transformation process* form the necessary conditions for contractibility of performance. The measures are summed according to the weights that are calculated with the MIMIC model, which can be found in appendix A at the end of this chapter on page 52.

[18]For the differences between formative and reflective indicators, and for a discussion of the consequences of these differences for model specification and measurement, see Bisbe *et al.* (2007), Diamantopoulos & Siguaw (2006), and Jarvis *et al.* (2003).

[19]Hair *et al.* (2010:92) claim that a Cronbach's alpha value of 0.60 is the lower limit of acceptability. Also Hair *et al.* (2010) argue that we need to consider multiple diagnostics to assess a measure's internal consistency, including inter-item correlations and confirmatory factor analysis.

Table 2.5: Items for Contractibility

Panel A - Items for Goal Clarity (Q17-20)

	Item description	Component loadings
a.	Clarity of goals	0.758
b.	Specificity of goals	0.734
c.	Difficulty of explaining goals to outsiders (reversed)	0.572
d.	Clarity of goals to insiders	0.682
	Percentage variance explained	47.6%
	Cronbach's alpha	0.630

Panel B - Items for Measurability of Outputs (Q21-23)

a.	Match between performance measures and business unit goals	0.731
b.	Objective measurability of goals	0.724
c.	Match between performance measures system and results	0.809
d.	Goals consistency performance measurement system	0.760
	Percentage variance explained	57.3%
	Cronbach's alpha	0.744

Panel C - Items for Knowledge of the Transformation Processes (Q24)

a.	Transparency of the transformation process	0.839
b.	Understanding of adjustments of the transformation process	0.839
	Percentage variance explained	70.3%
	Cronbach's alpha	0.580

Emphasis on Firm-Level Measures The emphasis on firm-level measures construct captures the reliance on performance measures that surpass the boundaries of the business unit. Firm-level performance measures also capture the performance of other organizational parts (*e.g.*, other business units over which the business unit manager has no control). Measures that include more than the organizational parts over which the business unit manager has control, introduce noise into the performance evaluation. I include the use of firm-level measures to control for this noise-increasing effect. 'Emphasis on firm-level measures' is measured as the summed weights of firm stock, firm value and firm profit measures in the performance measurement system (Q30).

Size (BU and Firm) The control variables business unit size (Q31) and firm size (Q32) are both based on the number of FTE and revenues. For both control variables, the natural logs of FTE and revenue were calculated. With business unit size measured as FTE and revenue, the Cronbach's alpha is rather low at a value of 0.634. However, the two items correlate positively ($r = 0.50$, $p < 0.01$). Additionally, the two factors load onto one component in the factor analysis (component loading = 0.866). Despite the low Cronbach's alpha, both items are combined into the measure business unit size. Although business unit size and firm size are significantly correlated with one another (Pearson correlation 0.479), they do not yield collinearity-issues in the multivariate analyses (as described in paragraph 2.4.3). Table 2.6 summarizes the questionnaire items and statistics on the size variables.

Table 2.6: Items for Business Unit Size & Firm Size (Q31-32)

	Item description	Component loadings
a.	Business unit FTE (Ln)	0.866
b.	Business unit revenue (Ln)	0.866
	Percentage variance explained	75.0%
	Cronbach's alpha	0.634
c.	Firm FTE (Ln)	0.952
d.	Firm revenue (Ln)	0.952
	Percentage variance explained	90.6%
	Cronbach's alpha	0.892

Sector Dummy Variables Murphy's (2001) findings indicate differences in RPE use at the executive level across industries. To filter out these effects from the analyses, I include dummy variables that distinguish between business unit sectors (Q33). The model contains dummies for production, financial services, and not-for-profit business units. The base model consists of service organizations, which represent the largest sector in the sample.

2.3.4 Descriptive Statistics

The descriptive statistics of the variables that are described above, are presented in table 2.7. The distribution analysis does not indicate normality issues for any of the variables. Kline (2005) and Widener (2007) suggest that skewness greater than three and kurtosis greater than ten may suggest a problem with the data. Kurtosis and skewness indicate that the data are within tolerable levels of univariate normality.

However, although the distribution of the RPE measures lies well within the boundaries of normality for the OLS regression (the skewness and kurtosis values are presented in table 2.7), the measures are actually left-censored. At the left end of the scale, a group of 21 (RPE-Use) or 19 (RPE-based-Targets) respondents do not use RPE at all. The rest of the responses are normally distributed and range from small use to extensive use of RPE. To address the censored nature of the dependent variables, I use Tobit regression as an additional analysis.

Table 2.7: Descriptive Statistics

	Mean	Min	Max	Std. Dev.	Skew-ness	Kurt-osis
RPE-Use	3.37	1.00	5.00	1.00	-0.85	0.31
RPE-based-Targets	3.49	1.00	5.00	1.04	-0.95	0.48
Common uncertainty	9.01	3.11	18.65	2.65	0.64	0.63
Interaction information asymmetry * comparability	-0.03	-2.69	2.95	0.86	0.23	1.96
Interaction uncertainty * information asymmetry	-0.01	-2.78	2.69	0.77	0.02	2.02
Information asymmetry	5.29	2.00	7.00	0.91	-0.61	0.57
Comparability	2.85	1.00	4.55	0.71	0.08	-0.31
Contractibility	-3.78	-14.76	3.30	3.06	-0.55	0.41
Emphasis on firm-level measures	0.16	0.00	1.00	0.21	1.83	3.46
Business unit size	10.27	2.30	14.97	2.26	-1.56	2.70
Firm size	13.41	3.69	19.41	2.43	-0.44	1.34

2.4 Analyses

This section presents the analyses. I test the hypotheses by using multivariate regression analysis[20]. The hypothesized main and interaction-effect variables are included as described previously. All interactions between information asymmetry and its moderator (*i.e.*, comparability) as well as the independent variable interaction between uncertainty and information asymmetry are calculated as multiplicative interactions (Gerdin & Greve 2008). Interaction effects are calculated by using the following procedure. First, all of the terms are standardized[21]. Then the terms are multiplied and added to the model.

In addition to the hypothesized main and interaction-effects, the statistical model consists of control variables (concerning contractibility, emphasis on firm-level measures, size and sector) and the unhypothesized main-effects of comparability and information asymmetry on RPE use. These main effects are included to ensure that the significant coefficients for the interaction effects are not caused by lower-order effects (see also: Hartmann & Moers 1999, Echambadi *et al.* 2006). The result is the statistical model presented in figure 2.3.

Overall, the regressions analyses provide mixed results with regard to the explanations of organizational reliance on RPE. The noise-based hypothesis (H1) is robustly supported, but the opportunism-based hypothesis (H2) and the joint effects of both perspectives (H3) yield inconsistent findings.

Before discussing the regression analyses in detail, the remainder of this section describes the prevalence of RPE amongst business unit managers and the correlations among the variables.

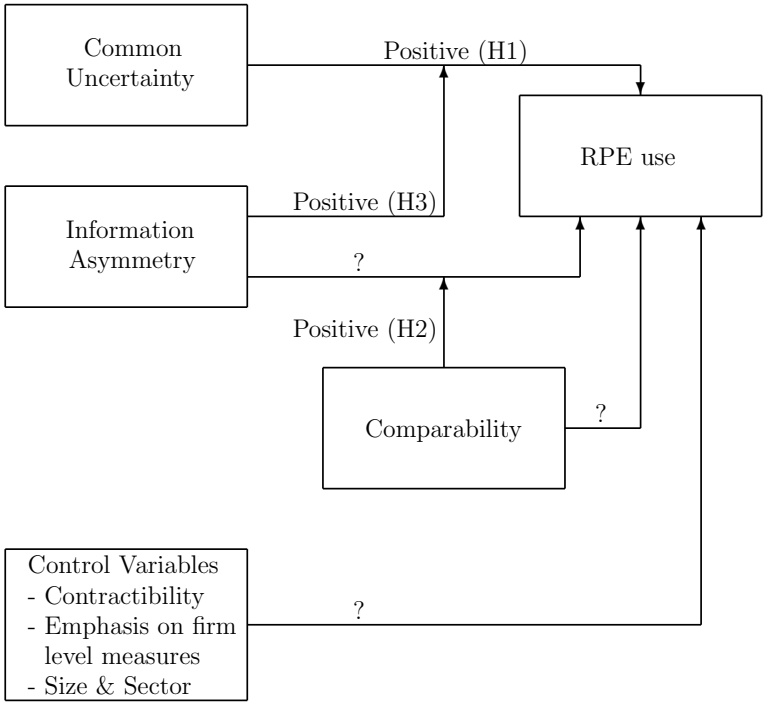
2.4.1 Empirical Results: the Use of RPE amongst Business Unit Managers

One of the goals of this study is to explore the use of RPE amongst business unit managers, as formalized in the introductory chapter of this thesis as the first research question. The answer to this first research question is discussed in the current subsection, which presents the extent to which RPE is used within the sample. Moreover, this subsection describes additional design characteristics of RPE use.

[20] For the multivariate analyses, I use EViews 6.

[21] Standardization and/or mean-centring of interaction terms is often used to alleviate multicollinearity problems, as recommended by Cronbach (1987) and Aiken & West (1991). According to Echambadi & Hess (2007), this is not a useful procedure. Echambadi & Hess analytically show that these linear transformations do not reduce the potential threat of multicollinearity. Rather, I used standardization to solve the inequality of scale-variances to prevent undesired weightings of the individual terms in the multiplied interaction construct.

Figure 2.3: Statistical Model



The data indicate that many organizations use peer performance information to evaluate the performance of their business unit managers, as summarized in table 2.8[22]. Based on the question regarding the extent to which peer performance functions as a point of reference for the respondents' performance evaluations, I find that approximately 88% of the business units in the sample use RPE to at least a small extent. Fewer than 8% of the respondents claim that peer performance does not influence their performance evaluations at all. In total, 52% of the respondents claim that peer performance is at least of great influence on their performance evaluations. The mean score of the broad and unspecific measure of RPE use (RPE-Use) is 3.37 on a scale ranging from 1 (no use) to 5 (maximum use), which indicates that RPE is part of the performance measurement systems of business unit managers. That is, peer performance is of some to great extent of influence on the performance evaluation. The descriptive statistics of the RPE-based-Targets construct (measuring explicit manifestations of RPE) provide a qualitatively similar picture regarding the dispersion of RPE amongst business unit managers, as shown in table 2.8.

[22]The figures in table 2.8 are presented as categorical data. For table 2.8 only, the original metric scales (as used throughout the chapter) are rounded to the nearest whole number.

Table 2.8: RPE use (RPE-Use and RPE-based-Targets)

	RPE-Use		RPE-based-Targets	
	Frequency	Percent	Frequency	Percent
No extent (1-1.5)	25	7.7%	23	7.1%
Little extent (1.5-2.5)	30	9.2%	19	5.8%
Some extent (2.5-3.5)	88	27.1%	63	19.4%
Great extent (3.5-4.5)	140	43.1%	126	38.8%
Very great extent (4.5-5)	28	8.6%	48	14.8%
Total valid	311	95.7%	279	85.8%
Missing	14	4.3%	46	14.2%

This study is not the first to empirically describe organizational reliance on RPE. Murphy (2001) and Gong *et al.* (2011) also explore to what extent organizations in various branches use RPE, albeit at the executive level. Dekker *et al.* (2012) study the use of (internal and external) benchmarking information for target setting at the middle-manager level. Comparing the descriptive statistics with prior findings suggests that RPE is used more at the lower echelons than at the executive level. Murphy (2001) and Gong *et al.* (2011) find that external-peer-group-based performance targets are frequently used amongst the top executives of large US corporations. However, the findings of my study indicate that these targets are used more extensively amongst middle-managers. Dekker *et al.* (2012) support the importance of relative targets by showing that they are widely used among middle-managers, especially if these targets are used for incentive compensation. Gong *et al.* report that approximately 25% of S&P 1500 firms explicitly use RPE in setting executive compensation. Providing more detailed information, Murphy finds that 6-15%[23] of the industrials use RPE. For the firms in the financial/insurance industry, RPE is used more extensively at 46-53%[23]. Within utility firms, 26-17%[23] of the firms set their executives' targets by using RPE.

Similar to Murphy's findings, the results of my study show differences across industries. In my sample, the business units in the financial services industry rely more on RPE for target setting purposes than the business units in other industries[24]. This corresponds to Murphy's findings. Also, RPE use for target setting purposes is lower than average

[23]The percentages refer to RPE used as the only performance criterion, or as part of a multitude of performance criteria, respectively.

[24]An independent sample T-test shows a significant difference in mean levels of RPE-based-Targets for financial industry respondents, compared to other business units. The means for RPE-based-Targets for financials and other industries are 3.75 and 3.44, respectively. The difference is statistically significant at a two-tailed p-value of 0.069 (T=1.829). However, the more general RPE-Use proxy shows no significant difference across industries.

amongst not-for-profit business units[25]. This is likely due to lower than average emphasis on output targets[26].

Although Murphy's (2001) findings indicate generally lower reliance on RPE, his findings do not contradict the results of my study, as both studies focus on different organizational levels, and use different operationalizations of RPE. Additionally, Murphy focuses only on external peer groups, whereas my data contain RPE using external, internal, and mixed peer groups.

My data show that most RPE adopters use both internal and external peer groups in their performance evaluation. Of the 325 respondents[27], 44% use the performance information from their internal peers (*e.g.*, other business units within the organization) and external peers. Table 2.9 presents the statistics on peer group selection.

Table 2.9: Use of Internal and External Peers within the Reference Group

	Frequency	Percent
Internal peers	94	28.9%
External peers	21	6.5%
Both internal/external peers	143	44.0%
Not applicable/no response	67	20.6%
Total valid	325	100.0%

[25]An independent sample T-test shows a significant difference in mean levels of RPE-based-Targets for not-for-profit business units, compared to other business units. The means for RPE-based-Targets for not-for-profits and other industries are 3.17 and 3.56, respectively. The difference is statistically significant at a two-tailed p-value of 0.013 (T=-2.528). Again, the more general RPE-Use proxy shows no significant difference across industries.

[26]An independent sample T-test shows a significant difference in mean levels of emphasis on targets (calculated as the mean of Q28 and Q29, for more details see the description of these questions on page 136) for not-for-profit business units, compared to other industries. The means of the emphasis on targets for not-for-profits and other industries are 3.08 and 3.80, respectively. The difference is statistically significant at a two-tailed p-value of 0.000 (T=-7.615).

[27]The number of included observations is presented in the tables that accompany the analyses. This number varies per analysis throughout the thesis because of the listwise deletion of missing values.

2.4.2 Correlations

Table 2.10 presents the correlation matrix of all of the metric variables[28]. The correlation table shows no significant theory-consistent associations between the dependent and independent variables. However, some significant correlations among the independent variables exist. I find a significant correlation between the two size variables (*i.e.*, business unit size and firm size). This finding might raise concerns regarding multicollinearity, because the Pearson correlation coefficient is 0.479. However, as discussed in the next paragraph (2.4.3), the multivariate models do not suffer from multicollinearity.

Table 2.10: Pearson Correlation Table

	1.	2.	3.	4.	5.	6.	7.
1. RPE-Use							
2. RPE-based-Targets	.519***						
3. Common Uncertainty	-.040	.004					
4. Information Asym.	.004	.014	-.039				
5. Contractibility	.205*	.144**	-.005	-.019			
6. Firm-level Measures	-.057	-.010	-.031	.086	-.064		
7. Business Unit Size	.002	.023	.068	.057	-.025	.066	
8. Firm Size	.039	.042	-.052	.135**	.118*	-.040	.479***

*** Correlation is significant at the 0.01 level (2-tailed).
** Correlation is significant at the 0.05 level (2-tailed).
* Correlation is significant at the 0.10 level (2-tailed).
Listwise N=244

2.4.3 Multivariate Analyses

This subsection presents the ordinary least squares regression analyses for the two dependent variables (RPE-Use and RPE-based-Targets). Additionally, Tobit and Logit regression analyses are included as additional analyses to improve and buttress the findings. The Tobit models have a slightly better fit because Tobit was designed to estimate limited dependent variables, such as the RPE use variables. However, the qualitative results are similar to those of the OLS models. All models are significant and show no indications of multicollinearity[29].

[28]This matrix excludes the sector controls. Because these controls are nominal variables, a Pearson correlation would not be appropriate.

[29]Variance Inflation Factors in the models are below 2.5.

2.4.3.1 Analyses with RPE-Use Measure

First, I test the model with RPE-Use, the broad measure of RPE that captures both implicit and explicit influence of peer performance on the performance evaluation. I find partial support for the hypotheses, as shown in table 2.11[30]. The OLS estimation shows that the model is significant and fits the data, albeit marginally. The reported R^2 is 7.5% (Adjusted R^2 = 3.1%). This model's ANOVA F-statistic is 1.718 (p = 0.063). As an additional analysis, a subsample of for-profit business units is analysed. The results are presented in appendix B at the end of this chapter on page 57. Potentially because of the limited size of the for-profit subsample, the resulting model is not significant (N = 224, ANOVA F-statistic is 1.461, p = 0.148).

Table 2.11: Results of OLS Regression Analysis (RPE-Use)

	Coefficient	Std. Error	t-Stat.	Prob.
Constant	3.445	0.715	4.816	0.00
Common uncertainty (H1)	0.358	0.269	1.333	0.09[a]
Interaction information asymmetry * comparability (H2)	0.126	0.096	1.332	0.09[a]
Interaction uncertainty * information asymmetry (H3)	-0.115	0.110	-1.070	0.14[a]
Information asymmetry	0.073	0.071	1.032	0.30
Comparability	-1.771	0.961	-1.841	0.06
Contractibility	0.049	0.021	2.287	0.02
Firm-Level Measures	-0.919	0.349	-2.629	0.00
Size of BU	0.025	0.032	0.800	0.42
Size of firm	-0.003	0.029	-0.121	0.90
Dummy production BU	0.122	0.194	0.626	0.53
Dummy financial services BU	0.071	0.154	0.460	0.64
Dummy not-for-profit BU	0.078	0.205	0.382	0.70
R^2 = 0.075			F-statistic = 1.718	
Adjusted R^2 = 0.031			Prob(F-statistic) = 0.063	

Included observations: 267

'a': variable based on directional hypothesis
significance calculated as one-tailed p-value

Both the effects of common uncertainty (hypothesis H1) and the interaction between information asymmetry and comparability of the business unit (hypothesis H2) hold in this model, although the effects are only marginally significant at t-values of 1.333 and 1.332, respectively. The combined effect of uncertainty and information asymmetry (H3) is not supported by the analysis.

[30]The presented p-values in tables 2.11 & 2.13 are one-tailed if the underlying hypothesis is directional. The one-tailed findings are marked as 'a'.

Concerning the control variables, I find a negative and significant influence of the unhypothesized main effect of comparability and the use of firm-level measures on RPE use. This former finding is counterintuitive. Although this study does not formulate explicit expectations concerning the direct effect of comparability on RPE use, it is plausible that a potential significant relation is positive. This is because comparability is a necessary condition for peer comparisons. Bivariately, the two variables are not significantly correlated (r = -0.08, p = 0.19).

Furthermore, the results show a positive significant effect of contractibility. Other control variables (business unit and firm size and the sector dummies) do not yield significant results. Neither does the information asymmetry variable, which was included to rule out the possibility that the significant coefficient of the interaction effect between information asymmetry and uniqueness is caused by a lower-order effect (*i.e.*, a main effect of comparability on RPE-Use) (conform: Hartmann & Moers 1999, Echambadi *et al.* 2006).

However, the data are censored. A group of 21 respondents does not use RPE. To better suit the left-censoring characteristic of the dependent variable RPE-Use, I estimate a Tobit model, which is a regression technique especially suited to analysing limited (censored) dependent variables. The Tobit model yields results similar to those of the OLS analysis. This model supports the hypotheses regarding the effects of common uncertainty (H1) and the interaction effect between information asymmetry and comparability (H2). To facilitate the comparison with the results of the OLS model, I report the McKelvey-Zavoina statistic. This statistic is a pseudo-R^2 measure that comes closest to the familiar R^2 from the OLS regression (Veall & Zimmermann 1994). According to the McKelvey-Zavoina statistic, the Tobit analysis explains 7.9% of the variance in RPE-Use. The results of the Tobit analysis are qualitatively similar to those of the OLS analysis, as is shown in table 2.12.

2.4.3.2 Analyses with RPE-based-Targets Measure

I also test the model with the RPE-based-Targets measure. RPE-based-Targets relates to the explicit impact of peer performance on the performance target difficulty. This model yields better test-statistics. As presented in table 2.13, I find a significant model that fits the data. The ANOVA F-statistic is higher than the RPE-Use model ($p < 0.01$), and the reported R^2 is higher at 12.2%. The qualitative results differ partially from those of the analyses with explicit RPE use. Similar to the previous model, I find a significant positive effect of common uncertainty (hypothesis H1), but the hypothesized interaction of information asymmetry and comparability (hypothesis H2) does not hold in this model. Instead, this model supports the interaction between common uncertainty and information asymmetry (hypothesis H3). The results are shown in table 2.13.

Table 2.12: Results of Tobit Regression Analysis (RPE-Use)

	Coefficient	Std. Error	t-Stat.	Prob.
Constant	3.315	0.634	5.226	0.00
Common uncertainty (H1)	0.373	0.283	1.316	0.09[a]
Interaction information asymmetry * comparability (H2)	0.134	0.102	1.306	0.10[a]
Interaction uncertainty * information asymmetry (H3)	-0.130	0.116	-1.116	0.13[a]
Information asymmetry	0.080	0.075	1.072	0.28
Comparability	-1.835	1.010	-1.818	0.07
Contractibility	0.049	0.023	2.130	0.03
Firm-Level Measures	-0.997	0.370	-2.690	0.01
Size of BU	0.031	0.034	0.919	0.36
Size of firm	-0.009	0.031	-0.294	0.83
Dummy production BU	0.058	0.177	0.332	0.77
Dummy financial services BU	0.006	0.173	-0.034	0.97
Dummy not-for-profit BU	0.049	0.183	0.270	0.79
McKelvey-Zavoina statistic = 0.079				
Left censored observations: 21			Right censored observations: 0	
Uncensored observations: 246			Total observations: 267	

'[a]': variable based on directional hypothesis
significance calculated as one-tailed p-value

Similar to the analyses with RPE-Use, the current model is also analysed with a subsample consisting of for-profit business units only. The results are presented in appendix B at the end of this chapter on page 57. Although the results explain less variance in the dependent variable RPE-based-Targets than the full sample analysis, the model is significant (N = 200, ANOVA F-statistic is 1.884, p = 0.044). The reduced explanatory power ($R^2 = 9.9\%$) is probably the result of the limited size of the for-profit subsample. The subsample analysis supports the qualitative findings of the full sample analysis of the RPE-based-Targets model.

Again, I find that the unhypothesized main effects of comparability (negative) and of contractibility (positive) significantly influence RPE use. This repeated finding renders these unhypothesized effects more robust against the different operationalizations of RPE use. The other results from the control variables are not significant, consistent with the RPE-Use model. The Tobit model, which corrects for the effects of the 19 observations of non-adopters of RPE-based-Targets, yields similar qualitative results. The corresponding McKelvey-Zavoina pseudo R^2 is 13.5%. The Tobit model shows qualitative similar results as the base model. The Tobit model is tabulated in table 2.14 on page 46.

Table 2.13: Results of OLS Regression Analysis (RPE-based-Targets)

	Coefficient	Std. Error	t-Stat.	Prob.
Constant	4.352	0.673	6.459	0.00
Common uncertainty (H1)	0.964	0.283	3.402	0.00[a]
Interaction information asymmetry * comparability (H2)	0.001	0.105	0.000	0.49[a]
Interaction uncertainty * information asymmetry (H3)	0.229	0.116	1.978	0.03[a]
Information asymmetry	-0.026	0.076	-0.345	0.73
Comparability	-3.851	1.023	-3.765	0.00
Contractibility	0.067	0.023	2.900	0.00
Firm-Level Measures	0.085	0.412	0.208	0.84
Size of BU	0.012	0.036	0.355	0.72
Size of firm	-0.008	0.032	-0.252	0.80
Dummy production BU	-0.279	0.204	-1.372	0.17
Dummy financial services BU	-0.102	0.165	-0.616	0.54
Dummy not-for-profit BU	-0.307	0.205	-1.502	0.13
$R^2 = 0.122$			F-statistic = 2.653	
Adjusted $R^2 = 0.076$			Prob(F-statistic) = 0.002	

Included observations: 242
'a': variable based on directional hypothesis
significance calculated as one-tailed p-value

2.4.3.3 Logit Model Robustness Check

To further buttress the findings of the base model analysis, I conduct two Logit analyses. I regress the hypothesized and control variables on dichotomized versions of both measures of RPE use. The Logit model predicts whether individual respondents are either RPE adopters or non-/low-RPE users[31], based on their score on the hypothesized and control variables. The logic behind using a Logit regression is that a fundamental distinction may exist between high (or intensive) RPE use and no/less intensive RPE use. As presented in section 2.4.1, almost all organizations consider the performance of peers when evaluating their employees. However, relying on RPE to a greater degree may be a deliberate choice that is driven by noise- and opportunism-reduction considerations, as predicted by the theory. In contrast, relying less on RPE by merely incorporating 'the world that's out there' into the background of the performance evaluation, may simply be 'common practice', not driven by the specific problem-solving capabilities of RPE per se. Logit regression analyses require the existence of two distinct groups and focus the analyses on this distinction.

[31]The mean scores of the RPE-Use and RPE-based-Targets scales are used as cut-off values. The cases with observed RPE scores at or below the mean are considered non-/low-RPE users. Higher scores indicate a high degree of RPE adoption. Additionally, analyses are run using the medians and modes of the RPE scales. These analyses yield similar results.

Table 2.14: Results of Tobit Regression Analysis (RPE-based-Targets)

	Coefficient	Std. Error	t-Stat.	Prob.
Constant	4.197	0.683	6.146	0.00
Common uncertainty (H1)	1.016	0.297	3.423	0.01[a]
Interaction information asymmetry * comparability (H2)	-0.015	0.111	-0.144	0.44[a]
Interaction uncertainty * information asymmetry (H3)	0.261	0.122	2.129	0.02[a]
Information asymmetry	-0.037	0.080	-0.466	0.64
Comparability	-3.984	1.069	-3.725	0.00
Contractibility	0.071	0.024	2.897	0.00
Firm-Level Measures	-0.073	0.433	-0.171	0.86
Size of BU	-0.020	0.038	-0.510	0.61
Size of firm	-0.009	0.033	-0.270	0.79
Dummy production BU	0.185	0.186	-0.998	0.31
Dummy financial services BU	0.241	0.188	1.285	0.19
Dummy not-for-profit BU	-0.222	0.188	-1.176	0.24
McKelvey-Zavoina statistic = 0.135				
Left censored observations: 19			Right censored observations: 0	
Uncensored observations: 223			Total observations: 242	

'[a]': variable based on directional hypothesis
significance calculated as one-tailed p-value

However, a drawback of this method is that it ignores a great deal of information about RPE use within these groups, which renders the analyses statistically less powerful.

The Logit model that predicts RPE-adoption based on the RPE-Use measure is significant ($chi^2 = 25.8$, $p = 0.011$). A significant chi^2 indicates that the predictors have a significant effect. The model fit is also confirmed by an insignificant Hosmer & Lemeshow statistic ($p = 0.593$). Similar to the base model, the explained variance indicators are low. The Cox & Snell R^2 is 0.09 and the Nagelkerke R^2 is 0.121. The model correctly predicts 61.2% of the cases whereas an 'empty' model that contains no predictors only corrects 53.1%. This difference is a small but significant increase. The second model, which uses a dummy variable for RPE-based-Targets, also fits the data. The model is significant ($chi^2 = 33{,}2$, $p = 0.001$, Hosmer & Lemeshow $p = 0.390$). The Cox & Snell R^2 is 0.127 and the Nagelkerke R^2 is 0.172. The model correctly predicts 70.6% of the cases, whereas an 'empty' model that contains no predictors correctly predicts 61.6%. This finding is, again, a small but significant increase. The test statistics are summarized in table 2.15.

Qualitatively, both Logit models yield results that are largely similar to those of the overall OLS findings. Similar to the OLS analysis of RPE-Use, the Logit analyses of both the

Table 2.15: Summary Statistics of Logit Regression Analysis

Panel A - RPE-Use

Correct classification (empty model)	53.1%
Correct classification (full model)	61.2%
$Chi^2 = 25.8$ (p = 0.011)	Homer & Lemeshow p = 0.593
Cox & Schnell $R^2 = 0.090$	Nagelkerke $R^2 = 0.121$

Panel B - RPE-based-Targets

Correct classification (empty model)	61.6%
Correct classification (full model)	70.6%
$Chi^2 = 33.2$ (p = 0.001)	Homer & Lemeshow p = 0.390
Cox & Schnell $R^2 = 0.127$	Nagelkerke $R^2 = 0.172$

RPE-Use and RPE-Based-Targets support the first and second hypotheses, at the 10% level (one-tailed significance of the Wald statistic). These hypotheses refer to the effect of common uncertainty on RPE use, and the interaction effect of information asymmetry and comparability on RPE use, respectively. However, neither of the Logit analyses support the third hypothesis, which refers to the interaction effect of common uncertainty and information asymmetry on RPE use. Although these results are not entirely similar to the base model's results, the findings substantiate the finding that both the noise- and opportunism-reduction perspectives on RPE, at least to some extent, help to clarify whether firms use RPE or not.

2.4.4 Summary of the Findings

The noise- and opportunism-based explanations of RPE use find (at least some) support in either one of the models. However, the primary (OLS and Tobit) analyses show only partially consistent findings for both operationalizations of RPE. Firstly, the positive effect of common uncertainty on RPE (H1) is robustly supported by all of the analyses. Secondly, I find only partial support for the hypothesized positive effect of the interaction between information asymmetry and comparability of the business unit (H2). Only the RPE-Use model supports this hypothesis. The opportunism mitigating characteristics of RPE (which is the underlying motivation for this hypothesis) may be more prevalent if RPE is not used for explicit performance target setting, but rather on the background of the performance evaluation. More generally, the marginal support for the opportunism-reducing characteristics of RPE may be due to the limited use of external peer groups (as presented in section 4.1). Both internal and external peer groups may be relevant for noise-reduction, as both groups can be subject to similar noise-inducing external events.

However, the reference groups with internal peers may be more sensitive to the risk of collusion: managers may collaborate with other managers to jointly limit their individual performances in order to achieve an easier benchmark. External peer groups may be less sensitive to this behaviour because of, for example, less intensive and more competitive relationships amongst peers. In chapter 4 of this dissertation, I study more in-depth the various applications of RPE and their effects on opportunism.

Additionally, the third hypothesis concerning the combined effect of uncertainty and information asymmetry on RPE use, is only supported by one operationalization (RPE-based-Targets). However, this finding can be explained by the nature of the RPE measure. The current measure (RPE-based-Targets) specifically focuses on the determination of explicit performance targets/budgets, whereas RPE-Use is less specific. The third hypothesis builds on the idea that organizations can choose from a variety of performance target-/budget-setting methods, of which RPE is one. Because both the hypothesis and the RPE-based-Targets measure refer to explicit target setting, the RPE-based-Targets model is the most suitable model for supporting this hypothesis.

With respect to the control variables, the results consistently show that comparability of the business unit and the contractibility of performance have significant effects. Comparability of the business unit is negatively associated with the use of RPE. This finding is counterintuitive. Although this study does not formulate explicit expectations concerning the direct effect of comparability on RPE use, it is plausible that a correlation between these two variables would be positive because comparability is a necessary condition for peer comparisons. Bivariately, the two variables are not significantly correlated with one another. This finding indicates that this negative correlation is due to covariances with other variables in the multivariate analyses. Contractibility is positively associated with the use of RPE. Although I did not hypothesize this correlation, it seems plausible. Contractibility can function as a necessary condition for result oriented control structures/instruments, such as RPE. The use of firm-level measures is only negatively associated with RPE in the RPE-Use model. The analyses do not yield any other significant influences of the control variables. The use of RPE is not influenced by (business unit or firm) size or sector. Additionally, the emphasis on targets in the performance measurement system has no effect on the use of RPE.

The Logit models consistently support the noise- and opportunism-reduction based research hypotheses (H1 and H2) but not the third hypothesis concerning the joint effects of uncertainty and information asymmetry. The Logit models provide additional support for both RPE-explanations, noise and opportunism. However, the explanatory power of the Logit analyses is limited. Additionally, the Logit analyses contain less information about the degree to which RPE is used, as the analyses distinguish only between no/low use and high use. The reduced informativeness of the dependent variable demands some caution for the interpretation of the findings.

2.5 Conclusions and Discussion

As one of the first empirical studies on the use of Relative Performance Evaluation (RPE) at the middle-management level (see also: Dekker *et al.* 2012), this research examines why and to what extent RPE has empirical relevance for the performance evaluation praxis of business unit managers. RPE is a means of determining the performance standard by using a peer group benchmark. The performance of an employee (or: business unit manager) is compared with the performance of a peer group. Peer performance functions as an (implicit or explicit) performance target. The peer group typically consists of individuals facing similar tasks and circumstances. The rationale behind this comparison is that the performance of the reference group provides information on the performance potential of the employee. RPE can be used as an explicit performance target or budget based on peer performance. However, RPE can also be used more implicitly in the background, where peer performance is used as a point of reference for the performance evaluation.

Prior literature has extensively studied RPE from both analytical and empirical standpoints. Despite ample analytical proof of the benefits of RPE, the empirical support for RPE is scarce. Focusing on the executive level, prior studies provide mixed results concerning organizational reliance on RPE. In contrast to the extant literature, the current study does not examine executive compensation praxis with archival data. Instead, I use detailed survey data to study the use of RPE at the lower echelons. I find that RPE is widely used to evaluate business unit managers, as more than half of the 325 respondents in this study claim to use RPE to a large or very large extent.

This study explains the use of RPE along two theoretical lines. The first line of reasoning is the noise-reduction perspective. The noise-reduction perspective on RPE is common in the literature and states that RPE filters out noise due to external events from the overall performance evaluation (Holmstrom 1982). These external events (*e.g.*, rising prices of raw materials or declining market demand) affect not only the performance metrics of the employee but also the performance metrics of the reference group. This effect increases noise in the performance metrics, which no longer reflect only effort. Instead, the metrics also reflect the performance effects of uncontrollable external events. Comparing the performance of a manager to the performance of peers informs the superior about the quality of the employee's performance, given the external events that have occurred. This comparison reduces the noise levels in the performance evaluation. This study finds support for this noise-reducing effect of RPE by analysing whether increased levels of common uncertainty (as a proxy for shared external events) positively affect the reliance on RPE.

The second explanation for RPE use is its ability to reduce the evaluated manager's room for opportunistic behaviour. This opportunism-mitigating perspective has received little

attention in the empirical literature. Following Murphy (2001), I argue that RPE can reduce the risk of managerial opportunism because RPE determines the difficulty of the manager's performance target outside his own sphere of influence. Murphy (2001) argues that managers can try to lower their performance targets if this lies within their reach. Managers can do so if, for example, the performance targets are based on prior-year performance or internal negotiations. However, if the target-setting process lies outside the manager's sphere of influence, the target is better insulated against such opportunistic behaviour. This is the case with RPE, where the performance target is based on the performance of a peer group and not on some internal process. According to this reasoning, RPE mitigates the room for managerial opportunism. I analyse whether information asymmetry (as a proxy for room for opportunism) combined with sufficient comparability to other organizations (as a necessary condition for relevant comparisons) positively affects the reliance on RPE. However, I find only marginal and mixed support for the opportunism-reducing effect of RPE. In chapter 4 of this thesis, I analyse in greater detail the opportunism-mitigating properties of RPE use to further study this perspective on RPE use.

This study suggests that RPE is used in practice to reduce noise in the performance evaluation of managers. Noise reduction improves the quality of the performance evaluation and the efficiency of the performance contract between the principal and agent. Albeit more carefully because of the inconsistency in the findings of this study, the findings suggest that RPE may help to reduce the room that managers have to opportunistically influence their performance targets. This function is important because the presence of room for opportunism potentially harms the realization of organizational goals. However, the evidence for this application of RPE is inconsistent and requires further research. In the remainder of this thesis, I study whether RPE is effective at reducing noise and opportunism to further examine these perspectives on RPE.

As with any study, this research is subject to a number of limitations. For instance, this study does not assess implications that RPE use may have on incentive compensation (Dekker *et al.* 2012).

Also, there are several data-related limitations. For example, the data are cross-sectional, which limits the research to analysing the associations between reliance on RPE and the drivers of noise and opportunism. Instead, a study of the longitudinal effects of RPE adoption would allow one to more directly observe the causality between these factors. Additionally, the data limit further exploration of the marginal evidence for the opportunism-reducing characteristics of RPE. The marginal evidence for the opportunism perspective may be caused by the risk of collusion amongst the evaluated managers. Holmstrom (1982) argues that RPE is vulnerable for collusion amongst agents. If agents collude, they collectively reduce their efforts, resulting in a lowered performance benchmark. This possibility

seems especially relevant in situations in which the agent's reference group consists of internal peers. Managers may collaborate rather easily with other managers within the same organization. If the evaluated manager can influence his relative performance target by colluding with other (internal) peers, the target can no longer be considered a purely external target. In contrast, external peer groups may be less vulnerable to collusion because they have less intensive and more competitive relationships amongst the peers. Using internal peer groups would not reduce the room that managers have to behave opportunistically. Unfortunately, the survey instrument contains only limited information about the use of external versus internal peer groups. More detailed data about the peer group composition would allow for further analysis of the effect of peer group choice on opportunism reduction. Despite these limitations, this study adds to our knowledge regarding RPE specifically, and performance target determination in general.

2.6 Contractibility MIMIC Model (Appendix A)

This appendix describes the estimation of the contractibility measure, which is used in the analyses in chapter 2. Contractibility is a requirement for the effectiveness of result-oriented control systems. According to Speklé & Verbeeten (2008), contractibility consists of three dimensions that need to be met simultaneously for the effective use of output controls, such as RPE. These dimensions are the following: 1) the ability to unambiguously specify goals *ex-ante*, 2) the ability to measure outputs in a reasonably undistorted manner (*i.e.*, the output metrics correspond rather well with the actual goals that the unit needs to accomplish), and 3) the responsible individuals (in this case, the business unit managers) need to understand the process of transforming efforts into results[32]. In section 2.3.3.2, these three dimensions are operationalized as *goal clarity*, *measurability of outputs* and *knowledge of transformation process*. Additionally, section 2.3.3.2 describes the underlying survey items and descriptive statistics of these measures. Because the three dimensions simultaneously 'cause' contractibility, contractibility is a formative construct.

Contractibility is not observed directly by the survey-instrument. Instead, contractibility is estimated as a latent construct by using a MIMIC model. MIMIC (Multiple Indicators for Multiple Causes) is a special case of structural equation modelling. The estimation is based on a number of measures, including measures that capture the three dimensions presented above. To specify the latent construct 'contractibility', one must first determine the nature and direction of the relationships between the construct and its indicators (Bisbe *et al.* 2007). Contractibility is estimated with two groups of constructs that differ in these aspects.

The first group consists of formative indicators that capture the three dimensions *goal clarity*, *measurability of outputs* and *knowledge of transformation process*. Formative indicators are causal in nature. That is, increases of the formative indicators result in higher degrees of contractibility. Together, the three dimensions 'define' contractibility. Leaving out one or more dimensions (*e.g.*, the unambiguous specification of goals) alters the content and meaning of the construct. If I only measure contractibility based on the two remaining dimensions, the resulting construct would be theoretically different from the abovementioned definition of contractibility. This important difference distinguishes formative constructs from the more commonly used reflective constructs[33]. Whereas with formative constructs, a census of indicators is required to describe the construct, reflective constructs may be

[32]Speklé & Verbeeten (2008) derive the list of conditions for contractibility from several seminal papers that apply the conditions (amongst others) to contract-oriented control structures, *e.g.* Otley & Berry (1980), Hofstede (1981), and Gibbons (1998).

[33]For a more elaborate discussion of the differences between formative and reflective constructs as well as the consequences of these differences for model specification and measurement, see: Jarvis *et al.* (2003), Diamantopoulos & Siguaw (2006), and Bisbe *et al.* (2007).

described by a sample of interchangeable indicators (Bisbe *et al.* 2007:792). With reflective indicators, one or more indicators are often excluded from the construct. This exclusion can be the result of a factor analysis and/or scale reliability analysis if a poorly fitting survey item is left out of the construct to enhance scale reliability. When measuring with formative indicators, this approach is not recommended because it alters the meaning of the construct.

The second group of constructs consists of reflective indicators. Whereas formative constructs cause the latent construct, the reflective measures are its *result*. The reflective measures in this model are the following: difficulty to assess work (Q25, reversed coding), difficulty to reach agreement on output quality (Q26, reversed coding), and knowledge about the business unit manager's job performance (Q27). In this case, the reflective indicators are all single-item constructs[34]. Table 2.16 presents the descriptive statistics for the formative and reflective contractibility indicators.

Table 2.16: Descriptive Statistics for Contractibility Indicators

	Mean	Min	Max	Std. Dev.	Skewness	Kurtosis
Goal clarity	2.34	1.00	5.00	0.57	0.43	1.19
Measurability of outputs	4.48	2.00	6.00	0.69	-0.45	0.38
Transformation process	4.95	2.50	7.00	0.93	-0.48	-0.04
Difficulty to assess work	2.58	1.00	5.00	0.78	0.38	0.08
Difficulty to reach agreement on output quality	2.58	1.00	5.00	0.66	0.18	0.12
Knowledge about job performance	3.42	1.00	5.00	1.05	-1.06	1.62

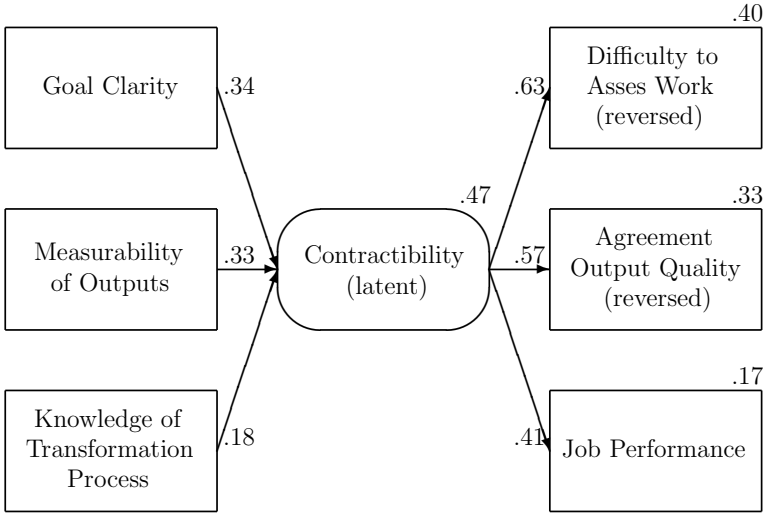
Formative constructs, such as contractibility, can be calculated by using various approaches. For one, it is possible to calculate a formative construct as a summated scale of the underlying formative indicators only. This calculation is achieved by summating the unweighted scores of the formative indicators. The resulting scale includes all of the underlying indicators, which is important because all of the indicators function as necessary conditions for the formative construct. Because the scores on the formative indicators do not need to covary, factor analysing the components of the formative construct is meaningless, as is the Cronbach-type analysis of scale reliability (Speklé & Verbeeten 2008). Because oftentimes no theoretical suggestions exist concerning how the underlying indicators should be weighted, this unweighted approach is justifiable.

[34]Because the reflective indicators are all single-item constructs, no factor analyses and/or scale reliability measures are reported.

I calculate contractibility not by using an unweighted summated scale but by using a more sophisticated MIMIC model. MIMIC also incorporates reflective indicators to calculate the relative weights of each of the formative indicators when forming the latent construct. As the relative weights of the three causal dimensions (goal clarity, measurability of outputs and knowledge of transformation process) cannot be derived theoretically, one may resort to structural equation techniques, such as MIMIC, to determine the appropriate weights (Bisbe *et al.* 2007). In the absence of a theoretical basis for the weights of the indicators, MIMIC at least provides an empirical basis for the index.

MIMIC estimates the latent formative construct contractibility, as depicted in figure 2.4. The MIMIC model calculates contractibility by using the formative (causal) indicators on the left-hand side of the figure and the reflective (outcome) indicators on the right-hand side of the figure. The numbers on the arrows are standardized regression coefficients. The numbers above the reflective measures (boxes) and the latent construct (contractibility) indicate the amount of variance explained. The model explains 47% of the variance in the latent construct contractibility.

Figure 2.4: Contractibility MIMIC Model



Overall, the MIMIC model fits the data. All of the formative and reflective indicators are significantly different from zero at a confidence level of 5%, as displayed in table 2.17 - Panel A. This finding implies that all of the indicators need to be included in the index. Table 2.17 also provides goodness-of-fit statistics for the model. To assess the fit of a SEM, one must assess multiple measures (Kline 2005). As the indicators of model fit, I use the chi^2, the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the comparative fit index (CFI), the root mean square residual (RMR) and the root mean square error of approximation (RMSEA). Kline (2005) suggests that the following values indicate a good fit: an insignificant chi^2; a GFI, AGFI, and CFI greater than 0.90; a RMR close to 0; and an RMSEA less than 0.045. As shown in table 2.17 - Panel B, the indicators show adequate model fit.

Table 2.17: Results of MIMIC Analysis

Panel A - Contractibility MIMIC			Coefficient	Std. Error	Critical Ratio	Prob.
Goal clarity	>	Contractibility	1.000			
Measurability of outputs	>	Contractibility	0.892	0.370	2.413	0.02
Transformation process	>	Contractibility	0.431	0.212	2.027	0.04
Contractibility	>	Difficulty to assess work	0.238	0.058	4.129	0.00
Contractibility	>	Agreement output quality	0.184	0.046	4.027	0.00
Contractibility	>	Job performance	0.193	0.052	3.697	0.00
'a': variable based on directional expectation; significance calculated as one-tailed p-value						

Panel B - Goodness-of-Fit Statistics Contractibility MIMIC	
$Chi^2 = 6.064$	*goodness-of-fit statistics*
Degrees of freedom = 6	GFI = 0.994
Probability level = .416	AGFI = 0.978
	CFI = 1.000
Multivariate normality	RMR = 0.011
Kurtosis 2.046 (c.r. 1.862)	RMSEA = 0.006

To estimate contractibility, I calculate the factor score weights in AMOS. Table 2.18 gives these weights for predicting the unobserved variable from the (formative and reflective) observed variables, with which contractibility is calculated.

Table 2.18: Weights for Construction of Contractibility Measure

Goal clarity	.585
Measurability of outputs	.522
Knowledge of transformation process	.252
Difficulty to reach agreement output quality (reversed)	.807
Difficulty to assess work (reversed)	.862
Job performance	.319

Based on the estimated weights, contractibility is calculated using the formula presented below. Contractibility is added to the analyses in section 4 of this chapter.

Contractibility = 0.585 * goal clarity + 0.522 * measureability of outputs + 0.252 * knowledge of transformation process + 0.807 * difficulty to reach agreement output quality + 0.862 * difficulty to assess work + 0.319 * job performance

2.7 Analyses for For-Profit BUs (Appendix B)

This appendix presents robustness checks of the models presented in section 2.4.3.1 (page 42) and 2.4.3.2 (page 43). The robustness checks analyse the base models using a for-profit subsample of the data. Although I do not have *a priori* different theoretical expectations for RPE antecedents among not-for-profit business units, it is possible that the empirical results are qualitatively or quantitatively different for not-for-profit business units from for-profit sectors. To control for the influence of not-for-profit business units on my results, I estimate both the RPE-Use and the RPE-based-Targets models with for-profit business units only.

However, leaving out not-for-profit business units has one important drawback; it reduces the sample size. This is potentially harmful for the significance of the models, because the models contain interaction effects that require larger sample sizes. The testing of hypotheses using interaction effects with relatively small datasets can easily result in statistical null findings (see: Aguinis 1995:1142, Cohen *et al.* 2003:297). This risk is especially relevant in the case of the RPE-Use model, since the full-sample model (table 2.11 on page 42) already has limited explanatory power (ANOVA F-statistic = 1.718, p = 0.063, R^2 = 0.075, adjusted R^2 = 0.031). As the results of the for-profit analyses show, the RPE-Use model is insignificant. The ANOVA F-statistic of this model, presented in table 2.19 - panel A is 1.461 (p = 0.148). As argued above, this may be the result of the decreased sample size: the full-sample model has 267 included observations, the subsample analysis contains 224 observations. Due to the insignificance of the model, no inferences can be drawn from this analysis.

Unlike the RPE-Use model, the for-profit robustness analysis of the RPE-based-Targets model is significant (F-statistic = 1.884, p = 0.044). The results of this analysis are shown in table 2.19 - panel B. Similar to the RPE-Use model, the RPE-based-Target model has a weaker F and R^2-statistic than the full-sample analysis. Again, this may be the result of the reduced number of observations. Whereas the original full-sample RPE-based-Targets model (table 2.13 on page 45) contains 242 observations, the subsample analysis includes 200 observations. The full-sample analysis explained 12.2% of the variance in RPE-based-Targets (adjusted R^2 = 0.076). The current subsample analysis results in an R^2 of 0.099 (Adjusted R^2 = 0.047). Qualitatively, the subsample analysis yields the same results as the full-sample analysis. The results in table 2.19 - panel B support hypotheses H1 (the effect of common uncertainty on RPE use) and H3 (the interaction effect of uncertainty and information asymmetry on RPE use), but do not find siginficant support for the second hypothesis (the effect of the interaction between information asymmetry and comparability on RPE use).

Concluding, the for-profit subsample analyses provide additional support for the hypotheses, for as far as the models themselves are significant. The insignificance of the RPE-Use model is probably the result of reducing the samplesize of the dataset.

Table 2.19: Results of OLS Regression Analysis for For-Profit BUs

Panel A - RPE-Use For-Profit Subsample Analysis

	Coefficient	Std. Error	t-Stat.	Prob.
Constant	3.793	0.633	5.940	0.00
Common uncertainty	0.532	0.304	1.749	0.04[a]
Interaction information asymmetry * comparability	0.119	0.113	1.049	0.15[a]
Interaction uncertainty * information asymmetry	-0.105	0.130	-0.809	0.20[a]
Information asymmetry	0.015	0.076	0.194	0.85
Comparability	-1.652	1.108	-1.491	0.14
Contractibility	0.033	0.024	1.383	0.17
Firm-Level Measures	-1.125	0.377	-2.983	0.00
Size of BU	-0.006	0.035	-0.172	0.86
Size of firm	-0.012	0.031	-0.396	0.69
Dummy production BU	0.011	0.174	0.065	0.95
Dummy fin. services BU	0.072	0.172	0.417	0.68
$R^2 = 0.070$			F-statistic = 1.461	
Adjusted $R^2 = 0.022$			Prob(F-statistic) = 0.148	
Included observations: 224				

Panel B - RPE-based-Targets For-Profit Subsample Analysis

Constant	4.481	0.714	6.279	0.00
Common uncertainty	0.765	0.332	2.312	0.01[a]
Interaction information asymmetry * comparability	-0.017	0.128	-0.129	0.44[a]
Interaction uncertainty * information asymmetry	0.307	0.144	2.127	0.02[a]
Information asymmetry	-0.072	0.084	-0.856	0.39
Comparability	-2.712	1.220	-2.223	0.03
Contractibility	0.053	0.026	2.039	0.04
Firm-Level Measures	-0.129	0.467	-0.276	0.78
Size of BU	-0.008	0.041	-0.193	0.85
Size of firm	-0.008	0.034	-0.243	0.81
Dummy production BU	0.155	0.191	0.810	0.42
Dummy fin. services BU	-0.217	0.184	-1.179	0.24
$R^2 = 0.099$			F-statistic = 1.884	
Adjusted $R^2 = 0.047$			Prob(F-statistic) = 0.044	
Included observations: 200				

'[a]': variable based on directional hypothesis
significance calculated as one-tailed p-value

Chapter 3

Does Relative Performance Evaluation Reduce Noise in the Performance Evaluation?

Preamble to Chapter 3

The current chapter addresses the claim that RPE reduces the room that managers have to behave opportunistically. By doing so, this study answers the first part of research question 3: *is RPE effective at reducing noise and opportunism?*

The noise-reduction claim is also assessed in chapter 2. Chapter 2 refers to the noise-reduction perspective to derive the effect of common uncertainty, a driver of noise, on RPE and finds support for this relation. However, the relation between these two variables may still be caused by mechanisms other than the reduction of noise in the performance evaluation. Prior literature has suggested several alternative perspectives on RPE use that could explain the findings in chapter 2.

The current chapter analyses the noise-reduction perspective in a more direct manner by seeking additional proof for the noise-reduction explanation of RPE use. Instead of analysing whether RPE use is increased by noise-inducing factors, this chapter studies whether RPE effectively reduces noise in the performance evaluation. Further study of the noise-reduction perspective is important for our understanding of RPE's consequences for Management Control. Together with chapter 2, the current chapter adds to our understanding of whether RPE use can be explained by an organizational wish to insulate the performance evaluation from the impact of external events, and to improve contractual efficiency between the manager and his superior.

3.1 Introduction

This chapter empirically assesses one of the fundamental claims of Relative Performance Evaluation theory: the causal relation between RPE use and the reduction of noise in the performance evaluation of employees. Noise decreases the performance metrics' informativeness about the efforts of the employee. Noise occurs if uncontrollable events affect the measured performance. Noisy performance measures do not purely reflect the efforts of the employee, as they are also affected by windfall gains and losses caused by uncontrollable events. For example, consider the situation of a sales-manager operating in a stagnating market. The uncontrollable market conditions reduce the sales manager's revenues. Although he zealously attempts to boost his sales, the manager's performance indicators suggest that he provided much less effort than last year, when the market conditions were favourable. In this situation external events (*in casu*: the stagnating market) reduce the manager's performance score. However, the opposite effect is also possible when the conditions are highly favourable. In both scenarios, noise results in a less functional performance measurement system, as it informs the principal less accurately with regard to the agent's efforts. This reduces the efficiency of the performance contract between the principal and the agent by imposing risk on the agent of not being rewarded for effort but for luck.

According to agency reasoning, RPE can restore the quality of the performance evaluation by filtering out noise from the evaluated manager's measured performance (Holmstrom 1982, Gibbons & Murphy 1990). RPE reduces noise by incorporating information about the attainable performance level into the evaluated manager's performance evaluation. This works because of peer comparison (Holmstrom 1982). RPE compares the manager's actual performance to the performance of a reference group that partially faces the same external events (*i.e.*, market stagnation and dropping revenues). Thus, the performance of the reference group is impacted in a similar manner by shared external events. The performance of the reference group in a given period, with its specific external events, informs the principal about the performance level that should be attainable for the agent in that period. The comparison with peer performance constructs a performance norm that is (at least partially) insulated from the impact of external events.

This causal relation between RPE use and noise in the performance evaluation is a cornerstone explanation for why organizations would use RPE. This noise-reduction explanation is prevalent in two streams of literature. The first stream is the analytical agency literature. In a seminal paper, Holmstrom (1982) shows that RPE reduces noise in the performance evaluation if benchmarked agents face some common uncertainties.

According to Holmstrom's analyses, RPE filters out the effects of common uncertainties such that the evaluation yields less noisy information about the agent's performance.

Holmstrom's conclusions concerning the noise-reducing properties of RPE are supported by other analytical research, such as the work of Gibbons & Murphy (1990).

The second stream of literature that explains RPE use from a noise reduction standpoint, is the empirical RPE literature. The empirical RPE literature focuses on why and to what extent organizations use RPE. Although the empirical support for RPE is limited in general, as argued in the previous chapters of this thesis, several studies confirm the prevalence of RPE based on noise-reduction arguments. Examples of these papers include Matsumara & Shin (2006), Albuquerque (2009), Gong *et al.* (2011), and chapter 2 of this thesis. However, whereas the analytical literature explicitly studies the extent to which RPE reduces noise, the empirical stream does not test noise reduction *directly*. Rather, the empirical literature *assumes* that RPE reduces noise and uses this assumption as an explanation for RPE use in practice. This assumption leaves the noise-reduction claim concerning the usefulness of RPE largely untested from an empirical standpoint.

However, to further examine a theory, one must assess its untested assumptions in as direct a manner as possible. Testing a theory's assumptions furthers our understanding about the quality of the applied theoretical perspective. For example, it is possible that, instead of the applied theoretical explanation, an alternative mechanism drove prior results. Specifically, based on the findings in chapter 2, I claim support for the noise-reduction explanation of RPE use. This conclusion is based on analyses of how an uncertainty-measure (derived from the noise-reduction perspective) increases RPE use. The empirical confirmation of this relationship, provided in the same chapter, supports the underlying noise-reduction theory. However, despite the theoretical grounding in prior studies and the empirical confirmation of the noise-reduction theory, it remains entirely possible that the relation between uncertainty and RPE use is, in fact, not due to noise reduction, but due to an alternative theoretical explanation.

Prior studies have suggested several alternative explanations for RPE use that may drive the results in chapter 2. For example, RPE may be used as a means to pit employees or organizational parts against each other to create a sense of internal rivalry that increases the efficiency of the individual agents (Holmstrom 1982). A second example suggested by Gibbons & Murphy (1990:33-S) is that RPE is used to facilitate organizational learning (*e.g.*, in the form of benchmarking the outputs and processes). Both of the presented applications can be especially useful in a highly uncertain environment, where environmental turbulence demands that firms perform optimal to survive.

An analysis of the relation between uncertainty and RPE use (as conducted in chapter 2 of this thesis) does not help to distinguish between the noise-reduction explanation and the two presented alternative explanations. Additionally, because the latter two explanations

would be confirmed in this type of analysis, this analysis falsely suggests support for the noise-reduction claim of RPE-theory. To test the noise-reduction assumption, I require a more direct test.

To test the noise-reduction assumption of RPE-theory in a more direct manner, the current research analyses noise-reduction effectiveness. This analysis assesses whether organizational reliance on RPE actually reduces the noise levels in the performance evaluation of employees.

For the empirical analysis, I use survey data of 325 business unit managers in the Netherlands. Overall, the results are consistent with the findings of the analytical agency literature and the conclusions of chapter 2. The analyses support the main claim of RPE theory: namely RPE reduces the amount of noise in an agent's performance evaluation.

This chapter is organized into five sections. After the introductory section, section 3.2 provides the development of the hypotheses and presents the research model. The measurement is described in section 3.3, and the analyses and results are described in section 3.4. Section 3.5 addresses the findings and the conclusions.

3.2 Development of the Model

This section describes the development of the research model, which consists of the research hypothesis and additional expectations. Firstly, the research hypothesis is presented in subsection 3.2.1. This hypothesis addresses the assumed noise-reduction properties of RPE. Subsection 3.2.2 presents a set of expectations consisting of noise-inducing and noise-reducing factors. These factors are included to address the measurement problems of the model's dependent variable rather than because of their theoretical relevance to the noise-reduction assumption.

The dependent variable *noise in the performance evaluation* is measured with a perceptual measure that refers to how respondents perceive their noise levels under their currently installed control systems. However, the currently perceived noise level includes the effects of potential noise-reducing control instruments, such as RPE. If I aim to estimate the effect of RPE on noise, it is necessary to predict a noise level 'theoretically' before it is reduced by RPE. To estimate the effect of RPE on noise, I extend the model with factors that affect the noise level outside the influence of RPE. These factors include environmental uncertainty, goal ambiguity, measurability of outputs, and the extent to which the performance measurement system relies on personal level and disaggregate measures. In section 3.2.2, I discuss how these factors (positively or negatively) affect the noise level that a manager experiences in his performance evaluation.

3.2.1 Research Hypothesis: RPE use and Noise in the Performance Evaluation

The research hypothesis concerns the negative effect of RPE use on the level of noise in the performance evaluation. As discussed in the introductory section of this chapter, RPE use reduces the amount of noise in the performance evaluation.

Noise decreases the informativeness of the performance metrics with regard to the agent's effort level. Because of noise, the measured performance of the business unit manager is not only the result of his effort but also the result of other factors that lie beyond the manager's control. However, according to RPE theory, noise in the performance evaluation can be reduced by using information about the performance of the manager's peers. To at least some degree, the agent's peers are exposed to the same factors that affect the agent's performance. These 'common factors' can for example negatively influence the performance of the agent and his peers in the same period. RPE filters out the noise caused by common factors by comparing the measured performance of the evaluated manager with the performance of the peer group. RPE does not adjust the score on the performance measure itself, but provides information about peer performance that aids the interpretation of the score on the performance measure. In other words, with RPE, the measured performance itself remains noisy, but the performance evaluation is insulated from this noise. Because of this insulation, the performance evaluation provides a more accurate description of the evaluated manager's effort level. This increased accuracy reduces the risk of the manager not being rewarded for his effort and reduces inefficient risk-sharing between the organization and the manager. The noise-reducing effect of RPE is formalized as follows:

H1: RPE use has a negative influence on the amount of noise in the performance evaluation.

3.2.2 Noise-Inducing and Noise-Reducing Factors

In addition to the hypothesis of theoretical interest presented above, the model contains additional expectations that address the measurement characteristics of the model's dependent variable, as explained in the introduction of this section. The additional expectations consist of noise-inducing and noise-reducing factors that constitute the noise that would exist in the performance evaluation, if RPE did not reduce this noise. In the current section, I argue that noise in the performance evaluation is affected by environmental uncertainty, goal ambiguity, measurability of outputs, emphasis on personal-level measures, and emphasis on disaggregated performance measures. The effect of the first of these noise-drivers (environmental uncertainty) is partially mitigated by RPE use. Before discussing the other factors, I first discuss the effects of environmental uncertainty on RPE use and on noise.

Environmental Uncertainty *Environmental uncertainty* has two effects in my model. Environmental uncertainty induces noise in the performance evaluation and drives the use of RPE to mitigate its noise-inducing effect. Environmental uncertainty[1] is the primary driver of noise in this study. In the discussion of the research hypothesis, I argue that noise is due to factors that lie beyond the manager's control. Environmental uncertainty refers to the prevalence of uncertain events that originate from outside the business unit. Almost by definition, these events lie outside the manager's control[2]. These external events affect the manager's performance metrics and thereby introduce noise into the performance evaluation. In the previous section, I argued that RPE can reduce noise by partially filtering out these external events. This reduction increases the attractiveness of RPE use in a situation characterised by high environmental uncertainty. If the amount of environmental uncertainty increases, the performance evaluation increasingly benefits from the deployment of RPE. This finding implies that environmental uncertainty drives the use of RPE. This expectation is formalized as follows: *environmental uncertainty has a positive effect on RPE use.*

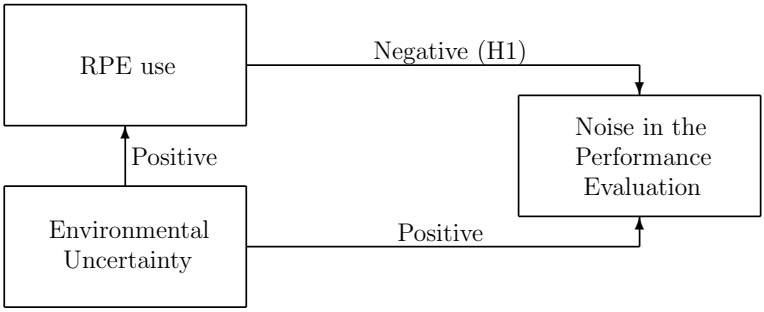
Although I argue that RPE can reduce the noise-driving effect of environmental uncertainty, this argument only applies to a specific type of environmental uncertainty: environmental uncertainty that is shared with one's peers. Only shared (or common) environmental uncertainty can be filtered out by comparing the performance with the performance of peers. Other idiosyncratic environmental uncertainties do not affect peer performance and cannot be filtered out from the performance evaluation by using RPE. In summary: although RPE partially filters out the effects of environmental uncertainty from the performance evaluation, a remaining idiosyncratic part of environmental uncertainty can still induce noise. This expectation is explicated as follows: *environmental uncertainty has a positive effect on noise in the performance evaluation.*

Figure 3.1 summarizes the relations among RPE use, environmental uncertainty, and noise in the performance evaluation.

[1] Whereas the previous chapter assesses how *common* uncertainty increases RPE use, the current chapter relies on a broader uncertainty measure that includes both common and idiosyncratic uncertainty. This broader measure is used to improve the noise estimation in the model as both common and idiosyncratic uncertainty induce noise in the performance evaluation. The noise-increasing effect is not limited to common uncertainty per se, although the noise-reducing effect of RPE is.

[2] It is possible that a business unit manager can influence his environmental uncertainty to some extent by altering his environment. For example, the manager may be able to choose to operate in a low-risk market, whilst avoiding operating in a more volatile, high-risk market. However, if the manager is part of a larger organization with an overarching company strategy, his influence may be limited. Therefore, this study assumes that the environment (and its uncertainty) is given for the business unit manager.

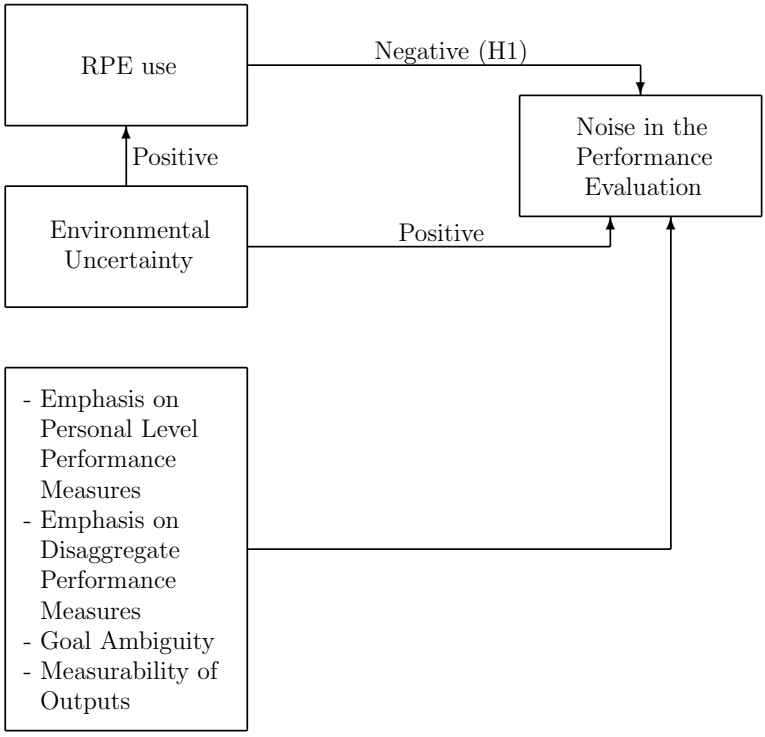
Figure 3.1: Conceptual Model



Emphasis on Personal-Level Measures In addition to environmental uncertainty, this study also includes other drivers of noise. The second driver of noise that is discussed in this chapter refers to the *emphasis on personal-level measures*. Performance information can be obtained with measures at various organizational levels, which range from personal-level performance indicators and business-unit-level measures to firm-level performance metrics. The organizational level at which the performance metrics are obtained, impacts the amount of noise in the measure. For example, performance information that is obtained at organizational levels higher than the business unit (*i.e.*, firm-level measures) include more than the organizational parts over which the business unit manager has control and are noisy per se, as the performance measurement includes factors beyond the control of the evaluated manager. In contrast, performance information at the personal level contains little to no noise, as it contains information about the manager's own actions and abilities. This information includes the manager's personal development of his professional and social skills, and experience and knowhow. The increased importance of personal-level performance measures in the Performance Measurement System as a whole reduces the room for noisy performance measures. This argument is expressed in the following expectation: *the emphasis on personal-level performance measures has a negative effect on noise in the performance evaluation.*

Emphasis on Disaggregated Performance Measures In addition to the organizational level of the performance metrics, the level of aggregation of the performance measures affects noise. Bouwens & Van Lent (2007) argue that a primary function of disaggregated measures is to reduce the noise in aggregated measures (*e.g.*, accounting returns and profits). Disaggregated measures, including cost, revenue, or cash flow measures, are less subject to exogenous events because disaggregated measures include fewer performance areas that can be affected by external factors. Therefore disaggregated measures contain less noise (Bouwens & Van Lent 2007). As a result, I form the following expectation: *the emphasis on disaggregated performance measures has a negative effect on noise in the performance evaluation.*

Figure 3.2: Causal Model



Goal Ambiguity & Measurability of Outputs The fourth and fifth expectations relate to the impacts of goal ambiguity and measurability of outputs on the noise in the performance evaluation. Both goal ambiguity and measurability of outputs are likely to influence the perceived noise level, because respondents may experience distortion (*i.e.*, the misalignment between the desired organizational results and the measured and achieved performance) as noise. If the organization cannot communicate or measure its desired results, then it can become unclear to the evaluated manager to what standards his achieved results are being compared. This ambiguity can increase the level of noise in the performance evaluation because if the goals and targets are unclear, then there is room for other factors to influence the manager's performance evaluation. To include the influence of unclear goals, or poorly measurable outputs, this study formalizes its effects as follows: *goal ambiguity has a positive effect on the level of noise in the performance evaluation* and *measurability of outputs has a negative effect on the level of noise in the performance evaluation.*

The conceptual model, which includes the research hypothesis and the associated expectations, is summarized in figure 3.2.

3.3 Sample and Measurement

The current section contains parts that are largely similar to parts of chapter 2. This similarity is due to the use of the same survey instrument and sample. Subsections 3.3.1 and 3.3.2 are marked with an asterisk to indicate this repetition. The reader may want to skip these parts.

3.3.1 Sample(*)

This study uses primary data collected from 325 business units with an extensive survey instrument. The questionnaire was filled out by business unit managers contacted through student's professional network. This approach has several benefits. The first benefit is respondent identification. The students were instructed to identify managers, who are responsible for an autonomous organizational body engaged in multiple activities (*e.g.*, purchasing, production, and sales). Furthermore, the respondent should have a superior within the organization (in other words: the respondent should not be the CEO). The second benefit is the possibility to conduct face-to-face interviews, which improves the valid interpretation of and careful responding to the questions by the respondent. During these interviews, the student can explain the survey questions (using a strict protocol provided by the researcher) to the business unit manager who answers the questions in the survey and fills out the questionnaire.

Table 3.1: Summary Statistics on Sample

Panel A - size

		N	Mean	Std. dev. of mean	Min.	Max.
BU size	FTE	325	220	439	10	5.000
	Revenue (mio EURO)	292	196	840	1	1.000
Firm size	FTE	321	14.730	41.771	31	470.123
	Revenue (mio EURO)	301	7.775	43.240	4	600.000

Panel B - sector

BU sector		
	Production	54
	Services	181
	Financial	52
	Not-for-profit	57

Firm sector		
	Production	71
	Services	170
	Financial	62
	Not-for-profit	64

The observations include data from production business units (17%), service BUs (56%), financial services (16%), and BUs operating in the not-for-profit sector (18%)[3]. Similarly diverse responses to the variance of sectors at the BU-level were provided on the question about the respondents firm's sector. Table 3.1 summarizes the background information on the respondents' business units and the firm to which each business unit belongs.

3.3.2 The Questionnaire(*)

The survey instrument was specifically designed for this research following the Tailored Design Method (Dillman 2000) where applicable to the specific research setting[4]. The survey instrument relies heavily on previously used and validated questions. Some of those measures are well established (*e.g.*, the measure for information asymmetry is largely based on Dunk (1993)). Other questions were specifically constructed for this survey. For example, the measures for the dependent variable -that measures the use of RPE- are purpose-developed.

To increase the reliability of the instrument, the survey instrument was thoroughly pre-tested using Hak's Three-Step Test Interview method (abbreviated: TSTI) (Hak *et al.* 2008). This method is an observational instrument for pretesting self-completion questionnaires[5]. TSTI works through observing actual instances of interaction between respondents and the instrument. Because this process mainly consists of cognitive processing and is therefore hidden from the observer, 'thinking aloud' is used as a technique for making the thought process observable (Hak *et al.* 2008). Through this process, TSTI helps to identify problems in questionnaires, which often leads to modification of the instrument. Three rounds of pre-tests with twelve managers[6] ensured a thorough and correct understanding of the questions and response categories.

[3]These percentages sum to 107% because some business units operate in more than one sector. Multiple answers to the related question are considered valid.

[4]The Tailored Design Method is specifically designed for mail and internet surveys. Many of the guidelines provided by Dillman (2000) are applicable to the survey design, but others specifically apply to mailed surveys instead of face-to-face survey interviews.

[5]Although the survey was not a self-completion questionnaire in the sense that the respondents did not fill out the survey by themselves, but together with the interviewing student, I still value the contribution of TSTI. Filling out the survey was outside of the direct sphere of influence of the researcher, but was carried out by -carefully briefed- students.

[6]The pre-test participants' organizations show great diversity, ranging from production to consulting, educational and financial services, and from very large to smaller business units, including both for-profit and not-for-profit organizations.

3.3.3 Measurement

This section describes the measurement of the variables used for this study. I discuss the variables related to the research hypothesis (*i.e.*, noise in the performance evaluation, and RPE-Use), as well as the noise drivers (*i.e.*, environmental uncertainty, goal ambiguity, measurability of outputs, emphasis on personal-level measures, and emphasis on disaggregated measures). Additionally, this section motivates and describes the control variables for size and sector. When references are made to the questionnaire items, please see the appendix at the end of the thesis for the exact phrasing, response categories, and, if applicable, sources.

Noise in the Performance Evaluation The dependent variable noise in the performance evaluation is based on a single question that asks the respondent to indicate the extent to which the performance evaluation that he receives is based on factors over which he has control (Q34). Throughout this chapter, this proxy is referred to as 'NOISE'. NOISE is measured by taking the reverse of the respondents' answers. In other words, NOISE measures the extent to which the performance evaluation is influenced by factors that lie beyond the business unit manager's control. I acknowledge that relying on a single indicator construct is less than ideal, especially when measuring the dependent variable in the model. Therefore, I seek additional validation for NOISE. The questionnaire provides additional items that support this metric. Specifically, the questionnaire provides two items that are associated with the variable of interest, though they do not measure noise in the performance evaluation per se. For one, I asked the respondents to indicate the extent to which the business unit's performance is substantially influenced by factors beyond his control (labelled: exogenous-performance-effects, Q35). Similar to NOISE, this question refers to the impact of exogenous factors. However, this time on the difficulty of the business unit performance instead of on the performance evaluation. Several reasons exist why NOISE and exogenous-performance-effects are not perfectly correlated. Firstly, the performance evaluation can be informed by things other than the business unit's performance (*e.g.*, contributions to firm performance, or personal development of skills). If so, the exogenous events that impact the business unit's performance, impact the performance evaluation only to a certain degree. Second, the performance evaluation can be insulated from exogenous events that influence business unit's performance through, for example, RPE. As argued before, RPE filters noise out of the performance evaluation. However, RPE does not remove the impact of exogenous events on the BU performance itself. Therefore, I expect NOISE and exogenous-performance-effects to correlate positively with one another, though not to a great extent.

The second questionnaire item to validate NOISE captures more broadly and informally whether the respondent is held accountable for uncontrollable factors, instead of whether

such factors specifically affect his (formal) performance evaluation. This item is labelled held-accountable (Q36). Specifically, held-accountable asks whether the respondent is frequently held accountable for certain (negative) results beyond his control. Because of its phrasing, this item focuses on the negative impact of external factors, whereas the primary measure NOISE measures both positive and negative noise. Although these two additional questionnaire items (exogenous-performance-effects and held-accountable) do not proxy noise in the performance evaluation per se, they can inform us about the quality of the NOISE measure. The correlations between NOISE and the two associated items are positive and significant ($p < 0.01$), but not high, as shown in table 3.2. The convergent validity of the NOISE measure is also analysed through a factor analysis of the three measures. Although the scale reliability does not suggest constructing a multi-item scale to proxy for noise in the performance evaluation (Cronbach's alpha < 0.6), factor analysis shows that the three measures can be combined into one measure by yielding a one-component solution (component loadings > 0.5). Both the correlations and factor analysis provide validation for the NOISE measure.

Table 3.2: Correlations between Noise-Measures

	1.	2.
1. NOISE		
2. Exogenous-Performance-Effects	.154***	
3. Held-Accountable	.334***	.165***

*** Correlation is significant at the 0.01 level (2-tailed).
Listwise N=325

RPE use The RPE use measure (labelled 'RPE-Use') asks about the extent to which peer performance functions as a point of reference for evaluating the quality of the agent's performance (Q1). The underlying questionnaire items focus on the *ex-post* nature of performance evaluation. This measure comprises both explicit coupling of the performance target to peer performance and more implicit applications of RPE. Implicit applications of RPE do not require that peer performance affects the performance target explicitly, but they can, for example, play a role in establishing an implicit performance standard or norm for the performance evaluation. This measure asks to what extent the respondent perceives that the performance of his peers is a point of reference for his superior when evaluating the respondent's performance. Additionally, the question differentiates between situations where the evaluated business unit performs substantially better versus substantially worse than its peers. This dinstinction controls for potential asymmetries in peer comparison. Throughout the sample, the asymmetries do not seem to be of great influence; the three settings (neutral/better/worse) lead to an internally consistent scale, as presented in table 3.3.

Table 3.3: Items for RPE-Use (Q1)

	Item description	Component loadings
a.	Peer performance point of reference	0.868
b.	Substantially better	0.923
c.	Substantially worse	0.902
	Percentage variance explained	80.6%
	Cronbach's alpha	0.879

Table 3.4: Items for Environmental Uncertainty (Q4-7)

	Item description	Component loadings (1)	Component loadings (2)
a.	External factors influencing business unit performance	0.544	0.472
b.	Unpredictable variation in the amount of work	0.667	-0.215
c.	Number of exceptions that arise in the unit	0.689	-0.496
d.	Differences in day-to-day situations	0.682	-0.423
e.	Need to react to outside pressure	0.672	0.326
f.	Long-range planning hindered by unpredictable events	0.602	0.495
	Percentage variance explained	41.6%	17.4%
	Cronbach's alpha	0.712	

Environmental Uncertainty Environmental uncertainty was reflectively measured using a multi-item construct, combining six items from earlier studies. These items stem from Van der Ven & Ferry (1980), Kruis (2008) and Kalleberg *et al.* (1996), where they have been applied to measure the impact of external factors, unpredictable events and difficulties in forecasting due to external factors. More specifically, I ask for the frequency of which external events occur that substantially influence the performance of the BU or the amount of work, the occurrence of exceptions that require different methods or procedures, and how often the same day-to-day situations/problems arise.

As is shown in table 3.4, principal component analysis yields two factors. However, the first factor has all component loadings higher than 0.500, and scale reliability analysis produces only one reliable scale with a high Cronbach's alpha. Therefore, I measure environmental uncertainty with one multi-item construct. The measure for environmental uncertainty is calculated as the mean of the six items[7].

[7]Before calculating the mean score, 7-point scales are recoded to 5-point Likert scales to prevent over-

Goal Ambiguity Goal ambiguity measures the extent to which firms clearly formulate their objectives. Goal ambiguity is measured with a multi-item construct, evaluated by four questionnaire items. As summarized in table 3.5, the items refer to how clear and specific the goals are defined, and how easily they can be communicated. The Cronbach's alpha of the four items together is rather low at a value of 0.630. The Cronbach's alpha cannot be improved by deletion of one of the items, as is shown in table 3.5. However, factor analysis shows that the four questions load onto one component, with all component loadings well above 0.500. Based on the adequate component loadings, I combine the items into one measure, labelled goal ambiguity.

Table 3.5: Items for Goal Ambiguity (Q17-20)

	Item description	Component loadings	Scale if item deleted
a.	Clarity of goals	0.758	0.484
b.	Specificity of goals	0.734	0.480
c.	Difficulty of explaining goals to outsiders (reversed)	0.572	0.552
d.	Clarity of goals to insiders	0.682	0.547
	Percentage variance explained	47.6%	
	Cronbach's alpha	0.630	

Measurability of Outputs Measurability of outputs refers to the extent to which firms are able to measure the desired results of the business unit in a goal consistent way. Measurability of outputs is a construct evaluated by four questionnaire items, asking about the objective measurability of the outputs, including aspects of completeness and goal consistency of the output measurement. The items load onto one factor with a scale reliability of 0.74 (Cronbach's alpha). The items and the corresponding statistics as summarized in table 3.6.

Emphasis on Personal-Level Performance Measures I measure *emphasis on personal-level measures* as the importance of personal-level performance measures to the respondent's performance evaluation (Q37). The respondents divided 100 points over a list of performance measures according to their relative weights in the performance evaluation system. The list of performance measures consists of measures at the firm, business unit, within-business-unit, and personal levels. I summate the relative weights of the various personal-level measures (professional skills, social skills, and experience & know-how), to calculate the emphasis on personal-level measures-construct.

representation of the 7-point scaled items in the uncertainty construct.

Table 3.6: Items for Measurability of Outputs (Q21-23)

	Item description	Component loadings
a.	Match between performance measures and business unit goals	0.731
b.	Objective measurability of goals	0.724
c.	Match between performance measures system and results	0.809
d.	Goals consistency performance measurement system	0.760
	Percentage variance explained	57.3%
	Cronbach's alpha	0.744

Emphasis on Disaggregated Performance Measures Similar to the measurement of the emphasis on personal-level performance measures, I measure the emphasis on disaggregated measures as their relative weights within the total set of performance measures (Q38). The importance of the individual disaggregated measures are measured at the business unit and the within-business-unit levels of the organization, and consist of costs, revenues, cash flow and non-financial disaggregated measures including market share, customer satisfaction, and quality measures. The weights of these measures are summated and jointly form the measure emphasis on disaggregated measures.

Control Variables This study controls for business unit and firm size, as well as for sector effects. This paragraph presents the measures for these control variables.

Size (BU and Firm) Size is often regarded as a potentially important determinant of performance measurement practices (Speklé & Verbeeten 2008). For example, larger organizations might be more effective in the use and implementation of sophisticated performance measures that are less vulnerable to noise. Control variables *business unit size* (Q31) and *firm size* (Q32) are both based on the number of FTE and revenues. For both control variables, the natural logs of FTE and revenue were calculated. With business unit size measured as FTE and revenue, the Cronbach's alpha is rather low. However, the two items correlate positively ($r = 0.50$, $p < 0.01$). Additionally, the two factors load onto one component in the factor analysis (component loading = 0.866). Despite the low Cronbach's alpha, both items are combined into the measure business unit size. Although business unit size and firm size are significantly correlated with one another (Pearson correlation 0.479), they do not yield collinearity-issues in the multivariate analyses. Table 3.7 summarizes the questionnaire items and statistics on the size variables.

Table 3.7: Items for Business Unit Size & Firm Size (Q31-32)

	Item description	Component loadings
a.	Business unit FTE (Ln)	0.866
b.	Business unit revenue (Ln)	0.866
	Percentage variance explained	75.0%
	Cronbach's alpha	0.634
c.	Firm FTE (Ln)	0.952
d.	Firm revenue (Ln)	0.952
	Percentage variance explained	90.6%
	Cronbach's alpha	0.892

Sector Dummy Variables To control for potential industry effects, dummy variables were included that distinguish between business unit sectors (Q33). The model contains dummies for production, financial services, and not-for-profit business units. The base model consists of service organizations, representing the largest sector in the sample.

3.3.4 Descriptive Statistics

Table 3.8 presents the descriptive statistics for all of the metrics variables described in the preceding section. According to the skewness and kurtosis indicators, all of the variables follow a normal distribution.

Table 3.8: Descriptive Statistics

	Mean	Min	Max	Std. Dev.	Skew-ness	Kurt-osis
NOISE	3.24	1.00	6.00	1.16	0.33	-0.31
RPE-Use	3.37	1.00	5.00	1.00	-0.85	0.31
Environmental Uncertainty	3.19	1.57	4.76	0. 60	-0.38	0.07
Goal Ambiguity	2.34	1.00	5.00	0.57	0.43	1.19
Measurability	4.48	2.00	6.00	0.69	-0.45	0.38
Emphasis on Personal-Level Measures	0.30	0.00	1.00	0.22	0.79	0.47
Emphasis on Disaggregated Measures	0.31	0.00	0.90	0.18	0.50	0.13
Business Unit Size	10.27	2.30	14.97	2.26	-1.56	2.70
Firm Size	13.41	3.69	19.41	2.43	-0.44	1.34

3.4 Analyses

This section presents the analyses. I analyse the noise-reducing effect of RPE with structural equation modelling (SEM). I use the structural model estimation of SEM to be able to estimate the indirect relation between environmental uncertainty and noise in the performance evaluation, via RPE use.

In addition to the aforementioned noise-reducing effect of RPE use and the effect of environmental uncertainty, via RPE use, on noise, I include the noise-drivers and control variables for size and industry. The result is the statistical model shown in figure 3.3.

3.4.1 Correlations

Table 3.9 presents the correlation matrix of all of the included variables. The table shows three significant correlations between the dependent and independent variables. Noise is negatively correlated with RPE-Use. Additionally, the correlations among noise and goal ambiguity, and noise and measurability of outputs are significant. The latter two correlations are consistent with the hypothesized directions. Furthermore, several significant correlations among the independent variables exist. RPE-Use is positively correlated with environmental uncertainty, and measurability of outputs. Additionally, I find significant correlations between the measures of goal ambiguity and measurability of outputs, as well as between the measures for business unit size and firm size. However, these correlations are not strong.

3.4.2 Multivariate Analyses

To test the noise-reducing effect of RPE, I analyse the model that is presented in figure 3.3. In addition to the expected paths, as presented in the causal model (figure 3.2 on page 66), I also include the control variables for size and business unit sector in the model. I use maximum likelihood estimation in AMOS 18 to estimate the structural model. Due to sample size limitations and to reduce the model complexity, I do not estimate the measurement model and the structural model simultaneously. The measurement is determined outside the model, using principal component analysis and scale reliability analysis, as presented in section 3.3.3.

I run two analyses. First, the model is tested with the RPE-Use measure. This model is referred to as the base model. Second, an alternative measure for RPE use is used as a robustness check. Both models fit the data and support the noise-reduction hypothesis (*i.e.*, RPE use reduces the amount of noise in the performance evaluation). The results of these analyses are discussed in section 3.4.2.1. 'Analyses with RPE-Use Measure' and section 3.4.2.2. 'Analyses with RPE-based-Targets Measure'.

Figure 3.3: Statistical Model

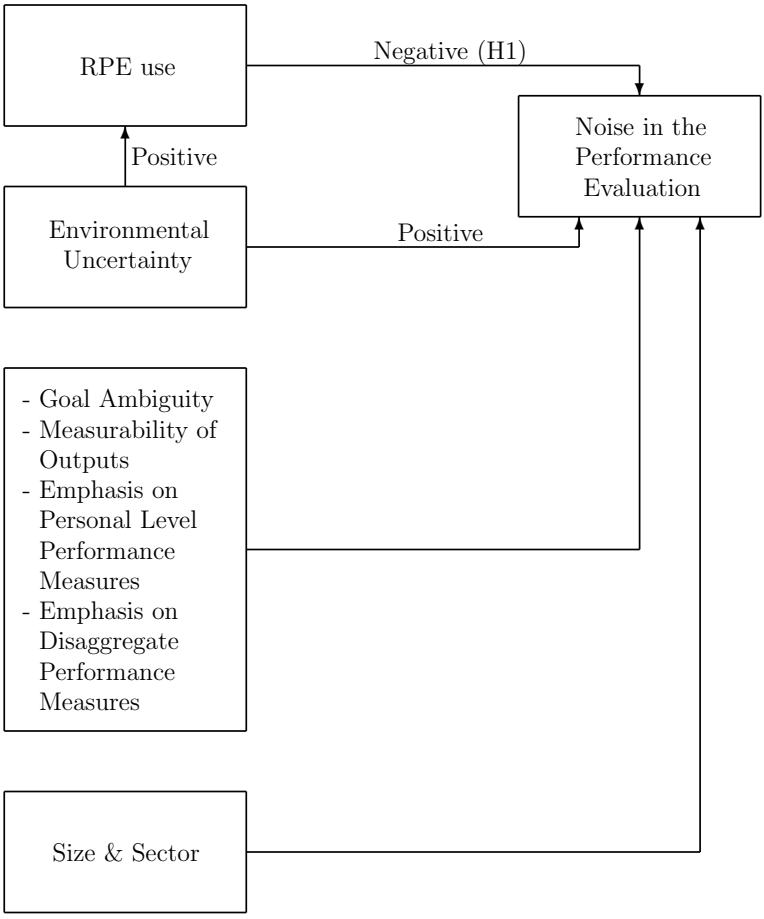


Table 3.9: Pearson Correlation Table

	1.	2.	3.	4.	5.	6.	7.	8.
1. NOISE								
2. RPE-Use	-.147**							
3. Envir. Unc.	-.093	.183***						
4. Goal Ambiguity	.264***	-.039	-.068					
5. Meas. of Outputs	-.212***	.108*	.080	-.559***				
6. Emphasis on PLM	-.006	.083	.034	.065	-.167***			
7. Emphasis on DM	-.085	-.086	-.041	.095	-.111	-.328***		
8. Business Unit Size	.027	.024	.045	.029	.075	-.132**	-.002	
9. Firm Size	-.010	.024	-.050	-.082	.221***	-.017	.010	.485***

*** Correlation is significant at the 0.01 level (2-tailed).
** Correlation is significant at the 0.05 level (2-tailed).
* Correlation is significant at the 0.10 level (2-tailed).
Listwise N=244

3.4.2.1 Analyses with RPE-Use Measure

The first model to be analysed is the base model. These analyses rely on the primary operationalization of RPE use for the evaluation of business unit managers ('RPE-Use') to test the noise-reduction properties of RPE. The base model fits the data and explains 11% of the variance in the dependent variable noise in the performance evaluation. Before presenting the model fit statistics, I discuss the qualitative results.

The base model supports the noise-reduction hypothesis, which is formulated in hypothesis H1. The analyses also support the second expectation; environmental uncertainty positively influences the reliance on RPE.

The model does not show a correlation between environmental uncertainty and noise in the performance evaluation. This null result suggests that the idiosyncratic part of environmental uncertainty does not have a significant effect on noise in the performance evaluation. This finding implies that on average, RPE insulates the performance evaluation from practically all environmental events.

Concerning the noise-drivers, the model shows support for all of the expectations. Only the control variables for size and sector do not influence noise in the performance evaluation. The results are summarized in table 3.10 - panel A.

The fit of the SEM model is assessed with a number of goodness-of-fit statistics. To assess the fit of a SEM, one must assess multiple measures (Kline 2005). As the indicators of model fit, I use the chi^2, the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the comparative fit index (CFI), the root mean square residual (RMR) and the root mean square error of approximation (RMSEA). Kline (2005) suggests that the following indicate a good fit: an insignificant chi^2; a GFI, AGFI, and CFI greater than 0.90; a RMR close to 0; and an RMSEA less than 0.045. As presented in table 3.10 - panel B, the indicators show good model fit.

Despite the good model fit statistics, the model shows potential non-normality issues. The multivariate kurtosis measure lies above the tolerable level. Kline (2005) argues that a kurtosis greater than ten may suggest a problem with the data. The kurtosis in the current model is 19.114. Because multivariate normality is a strict assumption of maximum likelihood estimation (MLE) in SEM models, I use bootstrapping with 500 samples, which does not assume multivariate normality, to analyse whether the kurtosis affects the MLE estimates (Kline 2005, Bryne 2001). I find that the results are qualitatively similar, and that normality issues do not drive the results.

Table 3.10: Results of SEM Analysis (RPE-Use)

Panel A - base model results			Coefficient	Std. Error	Critical Ratio	Prob.
RPE use	>	NOISE	-0.132	0.064	-2.300	0.01'a'
Envir. unc.	>	NOISE	-0.007	0.111	-0.117	0.45'a'
Envir. unc.	>	RPE use	0.165	0.102	2.784	0.00'a'
Goal Ambiguity	>	NOISE	0.190	0.141	2.824	0.00'a'
Meas. of Outputs	>	NOISE	-0.140	0.123	-1.951	0.03'a'
Emph. on PLM	>	NOISE	-0.101	0.327	-1.587	0.06'a'
Emph. on DM	>	NOISE	-0.110	0.384	-1.729	0.04'a'
BU size	>	NOISE	0.011	0.034	0.160	0.87
Firm size	>	NOISE	0.050	0.031	0.722	0.47
Dummy prod.	>	NOISE	-0.021	0.175	-0.344	0.73
Dummy fin. serv.	>	NOISE	0.032	0.188	0.444	0.66
Dummy N.F.P.	>	NOISE	0.028	0.182	0.539	0.59
'a': variable based on directional expectation; significance calculated as one-tailed p-value						

Panel B - base model fit statistics	
$Chi^2 = 6.262$	*goodness-of-fit statistics*
Degrees of freedom = 9	GFI = 0.996
Probability level = .713	AGFI = 0.968
	CFI = 1.000
Multivariate normality	RMR = 0.015
Kurtosis 19.114 (c.r. 8.693)	RMSEA = 0.000

Panel C - bootstrap results	
Bootstrap samples = 500	Bollen-Stine bootstrap p = 0.752

Additionally, the Bollen-Stine bootstrap p does not reject the model. The Bollen-Stine bootstrap is a bootstrap modification of model chi^2, which is used to test model fit, by, for example, adjusting for the lack of multivariate normality. Similar to the chi^2, the Bollen-Stine bootstrap p is not significant at p = 0.752, which indicates a fit between the data and the model. The bootstrap statistics are presented in table 3.10 - panel C.

In sum, the base model supports all of the expected relations, except for the effect of environmental uncertainty on noise in the performance evaluation. Although the data show potential issues due to multivariate non-normality, additional analyses show that these potential issues do not substantively influence the results. The base model uses the primary operationalization of RPE use, which asks to what extent the respondents perceive that the performance of their peers is a point of reference for his superior when evaluating

the performance. The next subsection uses an alternative operationalization of RPE use to further assess the robustness of the findings.

3.4.2.2 Analyses with RPE-based-Targets Measure

To further buttress the findings of the base model analyses, I run a robustness check. The robustness check consists of rerunning the base model analysis with a different operationalization of RPE use (*i.e.*, RPE-based-Targets).

RPE-based-Targets specifically measures the explicit use of RPE for target setting purposes (*i.e.* how peer performance influences the difficulty of the performance standards used to evaluate the manager's performance). I measure RPE-based-Targets as the impact of peer performance on the performance target. I ask to what extent the performance of the peers influences the difficulty of the target and/or budget for a variety of performance measures as described in the appendix (Q2). To interpret the impact of peers on the evaluation as a whole, I weigh the various performance metrics. I weigh the impact of peers on each of the metrics by multiplying the scores with their individual importance for the performance evaluation (Q3). RPE-based-Targets does not solely measure *ex-post* peer comparison. The influence of peer performance on the agent's target can also take place at the beginning of the period. In this case, the target for the upcoming period is based on the actual performance of peers during the current period. I expect the second RPE measure to correlate with the primary measure 'RPE-Use', because they both include explicit RPE use and *ex-post* application of the peer comparison information.Because of the differences between the measures presented above, the correlation is not necessarily strong. Correlation analysis confirms this; the two measures for RPE are correlated, but not strongly (0.498; $p < 0.01$).

The analyses with the RPE-based-Targets measure support the findings of the base model's analyses. The robustness check yields results that are qualitatively similar to those of the model with the primary RPE measure (RPE-Use). The model is significant (chi^2 statistic $p = 0.274$). Additionally, the goodness-of-fit statistics suggest a sufficient model fit. The explained variance in NOISE is slightly higher than the explained variance in the base model at 13%. Similar to the base model, the model with the RPE-based-Targets proxy supports the main research hypothesis regarding the noise mitigating effect of RPE (H1) and the additional expectations regarding the drivers of noise. Additionally, in accordance with the base model, environmental uncertainty does not increase the amount of noise in the performance evaluation. The results of the robustness check are summarized in table 3.11. According to the kurtosis, the current model also has potential non-normality issues. Similar to the base model, a Bolen-Stine bootstrap ($p = 0.305$) indicates the fit between the data and the model. In other words, the kurtosis does not significantly affect the results of the estimated model. In sum, the robustness check supports the validity of the findings in the base model's analysis. Although minor quantitative changes occur, the implications of the models are qualitatively similar.

Table 3.11: Results of SEM Analysis (RPE-based-Targets)

Panel A - robustness check results			Coefficient	Std. Error	Critical Ratio	Prob.
RPE use	>	NOISE	-0.132	0.066	-2.187	0.01[‘a’]
Envir. unc.	>	NOISE	-0.026	0.117	-0.430	0.33[‘a’]
Envir. unc.	>	RPE use	0.221	0.108	3.589	0.00[‘a’]
Goal Ambiguity	>	NOISE	0.226	0.148	3.139	0.00[‘a’]
Meas. of Outputs	>	NOISE	-0.132	0.130	-1.583	0.06[‘a’]
Emph. on PLM	>	NOISE	-0.105	0.369	-1.598	0.06[‘a’]
Emph. on DM	>	NOISE	-0.162	0.419	-2.480	0.01[‘a’]
BU Size	>	NOISE	.0.034	0.038	0.488	0.63
Firm Size	>	NOISE	0.037	0.029	0.564	0.57
Dummy prod.	>	NOISE	-0.054	0.182	-0.873	0.38
Dummy fin. serv.	>	NOISE	0.001	0.196	0.009	0.99
Dummy N.F.P.	>	NOISE	0.005	0.192	-0.080	0.94

[‘a’]: variable based on directional expectation; significance calculated as one-tailed p-value

Panel B - robustness check model fit statistics	
$Chi^2 = 11.025$	*goodness-of-fit statistics*
Degrees of freedom = 9	GFI = 0.993
Probability level = .274	AGFI = 0.938
	CFI = 0.994
Multivariate normality	RMR = 0.027
Kurtosis 18.003 (c.r. 7.780)	RMSEA = 0.030

Panel C - bootstrap results	
Bootstrap samples = 500	Bollen-Stine bootstrap p = 0.305

3.4.2.3 Endogeneity Analysis

A special concern with this chapter's model is the issue of endogeneity. Endogeneity is an econometric problem that leads to biased and inconsistent estimators within the equations used to test theoretical propositions. This bias renders any inferences problematic and consequently reduces the confidence in the conclusions drawn from the research (Chenhall & Moers 2007). Although endogeneity is a key limitation in most empirical research in general (Ittner & Larcker 2001), the current study has specific reasons for analysing endogeneity because simultaneity bias is particularly likely to occur. Simultaneity arises if one (or more) of the explanatory variables is jointly determined with the explained variable. In this case, the causal relationship between an explained and an explanatory variable runs both ways (Chenhall & Moers 2007). This is the case with noise in the performance

evaluation and RPE use, since RPE use can reduce noise in the performance evaluation. However, noise can also increase organizational reliance on RPE, as argued in chapter 2.

I partially mitigate this problem in the models presented in the preceding sections by including the effect of environmental uncertainty on RPE use. I argue that environmental uncertainty is a factor that simultaneously affects noise in the performance evaluation and RPE use. Additionally, I conduct a robustness check on the models analyses to address the issue of potential remaining endogeneity. Following the recommendations of Antonakis *et al.* (2010), who review threats to the validity of SEM, I conduct a Hausman test. I assess endogeneity for both RPE measures (RPE-Use and RPE-based-Targets). However, as the procedure and the results are similar for both RPE measures, only the test for the primary RPE measure (RPE-Use) is described in this section.

The Hausman statistic is computed by conducting two OLS regressions after one another. The first regression predicts RPE-Use based on one or more instrumental variables, and all other variables in the original equation. The instrumental variable(s) should correlate with RPE-Use, but not with the error term of the structural model (Larcker & Rusticus 2010). The second regression includes the original (observed) RPE-Use measure, the predicted RPE-Use measure from the first stage and the original independent variables. The Hausman test assesses the significance of the predicted RPE-Use in the second regression.

Proper instrumental variables should be both *relevant* (*i.e.*, they should be related to the endogenous explanatory variable RPE use) and *exogenous* (*i.e.*, they should be unrelated to the undefined residual of noise in the structural model). In my analyses, I select one instrumental variable from the questionnaire that I associate with RPE use. This item asks to what extent peer performance is a point of reference for the respondent's superior when evaluating the human resources (HR) function in the respondent's business unit. Whereas RPE use refers to the performance evaluation of the business unit manager, the questionnaire item focuses on the general assessment of a smaller organizational part. Still, both RPE use and the selected questionnaire item relate to some kind of peer comparison for the evaluation of organizational parts. Accordingly, I believe that the selected questionnaire item is related to RPE use and that it is a relevant instrumental variable.

Arguing why an instrumental variable is exogenous, is, however, problematic in most research, because it requires one to demonstrate that an instrumental variable is not associated with the undefined structural residual. However, what researchers can do instead is try to find reasons that falsify this association's non-existence. To the best of my ability, I could not find a theoretical argument for how the use of peer performance as a point of reference for the evaluation of the HR function is related to the structural residual of noise in the performance evaluation (*i.e.*, the part of noise that is not explained by other exogenous factors in the structural model).

Larcker & Rusticus (2010) and Chenhall & Moers (2007) argue that, in addition to providing (at least a beginning of) theoretical argumentation for the instrument's relevance, a statistical specification test is required. The specification test for the instrumental variable's relevance is the partial F-test[8], which examines whether the instrument is significantly related to the endogenous explanatory variable. This test is conducted by evaluating the partial F-statistic of the first-stage regression analyses. This partial F-statistic indicates whether the addition of the instrumental variable contributes to the model. Given the fact that I use a single instrumental variable, a partial F-statistic value of above 8.96 suggests that the instrumental variable is relevant (Larcker & Rusticus 2010). The first-stage F-statistic in my model is 30.43 ($p < 0.01$). Considering both the theoretical and statisical assessment of the instrumental variable, the selected questionnaire item is used as a single-item instrumental variable.

The first-stage regression analysis is significant (ANOVA F-statistic is 4.140). This regression explains 14.7% of the variance of RPE-Use (Adjusted $R^2 = 11.2$). Additionally, the instrumental variable is significant. Following the recommendations of Larcker & Rusticus (2010), I present the coefficient estimates for the first-stage model (see table 3.12).

Table 3.12: Results of First-Stage Regression Analysis (RPE-Use)

	Coefficient	Std. Error	t-Stat.	Prob.
Constant	0.891	0.817	1.091	0.28
Instrumental Variable	0.180	0.033	5.508	0.00
Environmental Unc.	0.198	0.100	1.991	0.05
Goal Ambiguity	0.116	0.127	0.918	0.36
Measurability of Outputs	0.113	0.111	1.019	0.31
Emphasis on PLM	0.461	0.296	1.558	0.12
Emphasis on DM	0.092	0.346	0.265	0.79
BU Size	0.001	0.030	0.022	0.98
Firm Size	0.012	0.028	0.426	0.67
Dummy production	0.153	0.158	0.968	0.33
Dummy financial services	0.078	0.164	0.476	0.64
Dummy not-for-profit	-0.076	0.171	-0.444	0.66
$R^2 = 0.147$			F-statistic = 4.140	
Adjusted $R^2 = 0.112$			Prob(F-statistic) = 0.000	

Included observations: 276

[8]Larcker & Rusticus (2010) also refer to this test as the 'weak instruments' test.

The predicted value of RPE-Use that follows from this regression is then used as a proxy for RPE-Use in the second regression. In addition to the predicted RPE-Use variable from the first regression, the second regression includes the original (observed) RPE-Use measure, and the original independent variables (environmental uncertainty, goal ambiguity, measuruability of outputs, emphasis on personal-level measures, emphasis on disaggregated measures, and the control variables). The second regression, presented in table 3.13, is also significant (ANOVA F-statistic is 2.844). The model explains 11.5% of the variance in noise in the performance evaluation (Adjusted R^2 = 7.4%).

Finally, I conduct the Hausman test for endogeneity to assess the existence of an endogeneity problem. I follow Larcker & Rusticus' (2010) explanation, which argues that if the coefficient for the predicted RPE-Use is significant, the Hausman test rejects the null of no endogeneity problem. In my second regression, the predicted RPE-Use is insignificant (p = 0.378). This finding implies that the model does not suffer from endogeneity problems[9]. This finding reinforces the results of the previous analyses.

Table 3.13: Results of Second-Stage Regression Analysis (RPE-Use)

	Coefficient	Std. Error	t-Stat.	Prob.
Constant	3.683	0.957	3.848	0.00
Unstandardized Predicted Value	0.193	0.218	0.883	0.38
RPE use	-0.167	0.070	-2.385	0.02
Environmental Unc.	-0.067	0.125	-0.537	0.59
Goal Ambiguity	0.379	0.146	2.601	0.01
Measurability of Outputs	-0.267	0.131	-2.041	0.04
Emphasis on PLM	-0.647	0.357	-1.811	0.07
Emphasis on DM	-0.681	0.394	-1.728	0.09
BU Size	0.004	0.035	0.105	0.92
Firm Size	0.023	0.032	0.717	0.47
Dummy production	-0.070	0.180	-0.390	0.70
Dummy financial services	0.096	0.186	0.516	0.61
Dummy not-for-profit	0.103	0.194	0.530	0.60
R^2 = 0.115			F-statistic = 2.844	
Adjusted R^2 = 0.074			Prob(F-statistic) = 0.001	

Included observations: 276

[9]The absence of the endogeneity problem does not mean that RPE use and noise in the performance evaluation are not determined simultaneously. The Hausman test results do not imply that the dependent and independent variables are not interrelated; the test only suggests that the model estimation is unbiased (Abernethy *et al.* 2004).

3.4.3 Summary of the Findings

The analyses show robust support for the main claim of RPE theory (*i.e.*, RPE reduces the amount of noise in the performance evaluation). This finding is robust for two operationalizations of RPE that capture various applications of RPE in practice. The two RPE use measures include situations in which peer performance functions as an implicit standard when appraising the performance of the business unit manager as well as situations in which peer performance directly influences the difficulty of the performance target. Regardless of the operationalization of RPE use, I find that RPE affects the amount of noise in the performance evaluation. When RPE use increases, noise is reduced.

Additionally, the analyses show the effects of a variety of factors on noise. For example, goal ambiguity increases the amount of noise in the performance evaluation, whereas increased measurability of the outputs of the BU and the emphasis on personal and disaggregated performance measures reduce the amount of noise. The analyses show no direct effect of environmental uncertainty on the amount of noise in the performance evaluation.

3.5 Conclusions and Discussion

This chapter studies the noise-reduction claim that underlies RPE theory. Noise reduction is a primary explanation of RPE use in practice (see, for examples, Holmstrom 1982, Gibbons & Murphy 1990, Matsumara & Shin 2006, Albuquerque 2009, Gong *et al.* 2011, and chapter 2 of this thesis). Noise results from the inclusion of windfall gains and losses in the performance evaluation of the employee. By reducing noise in the performance evaluation, RPE use can benefit organizations. Noise reduction improves the contractual efficiency between the owner and the employee by improving the performance evaluation. According to RPE theory, RPE can restore the quality of the performance evaluation by filtering out the noise from the employee's measured performance (Holmstrom 1982, Gibbons & Murphy 1990). RPE reduces noise by incorporating information about the attainable performance level into the evaluatee's performance evaluation via peer comparison (Holmstrom 1982). RPE reduces noise by incorporating information about the attainable performance level into the evaluated employee's performance evaluation via peer comparison (Holmstrom 1982). RPE compares the employee's actual performance with the performance of a reference group that partially faces the same external events. The performance of the reference group informs the principal about the performance level that should be attainable for the agent under the external events that have occurred. The comparison with peer performance constructs a performance norm that is (at least partially) insulated from the impact of external events.

The noise-reduction effect of RPE finds support within both the analytical and empirical literature. However, the support for this relationship is based on rather indirect analyses,

leaving ample room for rival explanations. In chapter 2 of this thesis, the support for the noise-reduction explanation is based on the empirically significant relation between environmental uncertainty (*i.e.*, a driver for noise) and the extent to which organizations rely on RPE. However, other explanations in the literature explain the relation between environmental uncertainty and RPE. For example, Holmstrom (1982) argues that RPE can be used as a means to pit employees or organizational parts against each other to create a sense of internal rivalry that increases the efficiency of the individual employees. Additionally, Gibbons & Murphy (1990:33-S) show that RPE is used to facilitate organizational learning. Both of the presented applications can be especially useful in a highly uncertain environment, where environmental turbulence demands that a firm exhibits its optimal performance to survive. As a result, the analyses in chapter 2 cannot empirically distinguish between the noise-reduction explanation and the two alternative explanations.

To provide a more direct analysis of the noise-reduction explanation of RPE theory, this chapter analyses the effectiveness of the noise-reduction properties of RPE. This analysis assesses whether organizational reliance on RPE actually reduces the noise levels in the performance evaluation of employees. For the empirical analysis, I use a SEM model to analyse survey data of 325 business unit managers in the Netherlands. Overall, the results are consistent with the findings of the analytical agency literature and the conclusions of chapter 2. The analyses support the main claim of RPE theory (*i.e.*, RPE can reduce the amount of noise in an agent's performance evaluation). This finding is important because it further validates the noise-reduction explanation on RPE.

The evidence for the noise-reduction claim is robust over two operationalizations of RPE use. Similar to chapter 2, this chapter relies on both a broad, unspecific measure of RPE use and on a measure that focuses on the explicit application of RPE to performance targets. Both measures yield qualitatively and quantitatively similar results. The SEM models explain 11-13% of the variance in the dependent variable 'noise in the performance evaluation'. Additionally, both models indicate an adequate fit on a number of goodness-of-fit indicators, as suggested by Kline (2005).

However, the models do have important limitations. These limitations lie in the measurement of the dependent variable. Firstly, the dependent variable *noise in the performance evaluation* is measured with a single item construct. Using single-item measures increases measurement error in the estimation and potentially reduces the validity of the model and the findings. Although pre-testing and robustness checks support the convergent validity of the single-item construct, the use of single-item constructs remains less than perfect. Secondly, the responses on the noise-measure may be biased to the extent that weaker performance managers may attribute negative performance to exogenous factors more than high performing managers. This potential bias may introduce some noise in my noise measure. Despite these limitations, I argue that the current study furthers our confidence in the cornerstone explanation of RPE theory. Because of more direct evidence on the implied causal mechanism underlying RPE theory, we can be more certain about the effectiveness of RPE in reducing noise in the performance evaluation.

Chapter 4

Does Relative Performance Evaluation Mitigate Room for Managerial Opportunism?

Preamble to Chapter 4

The current chapter addresses the claim that RPE reduces the room that managers have to behave opportunistically. By doing so, this study answers the second part of research question 3: *is RPE effective at reducing noise and opportunism?*

The opportunism-mitigation claim is also assessed in chapter 2, together with the noise-reduction claim. Chapter 2 refers to the opportunism-mitigation perspective to motivate the effect of opportunism drivers on RPE, and finds mixed support for this relation. The current chapter analyses the opportunism-mitigation claim in a more direct manner. Instead of analysing whether RPE use is increased by factors that facilitate managerial opportunism, the current chapter studies whether RPE effectively reduces the room for managerial opportunism.

Further studying the opportunism-mitigation perspective is important for our understanding of the consequences of RPE use for Management Control. The analyses shed light on whether the inconclusive results in chapter 2 are due to RPE's limited added value to reduce opportunism, or whether the results are, for example, driven by the statistical properties of the analyses in chapter 2. Together with the assessment of the opportunism-mitigating properties of RPE, the current chapter adds to our understanding of whether RPE use can be explained by an organizational wish to restrain managers from engaging in behaviour that is not congruent with the organization's objectives.

4.1 Introduction

The current chapter empirically assesses whether Relative Performance Evaluation reduces managers' room for opportunistic behaviour, as argued by Murphy (2001). Murphy claims that RPE reduces opportunism because RPE makes it more difficult for members of the organization to opportunistically influence their own performance targets. The tendency to behave opportunistically is a central tenet in economics-based theorizing about Management Control. Economic theories provide us with a model of man, who shows self-interest seeking behaviour with guile (Williamson 1975:255). Opportunistic behaviour is not simply a theoretical matter but is highly recognizable in actual behaviour. Prior research indicated that opportunistic behaviour is a realistic human trait. For example, an experimental study by Maas & Van Rinsum (2011) shows that more than half of the study's participants act opportunistically if possible by overstating their performance to maximize their individual pay-offs. An earlier study by Hannan *et al.* (2006) provides a similar picture of participants of an experiment opportunistically misrepresenting their own performances.

The consequences of opportunistic behaviour within organizations can be significant. If the members of the organization seek their own self-interest and act against organizational goals (*e.g.*, by shirking or committing fraud), their behaviour reduces organizational gains. Another manifestation of opportunistic behaviour relates to the determination of performance targets. Murphy (2001) argues that organizational members often can opportunistically influence their own target difficulty. For example, if performance standards are based on prior-year performance, managers can affect their target-levels by avoiding unusually positive outcomes.

However, according to Murphy's reasoning, organizations can use RPE to reduce the room for managerial opportunism. RPE places the determination of the performance target outside the employee's sphere of influence, where it is more difficult for the employee to lower his target difficulty. Under RPE, the performance of the evaluated agent is compared with the performance of an external reference group of agents. This peer comparison determines a performance standard to which the agent's performance is benchmarked. The 'external determination' of the performance target renders it less vulnerable to managerial opportunism, than performance targets that are determined within the employee's 'reach'.

Murphy's opportunism perspective provides a compelling theoretical explanation for RPE use. However, the empirical support for this explanation of RPE is limited. Murphy finds empirical support for his reasoning in a study that does not explicitly focus on RPE, but on externally determined targets in general. Murphy shows that companies using budget-based and other internally determined performance standards have less variance in the bonus payouts and are more likely to smooth their earnings than companies using externally determined standards (Murphy 2001:245).

Moreover, chapter 2 of my thesis, which explicitly focuses on RPE, does not report robust support for the opportunism-mitigation perspective on RPE. To my knowledge, chapter 2 of this thesis is the first empirical study that focuses on the opportunism-reducing properties of RPE *within* the organization (*e.g.*, amongst business unit managers). The inconclusive results in chapter 2 can be the result of the limited added value of RPE to reducing opportunism at the lower echelons of the organization. However, the limited support may also be due to data limitations and the specifications of the statistical model, specifically with regard to the use of statistically demanding interaction effects.

In chapter 2, I study the opportunism-reducing properties of RPE by examining the interaction effect between information asymmetry and the comparability of the business unit. Information asymmetry enables employees to opportunistically misrepresent information to top management to achieve easier performance targets. High levels of information asymmetry increase the attractiveness of using RPE to mitigate this opportunism. However, informative peer comparison is only possible if the peers are sufficiently comparable. I model this condition as an interaction effect between information asymmetry and comparability that influences RPE use. The combination of these factors increases the attractiveness of using RPE to mitigate opportunism. However, the analyses did not robustly support this interaction effect. This finding can result in a potential rejection of the opportunism-mitigation explanation for RPE.

However, tests of hypotheses pertaining to the effects of interactions often have low statistical power (Aguinis 1995:1142, Cohen *et al.* 2003:297). Hypotheses using interaction effects with relatively small datasets can easily generate statistical null results that lead to false rejections of the research hypothesis. These false rejections are known as type II errors and can cause researchers to find effects that exist in the population to be statistically insignificant in the sample.

The current chapter re-examines the opportunism-mitigation explanation for RPE use amongst business unit managers by employing a different approach that does not require interaction effects. Similar to chapter 3, the current chapter analyses the consequences of RPE use. In this chapter, I do not study whether RPE use is more prevalent when increased room for managerial opportunism exists. Rather, I study whether business unit managers that use RPE have less room to behave opportunistically. By addressing the opportunism-mitigation perspective on RPE with a different approach, the current study contributes to our knowledge regarding the utility of RPE for Management Control in practice.

Contrary to the expectations, the findings of the current chapter do not support Murphy's explanation of the opportunism-mitigating properties of RPE. This result confirms and extends the overall findings of chapter 2. Finding similar results after respecifying the model strengthens the conclusions of chapter 2. Despite Murphy's (2001) own empirical support for his theory at the level of executive officers, my findings suggest that RPE is not an effective method for reducing opportunism at the level of business unit managers.

To my knowledge, this study is the first that maps the room managers have to engage in opportunistic behaviour. The findings of this chapter suggest that control systems are generally well equipped to mitigate opportunistic behaviour. In my sample, both the RPE-adopters and non-RPE-adopters claim to have limited room for opportunism under their currently installed control systems. This finding may have important implications for control system design, by informing practice about the necessity to implement additional opportunism-mitigating control instruments.

The remainder of this chapter is organized into four sections. Section 2 provides the development of the hypotheses and presents the research model. The measurement is described in section 3, and the analyses and results are described in section 4. Section 5 addresses the findings, and the conclusions.

4.2 Development of the Model

This section describes the development of the research model, which consists of the research hypothesis and additional expectations. Subsection 4.2.1 presents the research hypothesis, which addresses the assumed opportunism-mitigating properties of RPE. In subsection 4.2.2, I present a set of expectations consisting of opportunism-inducing and opportunism-reducing factors. These antecedents are included as a means to address a measurement problem related to the model's dependent variable rather than because of their theoretical relevance to the opportunism-reduction assumption per se. The dependent variable *room for managerial opportunism* has a measurement problem similar to that of the *noise in the performance evaluation* measure described in the previous chapter. This variable is based on a perceptual measure that relates to how respondents perceive the room to act opportunistically under their currently installed control systems. However, the currently perceived room for opportunism includes the effect of potential opportunism-mitigating control instruments, such as RPE. If I aim to estimate the effect of RPE on the room for managerial opportunism, I must predict the 'theoretical' room for opportunism before this space is reduced by RPE. To estimate the effect of RPE on the room for managerial opportunism, I extend the model by employing factors that affect the room for opportunism outside the influence of RPE. These antecedents include information asymmetry, goal ambiguity, measurability of outputs, and decentralization of decision rights. In subsection 4.2.2, I argue how these antecedents (positively or negatively) affect the room for managerial opportunism.

4.2.1 Research Hypothesis: Managerial Opportunism and RPE

The research hypothesis concerns the negative effect of RPE use on the room for managerial opportunism. According to Murphy (2001), an important source of room for opportunism

is the way in which performance standards are determined. Murphy argues that various standard-setting processes differ in their vulnerability to opportunism. The performance standard-setting process can assume two main forms: 1) standards can be determined by some internal, administrative process; or 2) standards can be determined externally (Murphy 2001). The difference between these alternatives lies in the extent to which agents can influence the target difficulty (Murphy 2001)[1], as described earlier in this thesis. Internally determined standards include standards based on prior-year performance and standards derived from firm plans or budgets. Such standards are affected by managerial actions and can have dysfunctional effects. For example, if standards are based on prior-year performance, managers have an incentive to avoid unusually positive outcomes, because good performance in the current period is penalized by an increased standard in the next period (Murphy 2001). Similarly, budget-based standards provide incentives to negotiate easy standards (Fisher, Frederickson & Peffer 2002) and disincentives to beat the budget, especially in a regime of incremental budgeting (Murphy 2001). In contrast, externally determined standards are less affected by managerial actions because the difficulty of the standard is based on something outside the managerial sphere of influence (*e.g.* the target is based on market conditions or peer performance). Murphy supports this theoretical claim with empirical results. His data show more year-to-year variance in the bonus payouts of top executives from companies using externally determined performance standards than those of top executives from companies using budget-based and other internally determined performance standards. This finding suggests that executives rewarded based on internal standards are more likely to opportunistically smooth earnings than executives rewarded based on externally determined standards (Murphy 2001:245).

RPE is an external standard-setting method (Murphy 2001:252). Under RPE, the performance of the evaluated agent is compared with the performance of an external reference group of agents[2]. This peer-comparison determines a (explicit or implicit) performance standard to which the agent's performance is benchmarked. Although the performance of the peer group generally lies well outside the managerial sphere of influence, their performance does constitute the evaluated agent's performance target. This RPE-based target is not subject to the business unit's prior-year performance, to negotiations between the business unit manager and the firm's top management, or to anything else that can be manipulated easily by the business unit manager. These considerations suggest that the room for managerial opportunism can be delimited by incorporating the performance of a reference group of agents into the compensation plan and that RPE can aid the standard-setting process. As a result, RPE-based standards are less vulnerable to managerial opportunism.

[1]Murphy (2001) notes that the distinction between internally and externally determined performance standards is not dichotomous but rather one of degree. Even externally determined standards are manipulable to some extent, especially when they are established for the first time and if major changes in the environment force a modification (Murphy 2001:249).

[2]'External' means that the reference group is external to the business unit. However, the group is not necessarily external to the firm. A reference group might consist of other business units inside the firm that somehow face (some of) the same external conditions (*e.g.*, they operate within the same market).

Based on Murphy (2001), this study expects a negative relation between the use of RPE and the room for managerial opportunism. The research hypothesis is formalised as follows:

H1: The use of RPE reduces the room for managerial opportunistic behaviour.

4.2.2 Antecedents to the Room for Managerial Opportunism

To analyse the effect of RPE on the room for managerial opportunism, this study includes antecedents to the room for managerial opportunism. These facilitators are included in the model as a means to deal with the measurement characteristics of the dependent variable, as explained in the introduction of this section. These facilitators constitute the room that managers have to behave opportunistically, *as if* RPE did not mitigate this space. Although facilitators of managerial opportunism have not been extensively studied in the empirical literature, prior research has provided a starting point for theorizing about these factors. A review of the literature on the relevant antecedents to opportunism has led to a paper by Wathne & Heide (2000), who study the relation between information asymmetry and opportunistic behaviour. Their main finding is discussed in the next paragraph and provides a starting point for assessing the room for opportunism.

Information Asymmetry Information asymmetry is expected to increase the room for managerial opportunism. Wathne & Heide (2000) argue that information asymmetry is one of the two conditions that facilitate opportunism[3]. Information asymmetry refers to the information gap between the principal and agent concerning, for example, their knowledge regarding the quality of a business unit's performance, the (internal and external) factors that influence the performance, and the technical processes and the transformation processes of the unit. A manager pursuing his own self-interest can exploit this information gap. Under asymmetric information, the superior's ability to detect opportunism is limited. Managers can realize individual gains by strategically disclosing information that is asymmetrically distributed throughout the organization. By misrepresenting the performance potential of the business unit, the manager can, for example, negotiate easy targets or build budgetary slack (Dunk 1993, Chow *et al.* 1988, Fisher, Maines, Peffer &

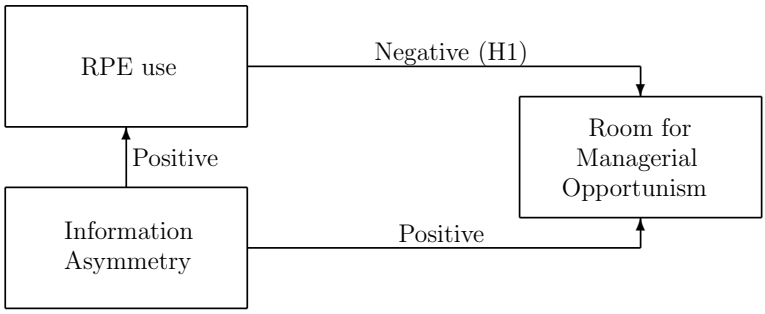
[3]The second condition is the presence of a lock-in situation between two parties, where one party is willing to accept a certain amount of opportunistic behaviour from the other party because the costs of retreating from the contract are too high. As the contract between the respondents and their superiors is placed within the hierarchy and not (as is the case with Wathne & Heide 2000) within the market or in a hybrid structure, the current study considers the lock-in situation to be present amongst all respondents. Consequently, for this study within the hierarchy, the extent to which the contracted parties are locked-in, is not considered to vary. Therefore, the degree to which a party is locked in is not a variable in the model and lies outside the scope of this study.

Sprinkle 2002). Based on Wathne & Heide (2000), the current study considers information asymmetry as the first antecedent to room for opportunism. This expectation is formalized as follows: *information asymmetry has a positive effect on the room for managerial opportunism.*

Following the hypotheses described in chapter 2 of this thesis, I also expect information asymmetry to increase organizational reliance on RPE[4], because RPE can partially reduce the room for managerial opportunism resulting from information asymmetries. If the superior has limited knowledge of the factors that influence performance because of high levels of information asymmetry, he can observe the effects of these factors through changes in peer performance by using RPE. Accordingly, I expect the following: *information asymmetry has a positive effect on the use of RPE.*

Figure 4.1 summarizes the relations among RPE use, information asymmetry, and the room for managerial opportunism.

Figure 4.1: Conceptual Model



Goal Ambiguity The second antecedent to the room for managerial opportunism is goal ambiguity. Goal ambiguity refers to the degree to which the organization is able to formulate the ultimate goals of the business unit to the business unit manager. Whereas specific goals are a clear statement of the firm's expectations from the business unit, ambiguous goals leave room for interpretation and discussion from the side of the business

[4]Whereas chapter 2 assesses the relation between RPE and information asymmetry, which interacts with the level of comparability between the business unit and its peers, the current chapter refrains from using interaction effects. This interaction effect is not required here because information asymmetry increases the room for managerial opportunism regardless of the business unit's comparability to its peers.

unit manager. A situation characterised by high goal ambiguity can be exploited by a self-interested manager. Ambiguous goals provide the manager with room to argue for his own preferred course of action, which may not be aligned with the goals of the firm. Therefore, this study considers goal ambiguity as the second facilitator of managerial opportunism. I formalize this expectation as follows: *goal ambiguity has a positive effect on the room for managerial opportunism.*

Measurability of Outputs The third opportunism-driving factor in the model is measurability of outputs, which refers to the degree to which an organization can measure its output in a reasonably undistorted manner. If the measurability of outputs is high, the business unit's output metrics correspond well with the actual goals that the business unit needs to accomplish. High levels of output measurability facilitate the manager's ability to communicate goal achievement to his superiors. However, if the measurability of outputs is low, monitoring becomes more difficult (Ouchi 1979, Langfield-Smith & Smith 2003). Because the actions of the business unit manager are less observable under conditions of low output measurability, low measurability creates leeway for the manager to act opportunistically without being caught by his superior. Based on this reasoning, this study argues that a low level of output measurability increases the room that managers have to behave opportunistically. This expectation is formalised as follows: *measurability of outputs has a negative effect on the room for managerial opportunism.*

Decentralization of Decision Rights Decentralization of decision rights refers to decisions at the business unit level concerning, for example, the business unit's investments or marketing choices. If an organization transfers decision rights from the CEO to the business unit manager, the business unit manager has more room to engage in opportunistic behaviour. The choices that are in the BU manager's best interest are not necessarily in the organization's best interest. For example, the business unit manager may choose to implement projects that he finds interesting but that are not in the firm's best interest. Additionally, the manager might exploit the firm's high-quality brand name to market the business unit's lower-quality products to maximize the BU's profits and/or the manager's individual bonus. By delegating the decision rights to the BU manager, the organization provides the manager an opportunity to make self-interested decisions. Therefore, this study argues that the decentralization of decision rights increases the room for managerial opportunism. The final expectation in this study is formalized as follows: *decentralization of decision rights has a positive effect on the room for managerial opportunism.*

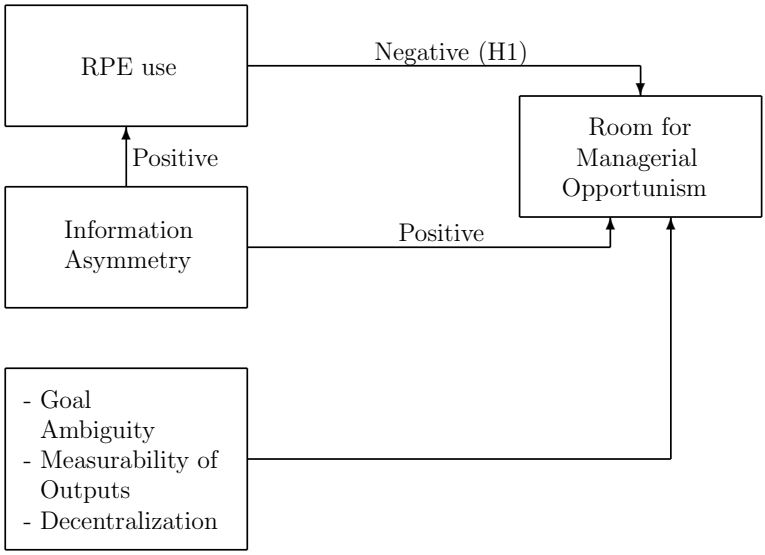
4.2.3 Summary of the Model

The model, which consists of the research hypothesis and the additional expectations, is summarized in figure 4.2.

4.3 Sample and Measurement

The current section contains parts that are largely similar to parts of chapters 2 and 3. This section is similar because it uses the same survey instrument and, in part, the same measures. Subsections 4.3.1 and 4.3.2 and parts of subsection 4.3.3 are marked with an asterisk to indicate this repetition. The reader may wish to skip these parts.

Figure 4.2: Causal Model



4.3.1 Sample(*)

This study uses primary data collected from 325 business units with an extensive survey instrument. The questionnaire was filled out by business unit managers contacted through students' professional networks. This approach has several benefits. The first is respondent identification. The students were instructed to identify managers who are responsible for an autonomous organizational body engaged in multiple activities (*e.g.*, purchasing, production, and sales). Furthermore, the respondent should have a superior within the organization (in other words, the respondent should not be the CEO). The second benefit is the possibility to conduct face-to-face interviews, which improves the respondent's valid interpretation of and careful responding to the questions. During these interviews, the student can explain the survey questions (using a strict protocol provided by the researcher)

to the business unit manager who answers the questions in the survey and fills out the questionnaire.

The observations include data from production business units (17%), service BUs (56%), financial services (16%), and BUs operating in the not-for-profit sector (18%)[5]. Similarly diverse responses to the variance of sectors at the BU level were provided on the question about the respondents firm's sector. Table 4.1 summarizes the background information on the respondents' business units and the firm to which each business unit belongs.

Table 4.1: Summary Statistics on Sample

Panel A - size

		N	Mean	Std. dev. of mean	Min.	Max.
BU size	FTE	325	220	439	10	5.000
	Revenue (mio EURO)	292	196	840	1	1.000
Firm size	FTE	321	14.730	41.771	31	470.123
	Revenue (mio EURO)	301	7.775	43.240	4	600.000

Panel B - sector

BU sector		
	Production	54
	Services	181
	Financial	52
	Not-for-profit	57

Firm sector		
	Production	71
	Services	170
	Financial	62
	Not-for-profit	64

4.3.2 The Questionnaire(*)

The survey instrument was specifically designed for this research, following the Tailored Design Method (Dillman 2000), where applicable to the specific research setting[6]. The survey instrument relies heavily on previously used and validated questions. Some of those measures are well established (*e.g.*, the measure for information asymmetry is largely based on Dunk (1993)). Other questions were specifically constructed for this survey. For example, the measures for the dependent variable that measures the use of RPE are purpose-developed.

[5]These percentages sum to 107% because some business units operate in more than one sector. Multiple answers to the related question are considered valid.

[6]The Tailored Design Method is specifically designed for mail and internet surveys. Many of the guidelines provided by Dillman (2000) are applicable to the survey design, but others specifically apply to mailed surveys instead of face-to-face survey interviews.

To increase the reliability of the instrument, the survey instrument was thoroughly pre-tested using Hak's Three-Step Test Interview method (abbreviated: TSTI) (Hak *et al.* 2008). This method is an observational instrument for pretesting self-completion questionnaires[7]. TSTI works through observing actual instances of interaction between respondents and the instrument. Because this process mainly consists of cognitive processing and is therefore hidden from the observer, 'thinking aloud' is used as a technique for making the thought process observable (Hak *et al.* 2008). Through this process, TSTI helps to identify problems in questionnaires, often leading to modification of the instrument. Three rounds of pre-tests with twelve managers[8] ensured a thorough and correct understanding of the questions and response categories.

4.3.3 Measurement

This section describes the measurement of the variables used for this study. First, I discuss the measurement of the dependent and independent variables (which include *room for managerial opportunism*, *RPE use*, and the antecedents to opportunism (*i.e.*, *information asymmetry*, *goal ambiguity*, *measurability of outputs*), and *decentralization of decision rights*). This section concludes with a description of the measurement of the control variables for size and sector. When this study refers to the questionnaire items, please see the appendix at the end of the thesis for the exact phrasing, the response categories, and, if applicable, the sources.

The Room for Managerial Opportunism Many prior studies and reviews have discussed the tendency of agents to act opportunistically (see, for example, Williamson 1985, Eisenhardt 1989, Speklé 2001, and Macher & Richman 2008). However, because the accounting & control literature generally does not empirically assess opportunism (Macher & Richman 2008), no measures of opportunism were readily available[9]. Without applicable measures available from prior studies, the variable room for managerial opportunism is originally developed for this study. The dependent variable Room for Managerial Opportunism (abbreviated: RFMO) is based on a series of four questions that ask the respondent to indicate the extent to which he can behave opportunistically under his current control system. The items and their mutual relations are summarized in table 4.2. Using these

[7]Although the survey was not a self-completion questionnaire in the sense that the respondents did not fill out the survey by themselves, but together with the interviewing student, I still value the contribution of TSTI. Filling out the survey was outside of the direct sphere of influence of the researcher, but was carried out by -carefully briefed- students.

[8]The pre-test participants' organizations show great diversity, ranging from production to consulting, educational and financial services, and from very large to smaller business units, including both for-profit and not-for-profit organizations.

[9]Studies from other fields that measure opportunism (most of which study mainly manufacturer and distributor exchange relationships in a channel setting) do not provide useful measures for this research (see, for example, John 1984, Dwyer & Oh 1987, and Gundlach *et al.* 1995).

items, I proxy for the extent to which the manager can behave opportunistically if he wanted to instead of whether he actually behaves opportunistically. This distinction is important because the room that the Management Control System leaves for potential opportunism indicates the quality of control.

Table 4.2: Items for Room for Managerial Opportunism (Q39)

	Item description	Component loadings
a.	Possibility to hide bad performance	0.848
b.	Possibility for shirking	0.816
c.	Possibility for window dressing	0.806
d.	Possibility to adopt pet projects	0.713
	Percentage variance explained	63.6%
	Cronbach's alpha	0.805

The questions for RFMO refer to the room for both active and passive forms of opportunism and include measures of misleading and lying behaviours. These behaviours include consciously providing effort that is not congruent with the firm's goals and committing fraud with accounting figures. RPE potentially restrains both forms of opportunism. Comparing a manager's performance with the performance of a peer group provides the manager's superior with information about the quality of the manager's reported performance. This information reduces the manager's ability to hide bad performance by using, for example, excuses about the market conditions under which the performance was realized. Moreover, unfavourable peer comparisons may indicate goal-incongruent behaviour, such as the adoption of a pet project. In contrast, a noticeably favourable comparison with peer performance that is not supported by a convincing explanation may lead the superior to detect, for example, fraud committed with accounting figures committed by the manager to increase his compensation.

Statistically, the individual questions combine into one factor. The four questions produce a single factor with adequate scale reliability. RFMO is calculated by averaging the scores on all of the items. The mean score on RFMO is 2.64 (St.Dev. 1.21) on a 7-point Likert-scale, which suggests that, on average, the respondents have limited room to behave opportunistically under their installed control systems.

RPE use(*) The RPE use measure (labelled 'RPE-Use') asks about the extent to which peer performance functions as a point of reference for evaluating the quality of the agent's performance (Q1). The underlying questionnaire items focus on the *ex-post* nature of performance evaluation. This measure comprises both explicit coupling of the performance target to peer performance and more implicit applications of RPE. Implicit applications

of RPE do not require that peer performance affects the performance target explicitly, but they can, for example, play a role in establishing an implicit performance standard or norm for the performance evaluation. This measure asks to what extent the respondent perceives that the performance of his peers is a point of reference for his superior when evaluating the respondent's performance. Additionally, the question differentiates between situations where the evaluated business unit performs substantially better versus substantially worse than its peers. This dinstinction controls for potential asymmetries in peer comparison. Throughout the sample, the asymmetries do not seem to be of great influence; the three settings (neutral/better/worse) lead to an internally consistent scale, as presented in table 4.3.

Table 4.3: Items for RPE-Use (Q1)

	Item description	Component loadings
a.	Peer performance point of reference	0.868
b.	Substantially better	0.923
c.	Substantially worse	0.902
	Percentage variance explained	80.6%
	Cronbach's alpha	0.879

Information Asymmetry(*) Information asymmetry is measured with a seven-item measure, largely based on Dunk (1993). There are two differences from Dunk's original instrument. First, the wording has been altered to fit better with the specific context of business unit managers. Second, one question was added to differentiate between knowledge about internal versus external factors that might influence the business unit performance (following Kruis 2008). All items load onto one factor. Overall, the construct for information asymmetry shows good internal reliability. Information asymmetry was calculated by averaging the scores on all items. Table 4.4 summarizes the construction of the information asymmetry variable.

Goal Ambiguity(*) Goal ambiguity measures the extent to which firms clearly formulate their objectives. Goal ambiguity is measured with a multi-item construct, evaluated by four questionnaire items. As summarized in table 4.5, the items refer to how clear and specific the goals are defined, and how easily they can be communicated. The Cronbach's alpha of the four items together is rather low at a value of 0.630, and cannot be improved by deletion of one of the items, as is shown in table 4.5. However, factor analysis shows that the four questions load onto one component, with all component loadings well above 0.500. Based on the adequate component loadings, I combine the items into one measure, labelled goal ambiguity.

Table 4.4: Items for Information Asymmetry (Q16)

	Item description	Component loadings
a.	Information about the activities undertaken	0.818
b.	Familiarity with the input/output relations	0.834
c.	Certainty about the performance potential	0.851
d.	Familiarity with the technical aspects of the work	0.711
e.	The impact of internal factors on the managers activities	0.697
f.	The impact of external factors on the managers activities	0.610
g.	Understanding of the achievements of the business unit	0.791
	Percentage variance explained	58.3%
	Cronbach's alpha	0.875

Measurability of Outputs(*) Measurability of outputs refers to the extent to which firms are able to measure the desired results of the business unit in a goal consistent way. Measurability of outputs is a construct evaluated by four questionnaire items, asking about the objective measurability of the outputs, including aspects of completeness and goal consistency of the output measurement. The items load onto one factor with a scale reliability of 0.74. The items and the corresponding statistics as summarized in table 4.6.

Decentralization of Decision Rights Following Abernethy *et al.* (2004) I measure the level of decentralization of decision rights by using a modified version of the instrument developed by Gordon & Narayanan (1984). The questionnaire asks the respondents to indicate their level of influence relative to their superior on five decision areas: strategy, investments, marketing, operations, and human resource management. As summarized in table 4.7, factor analysis and the Cronbach's alpha support the incorporation of all five items into a single unidimensional construct.

Control Variables(*) This study controls for business unit & firm size, as well as for sector effects. This paragraph presents the measures for these control variables.

Table 4.5: Items for Goal Ambiguity (Q17-20)

	Item description	Component loadings	Scale if item deleted
a.	Clarity of goals	0.758	0.484
b.	Specificity of goals	0.734	0.480
c.	Difficulty of explaining goals to outsiders (reversed)	0.572	0.552
d.	Clarity of goals to insiders	0.682	0.547
	Percentage variance explained	47.6%	
	Cronbach's alpha	0.630	

Table 4.6: Items for Measurability of Outputs (Q21-23)

	Item description	Component loadings
a.	Match between performance measures and business unit goals	0.731
b.	Objective measurability of goals	0.724
c.	Match between performance measures system and results	0.809
d.	Goals consistency performance measurement system	0.760
	Percentage variance explained	57.3%
	Cronbach's alpha	0.744

Size (BU and Firm)(*) Size is often regarded as a potentially important determinant of performance measurement practices (Speklé & Verbeeten 2008). For example, larger organizations might be more effective in the use and implementation of sophisticated performance measures that are less vulnerable to opportunism. Control variables *business unit size* (Q31) and *firm size* (Q32) are both based on the number of FTE and revenues. For both control variables, the natural logs of FTE and revenue were calculated. With business unit size measured as FTE and revenue, the Cronbach's alpha is rather low. However, the two items correlate positively ($r = 0.500$, $p < 0.01$). Additionally, the two factors load onto one component in the factor analysis (component loading = 0.866). Despite the low Cronbach's alpha, both items are combined into the measure business unit size. Although business unit size and firm size are significantly correlated with one another (Pearson correlation 0.479), they do not yield collinearity-issues in the multivariate analyses. Table 4.8 summarizes the questionnaire items and statistics on the size variables.

Table 4.7: Items for Decentralization of Decision Rights (Q40)

	Item description	Component loadings
a.	Influence on decisions regarding strategy	0.724
b.	Influence on decisions regarding investments	0.687
c.	Influence on decisions regarding marketing	0.774
d.	Influence on decisions regarding operations	0.759
e.	Influence on decisions regarding human resource management	0.638
	Percentage variance explained	51.5%
	Cronbach's alpha	0.762

Sector Dummy Variables(*) The analyses include dummy variables controlling for business unit industry (Q33). The model contains dummies for production, financial services, and not-for-profit business units. The base model consists of service organizations, representing the largest sector in the sample.

4.3.4 Descriptive Statistics

Table 4.9 presents the descriptive statistics. According to the skewness and kurtosis indicators, all of the variables follow a normal distribution. Furthermore, based on RFMO, we see that managers have limited room for opportunism. I measured RFMO with a 7-point scale in which the highest category (7) was awarded by only one of the respondents. On average, the respondents claimed that they somewhat disagree with the statements suggesting that they have room to act opportunistically. The mean score (2.64) lies well below the midpoint of the scale (4). Fewer than 15% of the respondents responded with neutral (4) or higher agreement with the opportunism statements. This observation implies that, to the extent that this score is due to the installed control system, control systems, on average, successfully reduce the room to act opportunistically. The following section analyses and discusses the extent to which the limited room for managerial opportunism observed in this study is due to the use of RPE.

Additionally, I find that this observation is consistent across sectors. I conducted additional analysis to explore the differences in RFMO across sectors. A one-way ANOVA analysis did not show any significant findings indicating cross-sector variance in RFMO,

Table 4.8: Items for Business Unit Size & Firm Size (Q31-32)

	Item description	Component loadings
a.	Business unit FTE (Ln)	0.866
b.	Business unit revenue (Ln)	0.866
	Percentage variance explained	75.0%
	Cronbach's alpha	0.634
c.	Firm FTE (Ln)	0.952
d.	Firm revenue (Ln)	0.952
	Percentage variance explained	90.6%
	Cronbach's alpha	0.892

Table 4.9: Descriptive Statistics

	Mean	Min	Max	Std. Dev.	Skew-ness	Kurt-osis
RFMO	2.64	1.00	7.00	1.21	0.73	-0.09
RPE-Use	3.37	1.00	5.00	1.00	-0.85	0.31
Information asymmetry	5.29	2.00	7.00	0.91	-0.61	0.57
Goal ambiguity	2.34	1.00	5.00	0.57	0.43	1.19
Measurability	4.48	2.00	6.00	0.69	-0.45	0.38
Decentralization	4.31	1.00	7.00	1.13	-0.12	-0.02
Business unit size	10.27	2.30	14.97	2.26	-1.56	2.70
Firm size	13.41	3.69	19.41	2.43	-0.44	1.34

as presented in table 4.10. The table shows an insignificant Levene's test for homogeneity of variances, which tests whether the variance in RFMO scores is the same for each of the four sectors. An insignificant Levene's test indicates that the assumption of homogeneity of variance is not violated.

4.4 Analyses

This section describes the analyses. Here I discuss the bivariate correlations and the multivariate analyses. I also analyse the opportunism-mitigating effect of RPE by performing OLS regression analyses. In addition to the opportunism-mitigating effect of RPE, I include a set of factors driving opportunism (*i.e.*, information asymmetry, goal ambiguity, measurability of outputs, and decentralization of decision rights), dummies to control for the sector in which the business unit operates, and size control variables. This results in the statistical model shown in figure 4.3.

Table 4.10: One-way ANOVA of RFMO-levels across sectors

ANOVA between groups	
ANOVA F-statistic	1.706
Prob.	0.17
Levene's test for homogeneity of variances	
Levene's statistic	0.897
Prob.	0.44

Table 4.11: Pearson Correlation Table

	1.	2.	3.	4.	5.
1. RFMO					
2. RPE-Use	.063**				
3. Information asymmetry	.043	-.026			
4. Goal ambiguity	.244***	-.081	-.202***		
5. Measurability	-.278***	.165**	.242***	.586***	
6. Decentralization	.053	-.016	.342***	-.176***	.173***

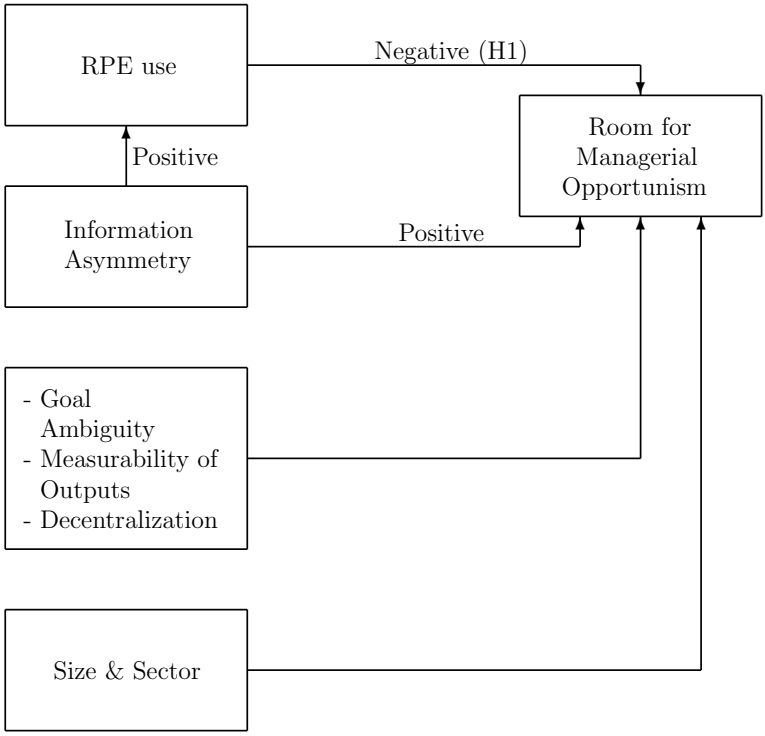
*** Correlation is significant at the 0.01 level (2-tailed).
** Correlation is significant at the 0.05 level (2-tailed).
* Correlation is significant at the 0.10 level (2-tailed).
Listwise N=274

4.4.1 Correlations

Table 4.11 presents the correlation matrix of the variables. The correlation table shows several significant associations between the dependent and independent variables. Contrary to my expectations, RPE-Use is positively correlated with RFMO[10]. Additionally, significant correlations exist between RFMO and goal ambiguity (positive) as well as between RFMO and measurability of outputs (negative). The latter two relations are theory-consistent. Furthermore, a strong and significant correlation between goal ambiguity and measurability of outputs exists. However, this correlation does not cause multi-collinearity, as shown in the next subsection. No significant relations exist between RFMO and information asymmetry, nor between RFMO and decentralization.

[10]In the conclusions and discussion section of this chapter, I seek an explanation for the positive relation between RFMO and RPE-Use.

Figure 4.3: Statistical Model



4.4.2 Multivariate Analyses

In order to test RPE's opportunism-mitigating effect, I analyse the model that is presented in figure 4.3. Two analyses are run: first, the model is tested with the RPE-Use measure. This model is referred to as the base model. Second, an alternative measure for RPE use is used as a robustness check. Both models fit the data, but do not support the research hypothesis; in my models, RPE use does not reduce the room for managerial opportunism. The results of these analyses and several robustness checks are discussed in the remainder of this section.

4.4.2.1 Analyses with RPE-Use Measure

This subsection presents the base model results of the multivariate analysis of the mitigating effect of RPE on the room for managerial opportunism, using the primary RPE-

measure. The multivariate analyses are conducted through structural equation modelling[11] (SEM). The SEM analysis shows overall support for the model. The model explains 14% of the variance in the room for opportunism. The model is presented in table 4.12 - panel A.

The model does not support the research hypothesis. Instead, RPE-Use has a positive effect on the room for managerial opportunism; more RPE use is associated with increased room for opportunism. In the conclusions and discussion section of this chapter, this unexpected finding will be discussed and confronted with prior literature.

The RPE-inducing effect of information asymmetry is also unsupported by the model. This finding is consistent with the results of chapter 2, where information asymmetry (interacting with the level of comparability of the business unit) does not increase RPE use.

Table 4.12: Results of SEM Analysis (RPE-Use)

Panel A - base model results			Coefficient	Std. Error	Critical Ratio	Prob.
RPE use	>	RFMO	0.100	0.058	1.883	0.06'a'
Inf. Asymmetry	>	RFMO	0.080	0.070	1.372	0.09'a'
Inf. Asymmetry	>	RPE use	0.051	0.062	0.901	0.18'a'
Goal Ambiguity	>	RFMO	0.198	0.127	3.137	0.00'a'
Meas. of Outputs	>	RFMO	-0.216	0.110	-3.265	0.00'a'
Decentralization	>	RFMO	0.084	0.056	1.446	0.07'a'
BU size	>	RFMO	-0.021	0.030	-0.339	0.73
Firm size	>	RFMO	-0.017	0.028	-0.266	0.79
Dummy prod.	>	RFMO	0.076	0.162	1.376	0.17
Dummy fin. serv.	>	RFMO	-0.033	0.164	-0.597	0.55
Dummy N.F.P.	>	RFMO	-0.052	0.170	-0.887	0.38

'a': variable based on directional expectation; significance calculated as one-tailed p-value

Panel B - base model fit statistics	
$Chi^2 = 5.656$	*goodness-of-fit statistics*
Degrees of freedom = 8	GFI = 0.997
Probability level = .686	AGFI = 0.973
	CFI = 1.000
Multivariate normality	RMR = 0.021
Kurtosis 13.524 (c.r. 7.017)	RMSEA = 0.000

Panel C - bootstrap results	
Bootstrap samples = 500	Bollen-Stine bootstrap p = 0.701

[11] The SEM analyses are run in AMOS 18.

The model does show significant relations for the antecedents for the room for managerial opportunism. The room for managerial opportunism is influenced by information asymmetry (positively), goal ambiguity (positively), measurability of outputs (negatively), and decentralization of decision rights (positively). The directions of these relations are consistent with my expectations. The size and industry control variables do not have significant effects on the room for managerial opportunism.

To evaluate the fit of the model, I calculated various fit indexes. Assessing the fit of a SEM is not as straightforward as, for example, an OLS regression analysis. The evaluation of the statistical properties of various fit indexes is still an active research topic, and no single measure provides a definite answer with regard to the fit between a model and the data, as a wide variety of goodness-of-fit statistics are available for SEM (Kline 2005). Kline suggests that multiple measures must be assessed to evaluate the goodness-of-fit. As the indicators of model fit, I use the chi^2, the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the comparative fit index (CFI), the root mean square residual (RMR), and the root mean square error of approximation (RMSEA). Kline (2005) suggests that the following factors indicate a good fit: an insignificant chi^2; a GFI, AGFI, and CFI greater than 0.90; a RMR close to 0; and an RMSEA less than 0.045. As presented in table 4.12 - panel B, the indicators show good model fit.

Despite the good fit, the model shows potential non-normality issues. The multivariate kurtosis measure lies above the threshold value of ten suggested by Kline (2005). Kline (2005) argues that kurtosis greater than ten may suggest a problem with the data. The kurtosis in the current model is 13.524. The high kurtosis is due to the inclusion of control variables for business unit and firm size[12]. Because multivariate normality is a strict assumption of maximum likelihood estimation (MLE) in SEM models, I use bootstrapping with 500 samples, which does not assume multivariate normality, to analyse whether the kurtosis affects the MLE estimates (Kline 2005, Bryne 2001). I find that the results are qualitatively similar, and that normality issues did not drive the results. Additionally, the Bollen-Stine bootstrap p does not reject the model. The Bollen-Stine bootstrap is a bootstrap modification of model chi^2 that is used to test model fit by, for example, adjusting for the lack of multivariate normality. Similar to the chi^2, the Bollen-Stine bootstrap is insignificant at p = 0.701, which indicates a fit between the data and the model. The bootstrap statistics are presented in table 4.12 - panel C.

4.4.2.2 Analyses with RPE-based-Targets Measure

To buttress the findings of the base model analyses, I run a robustness check. Here I follow the strategy of chapters 2 and 3 by rerunning the analysis with a different operationalization of RPE use: RPE-based-Targets.

[12]Excluding the size controls from the model reduces the kurtosis to 3.928, but does not yield qualitatively different conclusions.

RPE-based-Targets specifically measures the explicit use of RPE for target-setting purposes (*i.e.*, how peer performance influences the difficulty of the performance standards upon which the manager's performance is evaluated). I measure RPE-based-Targets as the impact of peer performance on the performance target. I ask to what extent the performance of the peers influences the difficulty of the target and/or budget for a variety of performance measures, as described in the appendix (Q2). To interpret the impact of peers on the evaluation as a whole, I weigh the impact of peers on each of the performance metrics by multiplying the scores with their individual importance for the performance evaluation (Q3). RPE-based-Targets does not solely measure *ex-post* peer comparison. Peer performance can also influence the agent's target at the beginning of the period. In this case, the target for the upcoming period is based on the peers' actual performance during the current period. I expect this second RPE measure to correlate with the primary measure 'RPE-Use' because they both include explicit RPE use and apply the peer comparison information *ex-post*. Because of the differences between the measures presented above, the correlation may not be strong. Correlation analysis confirms this inference; the two measures of RPE are correlated with one another but not strongly (0.498; $p < 0.01$).

The analyses with the RPE-based-Targets measure support the main findings of the base model analyses, albeit with a trimmed and marginally significant model. Similar to the base model, the model has a kurtosis that lies above the threshold value of 10 (Kline 2005). In contrast to the previous models described in this thesis, the kurtosis of 14.179 significantly affects the results of the estimated model. The Bolen-Stine bootstrap is significant ($p = 0.025$). This finding indicates that there is no fit between the data and the model. To reduce the non-normality of the data, I excluded the kurtosis-increasing size controls from the analyses. The trimmed model has a kurtosis of 2.783, which indicates that the trimmed model does not suffer from potential non-normality issues.

Both the original and trimmed versions of the robustness check model marginally fit the data. As argued in the previous section, the evaluation of model fit with SEM is not based on a single fit statistic. Relying on multiple statistics can lead to contradictory results, as shown by the current model. The model's chi^2, the adjusted goodness-of-fit index (AGFI), and the root mean square error of approximation (RMSEA) index indicate an inadequate fit between the model and the data. However, the goodness-of-fit index (GFI), the comparative fit index (CFI), and the root mean square residual (RMR) suggest that the model and the data do fit. As a result, no definite conclusion about the goodness of fit can be drawn. Still, because the inconclusive set of fit indexes does not refute the model, the results of this model are presented in this section as another illustration of the relation between RPE use and the room for managerial opportunism.

The robustness check yields results that are largely similar to those of the model with the primary RPE measure (RPE-Use). The current model explains 14% of the variance of RFMO. Similar to the base model, the model with the proxy for the RPE-based-Target does not support the research hypothesis, as the model shows a positive relation between

RPE and RFMO. Moreover, in line with the base model, RFMO is driven by goal ambiguity (positively), measurability of outputs (negatively), and decentralization of decision rights. However, the relation between information asymmetry and RFMO is marginally insignificant (one-tailed p-value = 0.13). The results of the robustness check are summarized in table 4.13.

Table 4.13: Results of SEM Analysis (RPE-based-Targets)

Panel A - base model results			Coefficient	Std. Error	Critical Ratio	Prob.
RPE use	>	RFMO	0.136	0.059	2.422	0.02'a'
Inf. Asymmetry	>	RFMO	0.071	0.075	1.150	0.13'a'
Inf. Asymmetry	>	RPE use	0.007	0.070	0.114	0.46'a'
Goal Ambiguity	>	RFMO	0.184	0.135	2.685	0.00'a'
Meas. of Outputs	>	RFMO	-0.206	0.119	-2.866	0.00'a'
Decentralization	>	RFMO	0.118	0.060	1.921	0.03'a'
Dummy prod.	>	RFMO	0.092	0.170	1.581	0.11
Dummy fin. serv.	>	RFMO	0.002	0.177	0.030	0.98
Dummy N.F.P.	>	RFMO	0.009	0.176	0.143	0.89

'a': variable based on directional expectation; significance calculated as one-tailed p-value

Panel B - base model fit statistics	
Chi^2 = 19.436	*goodness-of-fit statistics*
Degrees of freedom = 6	GFI = 0.985
Probability level = .003	AGFI = 0.888
	CFI = 0.948
Multivariate normality	RMR = 0.030
Kurtosis 2.783 (c.r. 1.643)	RMSEA = 0.090

Panel C - bootstrap results	
Bootstrap samples = 500	Bollen-Stine bootstrap p = 0.025

4.4.2.3 Endogeneity Analysis

Similar to the previous chapter, this chapter's dependent variable may be jointly determined with one of the independent variables. This relation exists between room for managerial opportunism and RPE use because RPE use can reduce the room for managerial opportunism, but increased room for opportunism can also induce organizational reliance on RPE. As described in chapter 3, this possibility raises endogeneity concerns. Endogeneity is a key limitation in most empirical research (Ittner & Larcker 2001). Endogeneity is

an econometric problem that leads to biased and inconsistent estimators within equations used to test theoretical propositions. This bias renders inferences problematic and consequently reduces the confidence in the conclusions drawn from the research (Chenhall & Moers 2007). Thus, endogeneity can cause the null-findings of the presented analyses. I partially mitigate this problem in the models presented in the preceding sections by including the effect of information asymmetry, mediated by RPE use, on room for managerial opportunism, because information asymmetry can influence both RPE use and room for managerial opportunism simultaneously. Additionally, I conduct a robustness check on the models analyses to address the issue of potential remaining endogeneity with a Hausman test. Endogeneity is assessed for both RPE measures (RPE-Use and RPE-based-Targets). However, as the procedure and the results are similar for both RPE measures, only the test for the primary RPE measure (RPE-Use) is described in this section.

Antonakis *et al.* (2010) recommend the Hausman test to assess endogeneity problems in SEM. The Hausman statistic is computed by conducting two OLS regressions after one another. The first regression predicts RPE-Use based on one or more instrumental variables, and all other variables in the original equation. The instrumental variable(s) should correlate with RPE-Use, but not with the error term of the structural model (Larcker & Rusticus 2010). The second regression includes the original (observed) RPE-Use measure, the predicted RPE-Use measure from the first stage and the original independent variables. The Hausman test assesses the significance of the predicted RPE-Use in the second regression.

Proper instrumental variables should be both *relevant* (*i.e.*, they should be related to the endogenous explanatory variable RPE use) and *exogenous* (*i.e.*, they should be unrelated to the undefined residual of room for managerial opportunism). In my analyses, I select one instrumental variable from the questionnaire that I associate with RPE use. This item asks to what extent peer performance is a point of reference for the respondent's superior when evaluating the HR function in the respondent's business unit. Whereas RPE use refers to the performance evaluation of the business unit manager, the questionnaire item focuses on the general assessment of a smaller organizational part. Still, both RPE use and the selected questionnaire item relate to some kind of peer comparison for the evaluation of organizational parts. Therefore, I believe that the selected questionnaire item is related to RPE use and that it is a relevant instrumental variable.

Arguing why an instrumental variable is exogenous, is, however, problematic in most research, because it requires one to demonstrate that an instrumental variable is not associated with the undefined structural residual. However, what researchers can do instead is to try to find reasons that falsify this association's non-existence. To the best of my ability, I could not find a theoretical argument for how the use of peer performance as a point of reference for the evaluation of the HR function is related to the structural residual of

room for managerial opportunims (*i.e.*, the part of RFMO that is not explained by other exogenous factors in the structural model).

Larcker & Rusticus (2010) and Chenhall & Moers (2007) argue that, in addition to providing (at least a beginning of) theoretical argumentation for the instrument's relevance, a statistical specification test is required. The specification test for the instrumental variable's relevance is the partial F-test[13], which examines whether the instrument is significantly related to the endogenous explanatory variable. This test is conducted by evaluating the partial F-statistic of the first-stage regression analyses. This partial F-statistic indicates if the addition of the instrumental variable contributes to the model. Given the fact that I use a single instrumental variable, a partial F-statistic value of above 8.96 suggests that the instrumental variable is relevant (Larcker & Rusticus 2010). The first-stage F-statistic in my model is 40.227 ($p < 0.01$). Considering both the theoretical and statisical assessment of the instrumental variable, the selected questionnaire item is used as a single-item instrumental variable.

The first-stage regression analysis is significant (ANOVA F-statistic is 4.784). This regression explains 14.0% of the variance of RPE-Use (Adjusted $R^2 = 11.0$). Additionally, the instrumental variable is significant. Following the recommendations of Larcker & Rusticus (2010), I present the coefficient estimates for the first-stage model (see table 4.14).

Table 4.14: Results of First-Stage Regression Analysis (RPE-Use)

	Coefficient	Std. Error	t-Stat.	Prob.
Constant	1.200	0.766	1.568	0.12
Instrumental Variable	0.192	0.030	6.342	0.00
Information Asymm.	0.053	0.065	0.807	0.42
Goal Ambiguity	0.134	0.118	1.136	0.26
Measurability of Outputs	0.147	0.102	1.435	0.15
Decentralization	-0.002	0.052	-0.044	0.97
BU Size	0.005	0.028	0.179	0.86
Firm Size	0.007	0.026	0.246	0.81
Dummy production	0.138	0.150	0.915	0.36
Dummy financial services	0.091	0.152	0.599	0.55
Dummy not-for-profit	-0.047	0.159	-0.299	0.77
$R^2 = 0.140$			F-statistic = 4.784	
Adjusted $R^2 = 0.110$			Prob(F-statistic) = 0.000	

Included observations: 306

The predicted value of RPE-Use that follows from this regression is then used as a proxy for RPE-Use in the second regression. In addition to the predicted RPE-Use variable from the first regression, the second regression includes the original (observed) RPE-Use

[13]Larcker & Rusticus (2010) also refer to this test as the 'weak instruments' test.

measure, and the original independent variables (information asymmetry, goal ambiguity, measurability of outputs, decentralization of decision rights, and the control variables). The second regression is also significant (ANOVA F-statistic is 4.647). The model explains 13.8% of the variance in RFMO (Adjusted R^2 = 10.5%).

Finally, I conduct the Hausman test for endogeneity to assess the existence of an endogeneity problem. I follow Larcker & Rusticus' (2010) explanation, which argues that if the coefficient for the predicted RPE-Use is significant, the Hausman test rejects the null of no endogeneity problem. In my second regression, as shown in table 4.15, the predicted RPE-Use is insignificant (p = 0.244). This finding implies that the model does not suffer from endogeneity problems[14]. This finding reinforces the results of the previous analyses.

Table 4.15: Results of Second-Stage Regression Analysis (RPE-Use)

	Coefficient	Std. Error	t-Stat.	Prob.
Constant	2.651	0.895	2.963	0.00
Unstandardized Predicted Value	-0.216	0.185	-1.168	0.24
RPE use	0.134	0.064	2.089	0.04
Information Asymm.	0.107	0.073	1.474	0.14
Goal Ambiguity	0.417	0.130	3.202	0.00
Measurability of Outputs	-0.318	0.118	-2.702	0.01
Decentralization	0.076	0.057	1.318	0.19
BU Size	-0.008	0.031	-0.248	0.80
Firm Size	-0.007	0.029	-0.244	0.81
Dummy production	0.232	0.165	1.406	0.16
Dummy financial services	-0.086	0.168	-0.513	0.61
Dummy not-for-profit	-0.169	0.174	-0.966	0.34
R^2 = 0.115			F-statistic = 2.844	
Adjusted R^2 = 0.074			Prob(F-statistic) = 0.001	

Included observations: 276

4.4.3 Summary of the Findings

Similar to the overall findings in chapter 2, the analyses in the current chapter do not support the claim that RPE mitigates opportunism. Based on Murphy (2001), this chapter

[14]The absence of the endogeneity problem does not mean that RPE use and noise in the performance evaluation are not determined simultaneously. The Hausman test results do not imply that the dependent and independent variables are not interrelated; the test only suggests that the model estimation is unbiased (Abernethy *et al.* 2004).

argues that RPE reduces the room for managerial opportunism by placing the standard-setting process outside the sphere of managerial influence. This study expected but did not find a negative relation between the use of RPE and the room that managers have to behave opportunistically. Instead, I find a positive relation between RPE-Use and RFMO. This finding is discussed in the next section.

In addition to this unexpected result, the analyses do not support a RPE-increasing effect of information asymmetry. Although a positive relation was expected, the null result conforms to the findings of chapter 2. In chapter 2, increased levels of information asymmetry (moderated by the comparability of the business unit) did not result in more RPE use.

Moreover, the analyses provide empirical evidence suggesting that a number of antecedents drive opportunistic behaviour. These antecedents include information asymmetry, goal ambiguity, measurability of the outputs, and decentralization of decision rights. These findings are largely robust over both analyses and may provide useful starting points for future research on opportunism-mitigating control instruments.

4.5 Conclusions and Discussion

This chapter studied the supposed opportunism-mitigating effect of RPE. Murphy (2001) claims that RPE reduces managerial opportunism by preventing managers from opportunistically influencing their own performance targets. The tendency to behave opportunistically is a central tenet in economics-based theorizing on Management Control. Opportunistic behaviour can significantly impact organizations. For example, if the members of the organization seek their own self-interest by shirking their duties or committing fraud, their behaviour will reduce the organizational gains. Additionally, opportunistically lowering the performance targets harms the organization's ability to realize its objectives. Although the literature is concerned with opportunism, few empirical studies have directly assessed opportunism (Macher & Richman 2008). The current chapter fills this gap by empirically studying opportunism and the extent to which RPE reduces opportunistic behaviour.

According to Murphy (2001), RPE reduces the room for managerial opportunism by determining the performance target outside the sphere of managerial influence. Doing so makes it more difficult for the manager to lower his target difficulty. This 'external determination' renders an RPE-based performance target less vulnerable to managerial opportunism than the performance targets that are determined within the reach of the employee. Murphy's

opportunism perspective provides a promising theoretical explanation for RPE use. However, the empirical support for this explanation of RPE is scarce. As the first study to my knowledge, chapter 2 of my thesis seeks empirical support for the opportunism-mitigation perspective on RPE at the level of business unit managers. Overall, the analyses in chapter 2 show no robust support for the opportunism-mitigation perspective on RPE. This finding may be the result of the limited added value of RPE to the reduction of opportunism at the lower echelons of the organization. However, the limited support may also be due to the model specification, which may reduce the statistical power of the analyses. The model specification in chapter 2 is needed to test the chapter's hypotheses. However, the current chapter takes a different approach. This chapter allows for more powerful and direct tests of the opportunism-mitigating explanation of RPE theory.

The current chapter develops a model that assesses whether organizational reliance on RPE mitigates the room for managerial opportunism. For the empirical analysis, I use a structural equation model to analyse survey data on 325 business unit managers in the Netherlands. Overall, the analyses yield no support for the opportunism-mitigation explanation of RPE theory. Rather, the findings indicate that using RPE increases the room for managerial opportunism. This result is robust over two operationalizations of RPE use. Similar to chapters 2 and 3, this chapter relies on a broad, unspecific measure of RPE use and on a measure that focuses on the explicit application of RPE to performance targets. Both measures yield qualitatively and quantitatively similar results. The structural equation models explain 14% of the variance in the dependent variable (*i.e.*, 'room for managerial opportunism').

A first explanation for not finding a negative effect of RPE use on the room for managerial opportunism is found in the work of Davila & Penalva (2006) and Ittner *et al.* (1997). Davila & Penalva (2006) describe how managers can influence the composition and weighting of their own performance measurement system. Managers self-select a set of performance targets that is highly controllable. Ittner *et al.* (1997) argue that managers potentially influence the composition of their performance measurement systems to obtain more manipulable performance measures. If managers can influence their performance measurement systems, they can also influence their performance standards. As a result, a manager can choose to be evaluated by those measures on which the manager compares favourably with his peers. For example, a manager's business unit may consistently score high on customer satisfaction and revenues but score poorly on costs. My data support this observation by Ittner *et al.* (1997). My respondents claim that they can negotiate or otherwise influence their performance criteria to, on average, some extent. The manager's influence on the performance criteria selection is positively related to the use of RPE for explicit target setting (*i.e.*, RPE-based-Targets), but not with less specific applications of RPE (*i.e.*, measured with RPE-Use)[15].

[15]The respondent's influence on the selection of the performance criteria is positivelly correlated with RPE-based-Targets (0.148, $p < 0.05$), but not with RPE-Use (0.022, $p = 0.70$).

By self-selecting the performance metrics (or the weights placed on the individual metrics in the performance measurement system), the manager bypasses the opportunism-mitigating properties of RPE. This post-hoc explanation suggests that RPE use may not reduce opportunism. However, it only partially explains the opportunism-*increasing* effect of RPE use found by this study. This positive relation between RPE use and opportunism requires an additional explanation, which may be found in the risk of collusion amongst agents[16]. Holmstrom (1982) argues that the possibility of collaboration among peers provides the evaluated employee with the opportunity to influence his performance target in his favour. If RPE is implemented in a manner that has a high risk of collusion, RPE use may actually increase the room for managerial opportunism. Following Gibbons & Murphy (1990), I argue that this case will seem increasingly possible if the peer group consists of other business units within the same organization. The data in my study provide some limited information about the composition of the peer group. In chapter 2 of this thesis, table 2.9 (page 40) shows that the majority of the RPE adopters rely on internal peers or both internal and external peers to determine their performance targets. The limited use of external peer groups renders RPE vulnerable to collusion amongst agents[17]. Doing so can increase (instead of decrease) the room that managers have to behave opportunistically.

Another empirical result concerns the room that managers have to engage in opportunistic behaviour under their installed control system. On average, the respondents of my study claimed to have limited room for opportunism. This result holds for both RPE adopters and non-users of RPE. However, the experimental literature supports the idea that self-interest seeking behaviour is highly prevalent amongst individuals. This notion may imply that, on average, the control systems within organizations are effective at reducing the room for opportunism either via RPE or via other control instruments.

As with any study, this research is subject to a number of caveats. The two most important limitations of this study are the following: 1) the fact that I focus on an isolated control instrument instead of taking a more holistic approach, and 2) the absence of information on the risk of collaboration and opportunistic peer group compositions. As I focus on a single control instrument, I do not have data on other potential control instruments that are used to reduce the room for managerial opportunism. This may affect the results of this study to some extent. For example, I did not gather data on other means of performance standard determination. Successfully using other externally determined performance targets to mitigate opportunism can affect the statistical relation between RPE and the room for managerial opportunism.

[16]Collusive behaviour amongst agents is documented by Towry (2003) and Zhang (2008). Both studies find in their experiments a positive relation between the possibility of agents to communicate amongst each other and the forming of collusive agreements. The possibility to communicate with other agents is especially relevant within a situation of internal peer comparison instead of with external peers.

[17]The consequences of using external peers to form a reference group is further explored in section 5.1.

With respect to the second limitation, it is possible that the unexpected opportunism-increasing effect of RPE use is due to the manner in which RPE is implemented in the control system, as suggested by the post-hoc explanation presented above. If the peer comparison is based on a reference group with whom the evaluated manager can collude, there is more room for opportunism. Because of the questionnaire design, I do not have sufficient data to further analyse this effect. More sophisticated measurement of the peer group composition would help to control for this effect.

Despite these limitations, I think that this study provides valuable empirical insights on the opportunism-mitigating properties of RPE at the level of business unit managers in addition to the findings in chapter 2 of this thesis. Whereas the mixed results of chapter 2 may be due to limited statistical power, the more direct analyses in the current chapter provide a clearer and more robust picture of the empirical relevance of Murphy's (2001) opportunism-mitigation explanation of RPE use. At least at the level of business unit managers, RPE does not seem to reduce the room for opportunism. Nonetheless, RPE may help to mitigate opportunism if the reference group is not composed of internal peers. This thesis leaves the effects of peer group composition to future research.

Chapter 5

Conclusion

5.1 Summary, Conclusions and Discussion

This thesis set out to take a closer theoretical and empirical look at Relative Performance Evaluation (RPE). RPE is a means to evaluate an employee's performance by comparing his performance with that of a reference group that consists of individuals facing similar tasks and circumstances. The performance of the reference group functions as a performance benchmark for the evaluated employee. The performance of the reference group constitutes an (explicit or implicit) performance target, to which the employee's actual performance is evaluated.

The resulting performance target incorporates external circumstances. If the performance target is set in absolute terms instead of relative to peer performance, external circumstances can influence how difficult it is for the employee to reach his performance target. A windfall loss would make achieving the preset absolute performance target unlikely, if not impossible. If the employee receives a (financial or non-financial) reward upon achieving a performance target, a windfall loss also reduces the employee's reward. This outcome is undesirable to the extent that targets and rewards can motivate the employee to maximize his efforts. If external circumstances reduce the achievability of the performance target, the target does not optimally motivate the employee. The opposite situation is also possible. If external circumstances render the performance target easy to achieve, the target does not optimally motivate the employee because easy targets do not require him to maximize his effort. In this case, even with relatively low effort levels, the target can be achieved. In both cases, external circumstances (positively or negatively) limit the organization's ability to realize its objectives because of the employee's reduced effort.

Although RPE also has other benefits, scholars have primarily described RPE in terms of its ability to reduce the problem of noise in the performance evaluation. The external circumstances described above induce noise in the performance evaluation. This noise reduces the quality of the performance evaluation, which aims to reward the agent for his effort but not for uncontrollable external circumstances. Reasoning from an analytical agency theoretical tradition, Holmstrom (1982) proves that RPE reduces noise in the performance evaluation by incorporating information about peer performance into the performance contract between the superior and the employee. The reference group's performance is informative because the group is exposed to the same uncontrollable events as the employee. Both the employee's organization and his peers' businesses may experience the same windfall losses (or gains) such that everyone experiences lower (or higher) performance. Comparing the performance of the agent with the reference group partially filters out the effects of external uncontrollable factors. By reducing noise, RPE improves the quality of the performance evaluation.

The second benefit that the literature ascribes to RPE lies in its opportunism-mitigating characteristics. This opportunism-mitigation perspective on RPE relates to the way in which performance standards are determined. Performance targets can be determined in two main ways. First, they can be set based on internal information by, for example, using prior-year performance or through internal negotiations between higher and lower management. According to Murphy (2001), this 'internal' determination of the performance target has an important drawback: it creates room for the evaluated employee to opportunistically influence his performance target. For example, if standards are based on prior-year performance, employees have an incentive to achieve their performance targets but to avoid outperforming them. Outperforming their performance target will result in an increased standard for the next period (Murphy 2001).

Performance targets can also be determined 'externally' (*i.e.*, outside the evaluated employee's sphere of influence) (Murphy 2001). External determination implies that the performance target is determined in such a way that the difficulty of the performance standard cannot be easily affected by the evaluated employee. For example, the target may be based on market conditions or peer performance. RPE is an example of an externally determined performance standard because it incorporates information about the performance of an external reference group. According to Murphy, the performance of a peer group lies outside the employee's sphere of influence. RPE-based targets cannot be easily manipulated by the evaluated employee. These considerations suggest that the room for opportunistic behaviour can be delimited by incorporating the performance of a reference group into the compensation plan and that RPE can aid the standard-setting process. Throughout the thesis, I explain and test the use of RPE with the two perspectives presented above.

Despite the compelling theoretical basis of RPE, the literature is inconclusive with regard to the empirical validity of RPE. Whereas the analytical research has found strong evidence for RPE (Holmstrom 1982), the empirical literature has not generally reported substantive organizational reliance on RPE. Triggered by this puzzle, my dissertation research approaches RPE from a partially different angle. Whereas most prior studies have analysed RPE solely from a noise-reducing perspective, this thesis also explains RPE from an opportunism-mitigation viewpoint. In addition, whereas prior research has almost exclusively studied RPE by using archival data of RPE for executive compensation, the current research project uses in-depth survey data on RPE use *within* organizations at the level of business unit managers. The remainder of this section presents the findings of my thesis, organized into four topics. First, I discuss the prevalence of RPE amongst business unit managers. The second and third topic address the extent to which we can understand the use of RPE at this level with noise- and opportunism-reduction explanations. The final topic refers to the possibility of combining these explanations to further our understanding of RPE in practice.

1. *RPE use amongst business unit managers*

   The first main finding of my research concerns the empirical puzzle presented earlier, which is that RPE seems to be theoretically and analytically sensible for business practice but lacks empirical support. The first topic that my thesis studies, is to what extent RPE is used for the performance evaluation of business unit managers.

   Overall, the body of RPE research on executive compensation suggests that RPE might not be widely used in practice. However, in chapter 2 of this dissertation, I report support for RPE use at the lower echelons of organizations. I find that RPE is widely used to evaluate business unit managers. More than half of the participants in my sample claim to use RPE to a large or very large extent. I observe that 88% of the business units in the sample use RPE to at least a small extent and that only approximately 8% of the respondents do not use RPE at all.

   The widespread organizational reliance on RPE seemingly contradicts the existing literature, which does not report large-scale RPE use. This difference is probably due to the level of analysis and type of data used for my research. My study focuses on the business unit level, where RPE may be more prevalent than at the executive level. Garvey & Milbourn (2003) argue that RPE may be inefficient if it is used to evaluate (relatively wealthy) executives because executives can more efficiently reduce noise in their incentive contracts by directly adjusting their own private investment portfolios. However, personal hedging is likely to be too costly for less wealthy employees. These high costs of hedging may increase the attractiveness of RPE to people working at the lower echelons, such as business unit managers.

Additionally, the type of data that my study uses may also have influenced the findings concerning RPE's prevalence. Most empirical studies that are currently available in the literature rely on archival data from before 2006, when the SEC did not yet require firms to disclose detailed information about their compensation plans. These data have the drawback of requiring indirect instead of direct analyses. Indirect analyses study RPE *as if* a manager's performance is evaluated relative to his competitors' performance instead of explicitly studying whether the performance evaluation system incorporates peer performance information. In contrast to most prior studies, this thesis uses detailed survey data on the performance evaluation praxis of 325 business unit managers. This survey instrument was specifically designed for the current study. This instrument provided the analyses with more detail and richness such that the study produced a more accurate description of RPE in practice.

Studying RPE at a lower organizational level by using detailed survey data may generate much more evidence on RPE use than prior studies. Therefore, I conclude that my findings concerning RPE adoption do not contradict the existing knowledge base on RPE but rather extend it to a different organizational level. This finding furthers our knowledge about the empirical validity of RPE.

2. *RPE use and the reduction of noise*

After the exploration of the prevalence of RPE at the level of business unit managers, the next goal of my research is to *explain* the use of RPE. This study adopts two explanations for RPE use. As presented earlier in this chapter, the first explanation is based on the notion that RPE use reduces noise by partially insulating the performance evaluation from the effects of external circumstances. This perspective on RPE is the primary explanation for RPE use found in the literature.

Chapter 2 of this thesis explains RPE use by studying factors that antecede RPE use. The first antecedent described in chapter 2 is common uncertainty, which follows from the noise-reduction perspective. Common uncertainty refers to uncertainties in the agent's external environment and his business unit that are common amongst his reference group. Uncertain events induce noise in the performance evaluation. RPE can reduce the effect of external uncertain events for as far as these events are common amongst the reference group (see, for example, Holmstrom 1982). The analyses conducted in chapter 2 robustly support the effect of this RPE driver by showing that common uncertainty positively influences the use of RPE. This finding suggests that RPE use can be explained from a noise-reduction perspective.

In chapter 3, I further analyse the noise-reduction effect of RPE by examining the consequences of using RPE. Whereas chapter 2 explains RPE use with assumed noise-reducing effects of RPE, chapter 3 analyzes this explanation in a more direct manner. Chapter 3 studies whether organizations that use RPE experience reduced levels of noise in the performance evaluation. This question is important because the results from chapter 2 may be driven by alternative explanations. The literature provides various rival explanations for the association between uncertainty and RPE use. These alternative explanations include the increased need for internal competition (*i.e.*, by pitting employees against each other through evaluating their performance relative to each other) and the desire to stimulate organizational learning via the benchmarking of performance. The results of chapter 3 reaffirm the findings of chapter 2 concerning noise reduction. The analyses show that RPE use yields reduced levels of noise in the performance evaluation. This finding strengthens the conclusion that RPE effectively reduces noise in the performance evaluation of business unit managers.

3. *RPE use and the reduction of room for managerial opportunism*

The third research topic concerns the opportunism-mitigation perspective on RPE, which is based on Murphy (2001). According to this perspective, RPE can reduce the room that managers have to opportunistically influence their performance target difficulty, by placing the target determination process outside their sphere of influence (see: Murphy 2001). This perspective is addressed in chapters 2 and 4.

As discussed in the previous section, chapter 2 explains RPE use with factors that antecede RPE use. In addition to the measure of common uncertainty that drives noise, chapter 2 includes an antecedent that is based on the opportunism-mitigation perspective. This antecedent is information asymmetry, which is combined with business unit comparability. Information asymmetry increases the room for the manager to behave opportunistically. I argue that the opportunism-increasing effect of information asymmetry can only be solved by RPE if the business unit is sufficiently comparable to a peer group[1]. However, the analyses do not show robust support for the opportunism-mitigating effect of RPE. Only one of two models supports the joint effects of information asymmetry and business unit comparability. However, this support is only marginally significant.

The opportunism-mitigation explanation receives additional empirical attention in chapter 4. Similar to the approach of chapter 3, the fourth chapter of my thesis

[1]If a business unit is not comparable to a relevant peer group because the unit produces highly specific products or services or relies heavily on specific processes, the unit's performance cannot be assessed by the performance of its peers. In this case, peer comparison is not informative about the performance target.

addresses the RPE claim in a more direct manner than chapter 2. The models in chapters 3 and 4 focus on testing a single explanation instead of including both perspectives simultaneously. The benefit of the analyses in chapters 3 and 4 lies in their statistical power. The analyses are statistically more powerful because they do not include or require statistically demanding interaction effects. Increasing statistical power is especially important for the opportunism perspective because limited statistical power could have driven the marginal findings in chapter 2.

However, chapter 4 confirms the conclusions of chapter 2 with regard to the opportunism-mitigating properties of RPE. Similar to chapter 2, the analyses in chapter 4 show that RPE has no opportunism-mitigating effect for business unit managers. Instead, the analyses of chapter 4 indicate that using RPE *increases* (instead of *decreases*) the room that managers have to behave opportunistically.

A partial, post-hoc explanation for why RPE does not reduce the room for managerial opportunism is found in the work of Davila & Penalva (2006) and Ittner *et al.* (1997). According to Davila & Penalva, managers can influence the composition and weighting of their own performance measurement system. Managers self-select a set of performance targets that is highly controllable. Ittner *et al.* argue that managers may influence the composition of their PMS to obtain more manipulable performance measures. If managers can influence their performance measurement system, they can also influence their performance standard. Thus, managers can choose those measures on which they compare favourably with their peers. By self-selecting the performance metrics (or the weight placed on the individual metrics in the performance measurement system), the manager bypasses the opportunism-mitigating properties of RPE. My data support this observation by Ittner *et al.* (1997). In response to a question not used earlier in this study, my respondents claim that they can negotiate or otherwise influence their performance criteria to, on average, some extent. Moreover, this influence increases with increasing use of RPE for explicit target setting purposes. My data show no association between less specific applications of RPE and the possibility to influence the set of performance criteria.

Although this explanation reasons why RPE may not be an efficient control instrument for opportunism reduction, it only partially explains the *positive* effect of RPE on the room for managerial opportunism. From the analyses in chapter 2, we learn that organizations implement RPE into the performance evaluation praxis of business unit managers mainly by relying on internal and mixed peer groups. Within my sample, only 8% of the RPE users rely exclusively on external peers, whereas the remaining 92% use internal peers at least to some extent to determine the performance targets. However, using internal peers may violate Murphy's (2001) 'influenceability' presumption. The reason why RPE may reduce opportunism because

the manager cannot easily influence his own performance target, but, as Gibbons & Murphy (1990) argue colluding with other managers within the organization to opportunistically lower the performance target may not be difficult [2].

Based on the overall findings of chapters 2 & 4, I conclude that RPE, as it is implemented in practice, is not an effective means to reduce room for opportunistic behaviour amongst business unit managers. However, the fact that RPE is not often used to mitigate opportunism by comparing managerial performance to that of external peers exclusively, does not falsify Murphy's claim that RPE can reduce opportunism. The use of internal peers is not a prerequisite for RPE. Rather, RPE as described in the executive compensation literature, most likely is based on performance information of external peers. After all, top managers do not have internal peers, whereas business unit managers do. Business unit managers may be comparable to other business unit managers within the organization. Notwithstanding my findings, RPE may still be an effective means of mitigating the room for opportunism amongst business unit managers if the peer comparison is based on an external reference group. Unfortunately, my data do not allow empirical analyses of this idea[3], as will be discussed in the limitations section at the end of this chapter.

4. *RPE use for both purposes simultaneously*

In the previous paragraph, we saw that, although many organizations use RPE, it is, on average, not used to mitigate opportunistic behaviour amongst business unit managers. Organizations may not use RPE for this purpose because in most cases, RPE adopters rely on reference groups containing internal peers. As argued previously, internal and mixed peer groups may not facilitate opportunism mitigation, at least not as well as external peer groups. This observation raises the question of why organizations, after investing in RPE as a control instrument, do not simply use external instead of internal peer groups and use RPE also to mitigate opportunism.

Here, I present two post-hoc explanations for this phenomenon. These explanations refer to cost-benefit trade-offs and to control instrument design considerations, respectively. First, the respondents report low room for managerial opportunism regardless of whether they use RPE. This finding, which was discussed in chapter 4, suggests that, on average, the respondents do not need (additional) control instruments to reduce opportunistic behaviour. Thus, the benefits of using RPE to reduce

[2]The risk of collusion is documented in the analytical agency literature (see: Holmstrom 1982) and in experimental studies (see: Towry (2003) and Zhang (2008).

[3]An independent sample T-test shows no significant difference in the mean levels of the room for managerial opportunism between the RPE users who rely exclusively on external peers and the RPE adopters who use mixed or purely internal peers. However, this null result can easily be driven by the limited subsample size of the 'external peers only' group (N=21).

opportunism may be limited. Additionally, using external instead of internal peers may incur significant costs. Making a comparison to a reference group requires performance data about peers. Internally, these data may be freely available or available at relatively low costs. However, externally, one needs to acquire data on a specific peer group. General performance information about competitors or the industry as a whole may be available for free. However, if a comparison to more specific measures at the business unit level or lower levels is preferred, these data become increasingly costly to obtain.

My data show that easily available company data only accounts for 16% of the information used in performance evaluations. In contrast, 61% of the information used in the performance evaluation of business unit managers consists of disaggregated or non-financial performance metrics at or below the business unit level. If these metrics are to be compared with those of external peers, significant costs would be involved to obtain this specific information. Overall, trading off the costs and benefits may result in not using RPE for opportunism-mitigation purposes.

However, it is entirely possible that if RPE is already used for the purpose of noise reduction, the additional use of RPE for opportunism-reduction purposes is not a matter of cost-efficiency. Instead, simultaneously using RPE for both purposes may be highly undesirable because the two purposes may have contradictory design requirements. This contradiction may be due to the nature of the preferred peer group composition. Whereas the use of RPE for opportunism-mitigation purposes may benefit mostly from an external peer group, the opposite may be true for noise-reduction purposes. To reduce noise, using an internal peer group would be most effective because of the *relevance* of the peer group's performance. The business unit may simply have more in common with its internal peers than with its competitors (*e.g.*, in terms of the back office, the costs of distribution or other organization-wide similarities), than with external peers who may organize their business differently. This consideration suggests that an organization has to choose whether it wants to incorporate peer information into the performance evaluation of its managers for noise- or for opportunism-reduction purposes but not for both.

5.2 Implications for Practice

A well-designed Management Control System is important for organizations because it helps to realize organizational goals. Literature claims that effective control system design consists of multiple dimensions. These dimensions include: 1) the choice of relevant performance metrics; 2) the strength of the relation between performing well on these metrics

and the compensation; and 3) the determination of the relevant difficulty of the performance target (*i.e.*, what performance level qualifies as 'good' performance?). RPE helps to determine the difficulty of the performance targets and/or budgets to which managers are held accountable. By providing organizations with a means to set the performance standards, RPE plays an important role in the design of the Management Control System.

This study shows that RPE is an effective and widely used praxis for evaluating business unit managers. RPE improves the quality of the performance evaluation if common external events affect the manager's measured performance. Reducing noise improves the contractual efficiency between the business unit manager and his superior. This improvement has important consequences for the motivation of the business unit manager. In turn, increased motivation can help the organization realize its goals.

If organizations wish to use relative performance targets, it is important to consider the composition of the peer group. Most organizations in my sample compare the performance of the evaluated manager to a group of internal peers. This has several benefits. Firstly, detailed performance data on internal peers is mostly available at low costs. Secondly, internal peers face, in general, highly similar external events. In this sense, the comparison to internal peers is highly efficient to reduce noise. However, using internal peers is probably not an effective option for another application of RPE use, namely the reduction of managerial opportunism.

Organizations may wish to use RPE to mitigate managerial opportunism. However, as it is currently implemented throughout my sample, RPE does not seem to be a suitable means of reducing the room that business unit managers have to engage in opportunistic behaviour. This is probably because of the fact that organizations primarily rely on internal peer groups. If organizations aim to use RPE for opportunism-reduction, the use of internal peers does not seem to be an effective design choice. This finding does not mean that the RPE cannot help organizations reduce the opportunism of their employees. In fact, RPE may still be effective in opportunism-mitigation if designed differently. My suggestion is that organizations use a benchmark peer group that cannot be easily influenced by the evaluated manager, by, for example, comparing the manager's performance to the performance of a group of competitors outside the organization. This use of RPE is similar to the application of RPE amongst top executives, where the effectiveness of RPE's opportunism-reduction properties is supported by empirical literature.

These considerations suggest that, when implementing RPE use in the Management Control System, organizations are well advised to make an *a priori* decision on wether they want to use RPE for the reduction of noise, or the mitigation of managerial opportunistic behaviour.

5.3 Limitations and Suggestions for Further Research

As any study, the research presented in this thesis is subject to a number of limitations. In this subsection, I discuss the limitations that I consider to be the most important for my study. These limitations include the drawback of studying control instruments in isolation, the potential issues caused by using survey data in terms of causality, and the quality of the measurement. These limitations and their consequences are discussed in the remainder of this chapter.

5.3.1 Studying Control Instruments in Isolation

Although RPE is part of a Management Control System that consists of many control instruments, this research project studies RPE as an isolated control instrument. I do so because of the design of the study in general and of the questionnaire in particular. Because I focus on a single control instrument, I do not have data (or theory) on other potential control instruments that are used, for example, to reduce noise in the performance evaluation or mitigate the room for managerial opportunism. This lack of data may affect the results of this study. According to Chenhall (2003), one drawback of studying the specific elements of a control system in isolation from the other organizational controls potentially leads to model underspecification.

Although I acknowledge Chenhall's warning, I deem it necessary to study this specific instrument in isolation first before taking on a more holistic approach. My study is one of the first to investigate RPE at the lower echelons. Before this study, we had little empirical understanding of how RPE works in the evaluation and compensation praxis of business unit managers. A focused study that analyses a single control instrument can invest the ample space and time that is required to approach the instrument with multiple measures and perspectives. Doing so is necessary to obtain a deeper understanding of the subject matter, which will, for example, allow one to establish measures that can be incorporated into other studies with broader scopes. Nonetheless, studying the interaction of RPE with other control instruments within the control system is an important next step to understanding RPE. This analysis may shed more light on the theories adopted throughout this thesis. This thesis leaves the holistic study of the role of RPE in the control system to future research.

5.3.2 Causality

The second limitation arises from the fact that I use cross-sectional data for my analyses, which measure the independent and dependent variables at one point in time. Although this type of quantitative research has several benefits (*e.g.*, relatively large-scale datasets, which increases the generalizability of the results), the method has drawbacks as well. With this method, I cannot fully test the causal character of the relations. A common trap in interpreting statistical results is to claim causation when the empirical methodology supports correlation (Echambadi *et al.* 2006:1811). I stress here that I theorize causation but primarily test associations. Research methods that incorporate longitudinal data (*e.g.*, longitudinal case studies) are better equipped to determine causality.

Nonetheless, I take steps to detect potential problems with causality. In chapters 3 and 4, I test for the endogeneity of the independent variables. Endogeneity occurs if the independent variable included in the model is correlated with the error terms of the estimated model. Sources of endogeneity include reverse causality, simultaneous causality, and omitted variables. Ignoring endogeneity may lead to biased and inconsistent estimates (Echambadi *et al.* 2006:1805). Hausman test results indicate that endogeneity is not a major concern in my models and that the results from my analyses are not biased by endogeneity. However, the endogeneity analysis cannot positively rule out reverse causality. Future research may wish to rely on panel data or design experiments to fully rule out reverse causality (Echambadi *et al.* 2006:1805).

5.3.3 Measurement

The final limitation that I discuss here refers to the measurement of my variables. This thesis relies on an extensive questionnaire. I invested considerable time and effort into the survey instrument. For example, I conducted several rounds of pretesting among the potential respondents. Doing so certainly improved the quality of the data that were obtained for this study. However, with the benefit of hindsight, several measures leave room for improvement. Most importantly, the RPE proxies that are used in this study can be improved. I use two different RPE measures that have specific strengths and weaknesses. Looking back, I should have developed an overall measure to function as a safety net under the measures that investigate specific manifestations of RPE. If I had developed this overall RPE measure, I would have invested in the other specific RPE measures such that they would show less overlap with each other and only overlap with the overall measure.

In addition to the RPE measures, other measures can be improved as well. For example, the measure for noise in the performance evaluation (*i.e.*, the dependent variable in chapter 3) is less than perfect. I established the noise measure by using a single-item construct.

Despite efforts to buttress the validity of the construct, the noise measure would have benefited from a broader operationalization.

Another important limitation in the measurement arises from the fact that my study cannot adequately distinguish between the extent to which peer comparison relies on internal versus external peer groups. The questionnaire provides some information on whether RPE adopters use internal peers, external peers, or both. However, these data are not sufficiently refined to allow for more sophisticated analyses of the opportunism-mitigating explanation of RPE use.

Despite these limitations, I believe that the models and the findings of this study contribute to our understanding of RPE. Citing the original research proposal for my PhD project: "This study sets out to build on and extend the literature on RPE. Its aim is to document and explain the use of RPE in performance evaluation and compensation plan design at the business unit management level, adding to our understanding of the incentive structure within the firm as one of the key components of the Management Control System." Whether I succeeded in this task, I leave to the reader to answer.

Appendix
Overview of Questionnaire Items

RPE-Use

Q1 a. Do you feel that the performance of your peers (inside or outside your organization) is a point of reference for your superior when evaluating your performance?
b. Consider a situation where your actual performance is substantially better than your peers (inside or outside your organization). Would this situation influence your performance evaluation?
c. Consider a situation where your actual performance is substantially worse than your peers (inside or outside your organization). Would this situation influence your performance evaluation?
scale: not at all (1) - a little (2) - somewhat (3) - quite a bit (4) - very much (5) - I don't know (0)

RPE-based-Targets

Q2 We wish to understand the relative importance of a number of performance measures used. Please indicate these weights for each of these measures.
a. Weights of the performance measures per level
items: company level - business unit level - within business unit level - individual level
b. Measures at the business unit level
items: return measures - profit measures - disaggregated measures - nonfinancial measures
c. Measures at the within business unit level
items: return measures - profit measures - disaggregated measures - nonfinancial measures
scale: percentage: weight of measures
source: Bouwens en Van Lent (2007) (modified)

Q3 To what extent does the performance of your peers influence the height of your target and/or budget, for each of the following performance measures?
items: business unit level return measures - business unit level profit measures - business unit level disaggregated measures - business unit level nonfinancial measures
scale: to no extent (1) - some extent (3) - very great extent (5) - n/a

Environmental Uncertainty

Q4 How often do factors, external to your business unit, substantially influence your business unit's performance?
scale: (almost) never (1) - occasionally (2) - regularly (3) - quite often (4) - continuously (5)

Q5 How often do the following things happen?
a. Unpredictably, the amount of work varies greatly over time (example: peak demands)
source: Kruis (2008)
b. Exceptions arise in your business unit that require substantially different methods or procedures
scale: never (1) - rarely (2) - sometimes (3) - frequently (4) - always (5)
source: Van der Ven & Ferry (1980)

Q6 How much the same are day to day situations, problems or issues you encounter in your business unit?
scale: very much the same (1) - mostly the same (2) - somewhat different (3) - quite a bit different (4) - completely different (5)
source: Van der Ven & Ferry (1980)

Q7 Please indicate whether you agree or disagree with the following statements:
a. "The business unit often needs to react to outside pressure"
b. "Making long-range plans for this business unit is hindered by the difficulty of predicting future events"
items:
scale: disagree strongly (1) - disagree somewhat (2) - neutral (3) - agree somewhat (4) - agree strongly (5)
source: Kalleberg *et al.* (1996)

Comparability of the Business Unit

Q8 How much of the equipment (*e.g.*, machinery, tools) in your business unit is especially designed or unique compared to equipment used for similar activities in other companies?
scale: none (1) - some (2) - about half (3) - most (4) - all (5)
source: Coles & Hesterly (1998), modified

Q9 How much of the stock (*e.g.*, materials, finished products, half-finished products) in your business unit is especially designed or unique compared to equipment used for similar activities in other companies?
scale: none (1) - some (2) - about half (3) - most (4) - all (5)

Q10 How unique are the systems (*e.g.*, communication systems, software) that are used in your business unit compared to systems used by other companies for similar activities?
scale: not at all unique (1) - unique to a little extent (2) - somewhat unique (3) - quite unique (4) - very unique (5)
source: Kruis (2008)

Q11 How difficult is it for a new (core-business) business unit employee to learn the ins and outs he/she needs to know to be effective?
scale: very easy (1) - quite easy (2) - neither easy nor difficult (3) - quite difficult (4) - very difficult (5)
source: Anderson & Schmittlein (1984), modified

Consider a new (core-business) employee who has experience in a similar job/profession. How much extra training would that person -on average- need to be effective in your business unit
scale: weeks of training
source: Anderson & Schmittlein (1984), modified

How unique are the skills and knowledge of (core-business) business unit employees compared to skills and knowledge of employees of other companies who work on similar activities?
scale: not at all unique (1) - unique to a little extent (2) - somewhat unique (3) - quite unique (4) - very unique (5)
source: Kruis (2008), modified

Q12 How unique is the knowledge that is used in your business unit compared to systems used by other companies for similar activities?
scale: not at all unique (1) - unique to a little extent (2) - somewhat unique (3) - quite unique (4) - very unique (5)
source: Kruis (2008), modified

Q13 This question deals with you to position your business unit relative to your competitors on several topics. This question refers to your actual situation, not to what you desire.
items: brand image; compared to your competitors', your brand image is
scale: significantly lower (1) - about the same (4) - significantly higher (7)
source: Van der Stede (2001)

Q14 When serving your clients, how important is the knowledge of customer specific requirements?
scale: relatively unimportant, my business unit delivers standard products/services - slightly important, my business unit delivers standard products/services, with minor adaptations to customer specific requirements - somewhat important, my business unit delivers standard products/services, adapted to customer specific requirements - important, my business unit delivers tailored products/services - very important, my business unit delivers highly customer specific products/services, developed to meet specific customer needs

Q15 We need to understand which 'assets' of your business unit (both tangible and intangible) are most critical to your business unit's success. We would like you to indicate weights for each of these elements, so that the most important element gets most weight. Your answers should total 100 points.
items: equipment (*e.g.*, machinery, tools) - stock (*e.g.*, materials, finished products, half-finished products) - systems (*e.g.*, communication systems, software) - core-business employees - knowledge - brand name/image - client portfolio

Information Asymmetry

Q16 Please compare the amount of information you have relative to your superior (the person responsible for your performance evaluation and/or bonus determination)
a. Who is in possession of better information regarding the activities undertaken in your business unit?
b. Who is more familiar with the input/output relationships (the transformation process) inherent in the internal operations of your business unit?
c. Who is more certain of the performance potential of your business unit?
d. Who is more familiar technically with the work of your business unit?
e. Who is better able to assess the potential impact on your activities of factors internal to your business unit?
f. Who is better able to assess the potential impact on your activities of factors external to your business unit?
g. Who has a better understanding of what your business unit has achieved?
scale: your superior (1) - about equal (4) - you (7)
source: Dunk (1993) (items a-d and f-g) and Kruis (item e) (2008)

Goal Clarity/Ambiguity

Q17 How clearly defined are the goals of this business unit?
scale: very unclear (1) - quite unclear (2) - somewhat clear (3) - quite clear (4) - very clear (5)
source: Rainey (1983)

Q18 How specific are the goals of this business unit?
scale: not at all specific; very general (1) - not very specific; mostly general (2) - very specific (3) - somewhat specific (4) - quite specific (5)
source: Kruis (2008)

Q19 How easy is it to explain the goals of this business unit to outsiders?
scale: very easy (1) - quite easy (2) - neither easy nor difficult (3) - quite difficult (4) - very difficult (5)
source: Rainey (1983)

Q20 Please indicate whether you agree or disagree with the following statement: "The goals of this business unit are clear to almost everyone who works in the business unit."
scale: disagree strongly (1) - disagree somewhat (2) - neutral (3) - agree somewhat (4) - agree strongly (7)
source: Rainey (1983)

Measurability of Outputs

Q21 To what extent do the performance measures match the goals which your business unit has to realize?
scale: to (almost) no extent (1) - little extent (2) - some extent (3) - great extent (4) - very great extent (5)

Q22 Can the outputs of the business unit be measured objectively and expressed in a number?
scale: not at all (1) - a little (2) - somewhat (3) - quite a bit (4) - very much (5) - I don't know (0)
source: Kruis (2008)

Q23 Please indicate whether you agree or disagree with the following statements:
a. "The total set of performance measures indicates which results my business unit should realize"
b. "The total set of performance measures is unambiguously linked to the goals of my organization"
scale: disagree strongly (1) - disagree somewhat (2) - neutral (3) - agree somewhat (4) - agree strongly (5)
source: Speklé & Verbeeten (2008)

Knowledge of Transformation Process

Q24 Please indicate whether you agree or disagree with the following statements:
a. "The transformation process (the causal relation between the deployment of assets and realization of goals) within my business unit is clear and transparent to me.
source: Speklé & Verbeeten (2008)
b. "In case that it is unlikely that my business unit will realize its goals, I generally know what actions to take to still achieve them"
scale: disagree strongly (1) - disagree somewhat (2) - neutral (3) - agree somewhat (4) - agree strongly (5)

Difficulty to Assess Work

Q25 How easy is it for your superior to know whether you and your employees do the work correctly?
scale: very easy (1) - quite easy (2) - neither easy nor difficult (3) - quite difficult (4) - very difficult (5)
source: Van der Ven & Ferry (1980)

Difficulty to Reach Agreement on Output Quality

Q26 How easy is it to reach agreement with superiors when judging/appraising the quality of business unit performance?
scale: very easy (1) - quite easy (2) - neither easy nor difficult (3) - quite difficult (4) - very difficult (5)
source: Kruis (2008)

Knowledge about Job Performance

Q27 Please indicate whether you agree or disagree with the following statement: "My superior is familiar with my job performance"
scale: disagree strongly (1) - neutral (4) - agree strongly (7)
source: Farh *et al.* (1997)

Emphasis on Targets

Q28 a. For which purposes does your superior deploy the performance measurement system in relation to the business unit?
scale: not at all (1) - frequently (5)
b. How capable do you think the performance measurement system is for these purposes?
scale: incapable (1) - capable (5)
items: performance evaluation, reward policies

Q29 a. To what extent does your performance influence your performance evaluation?
b. To what extent does your performance influence the size of your bonus?
scale: to no extent (1) - little extent (2) - some extent (3) - great extent (4) - very great extent (5)

Emphasis on Firm-Level Measures

Q30 We wish to understand the relative importance of a number of performance measures used. Please indicate these weights for each of these measures.
a. Weights of the performance measures per level
items: company level - business unit level - within business unit level - individual level
b. Measures at the company level
items: stock price related measures - return measures - profit measures
source: Bouwens en Van Lent (2007) (modified)

Business Unit Size / Firm Size

Q31 What is the size of your business unit?
items: number of employees (approximately) - revenues (approximately)
scale: FTE - EURO

Q32 What is the size of your company as a whole?
items: number of employees (approximately) - revenues (approximately)
scale: FTE - EURO

Sector dummy

Q33 What industry is your business unit in?
items: production of goods - services - financial sector - not-for-profit

Noise

Q34 NOISE (primary measure)
Please indicate whether you agree or disagree with the following statements: "The evaluation I receive is based on factors over which I have control"
scale: disagree strongly (1) - neutral (4) - agree strongly (7)
source: Hartmann (1997)

Q35 Exogenous-Performance-Effects (first alternative measure)
How often do factors, external to your business unit, substantially influence your business unit's performance?
scale: (almost) never (1) - occasionally (2) - regularly (3) - quite often (4) - continuously (5)

Q36 Held-Accountable (second alternative measure)
Please indicate whether you agree or disagree with the following statements: "It frequently happens that my superior holds me accountable for certain (negative) results that I cannot help"
scale: disagree strongly (1) - neutral (4) - agree strongly (7)
source: Hartmann (1997)

Emphasis on Personal-Level Measures

Q37 We wish to understand the relative importance of a number of performance measures used. Please indicate these weights for each of these measures.
a. Weights of the performance measures per level
items: company level - business unit level - within business unit level - individual level
b. Measures at the personal level
items: professional skills measures - social skills measures - experience and knowhow measures
source: Bouwens en Van Lent (2007) (modified)

Emphasis on Disaggregated Measures

Q38 We wish to understand the relative importance of a number of performance measures used. Please indicate these weights for each of these measures.
a. Weights of the performance measures per level
items: company level - business unit level - within business unit level - individual level
b. Measures at the business unit level
items: disaggregated measures (*e.g.* costs, revenues, cash flow) - nonfinancial measures (*e.g.*, market share, customer satisfaction, quality)
c. Measures at the within business unit level
items: disaggregated measures (*e.g.* costs, revenues, cash flow) - nonfinancial measures (*e.g.*, market share, customer satisfaction, quality)
source: Bouwens en Van Lent (2007) (modified)

Room for Managerial Opportunism

Q39 Please indicate whether you agree or disagree with the following statements:
a. "If I would really want to, it is possible for me to receive my bonus, even if I would not have realised my targets"
b. "If I would really want to, I would be able to hide bad performance, so my superior would think that my business unit has performed well"
c. "Probably, my superior would not notice if I'd 'take it easy' "
d. "If I would really want to, I could window dress my business unit's figures, and make my business unit's performance look better than what would be realistic"
e. "Probably, my superior would not notice if my business unit would implement projects which I find interesting, but are not in the best interest of the company ("pet projects")"
scale: disagree strongly (1) - neutral (4) - agree strongly (7)

Decentralization of Decision Rights

Q40 In this question, we would like you to compare your authority with the authority of your superior (the person responsible for your performance evaluation and/or bonus determination) on the following decisions.
items: strategic decisions (*e.g.*, development of new products; enter and develop new markets; your business unit's strategy) - investment decisions (*e.g.*, acquiring new plants, property and equipment, development of new information systems) - marketing decisions (*e.g.*, campaigns, pricing decisions) - decisions regarding your internal processes (*e.g.*, setting production/sales priorities, resource allocation) - human resource decisions (*e.g.*, hiring and firing; compensation and setting career paths for the personnel employed within your business unit; reorganizing your business unit; creation of new jobs)
scale: my superior has all authority (1) - my business unit and my superior have the same authority (4) - my business unit has all authority (7)
source: Gordon & Narayanan (1984) - modified

References

1. Abernethy, M.A., Bouwens, J.F.G.M. & van Lent, L. 2004. 'Determinants of control system design in divisionalized firms.' The Accounting Review, 79, 545-70.

2. Aggarwal, R.K. & Samwick, A.A. 1999. 'Executive compensation, strategic competition, and relative performance evaluation: theory and evidence.' The Journal of Finance, 54, 1999-2043.

3. Aguinis, H. 1995. 'Statistical power with moderated multiple regression in management research.' Journal of Management, 21, 1141-58.

4. Aiken, L.S. & West, S.G. 1991. 'Multiple regression: testing and interpreting interactions.' Newbury Park, CA.: Sage Publications.

5. Albuquerque, A. 2009. 'Peer firms in relative performance evaluation.' Journal of Accounting and Economics, 48, 69-89.

6. Anderson, E. & Schmittlein, D.C. 1984. 'Integration of the sales force: an empirical examination.' The RAND Journal of Economics, 15, 385-95.

7. Antle, R. & Smith, A. 1986. 'An empirical investigation of the relative performance evaluation of corporate executives.' Journal of Accounting Research, 24, 1-39.

8. Antonakis, J., Bendahan, S., Jacquart, P. & Lalive, R. 2010. 'On making causal claims: a review and recommendations.' The Leadership Quarterly 21, 1086-120.

9. Baker, G. 2002. 'Distortion and risk in optimal incentive contracts.' The Journal of Human Resources, 37, 728-51.

10. Bertrand, M. & Mullainathan, S. 2001. 'Are CEOs rewarded for luck? The ones without principals are.' The Quarterly Journal of Economics 116, 901-32.

11. Bisbe, J., Batista-Foguet, J.-M. & Chenhall, R.H. 2007. 'Defining management accounting constructs: a methodological note on the risks of conceptual misspecification.' Accounting, Organizations and Society, 32, 789-820.

12. Bouwens, J.F.G.M. & Kroos, P. 2011. 'Target ratcheting and effort reduction.' Journal of Accounting and Economics, 51, 171-85.

13. Bouwens, J.F.G.M. & van Lent, L. 2007. 'Assessing the performance of business unit managers.' Journal of Accounting Research, 45, 667-97.

14. Bouwens, J.F.G.M. & Speklé, R.F. 2007. 'Does EVA add value?' In Hopper, T., Northcott D. & Scapens R. (Eds.), 'Issues in management accounting.' (3rd edition) Financial Times/Prentice Hall.

15. Bryne, B.M. 2001. 'Structural equation modeling with amos.' Mahwah, NJ: Lawrence Erlbaum Associates.

16. Chenhall, R.H. 2003. 'Management control systems design within its organizational context: findings from contingency-based research and directions for the future.' Accounting, Organizations & Society, 28, 127-68.

17. Chenhall, R.H. & Moers, F. 2007. 'The issue of endogeneity within theory-based, quantitative management accounting research.' European Accounting Review, 16, 173-96.

18. Chow, C.W., Cooper, J.C. & Waller, W. 1988. 'Participative budgeting: effects of a truth-inducing pay scheme and information asymmetry on slack and performance.' The Accounting Review, 63, 111-22.

19. Cohen, J., Cohen, P., West, S.G. & Aiken, L.S. 2003. 'Applied multiple regression/correlation analysis for the behavioral sciences.' (3rd edition) Mahwah, NJ: Lawrence Erlbaum Associates.

20. Coles, J.W. & Hesterly, W.S. 1998. 'Transaction costs, quality, and economies of scale: examining contracting choices in the hospital industry.' Journal of Corporate Finance, 4, 321-45.

21. Cronbach, L.J. 1987. 'Statistical tests for moderator variables: flaws in analyses recently proposed.' Psychological Bulletin, 102, 414-17.

22. Davila, A. & Penalva, F. 2006. 'Governance structure and the weighting of performance measures in CEO compensation.' Review of Accounting Studies, 11, 463-93

23. De Lancer Julnes, P. & Holzer, M. 2001. 'Promoting the utilization of performance measures in public organizations: an empirical study of factors affecting adoption and implementation.' Public Administration Review, 61, 693-708.

24. Dekker, H.C., Groot, T.L.C.M. & Schoute, M. 2012. 'Determining performance targets.' Behavioral Research in Accounting, *forthcoming*.

25. Dewatripont, M., Jewitt, I. & Tirole, J. 1999. 'The economics of career concerns, part II: application to missions and accountability of government agencies.' Review of Economic Studies, 66, 199 - 217.

26. Diamantopoulos, A. & Siguaw, J.A. 2006. 'Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration.' British Journal of Management, 17, 263-82.

27. Dillman, D.A. 2000. 'Mail and internet surveys; the tailored design method. New York: John Wiley & Sons, Inc.

28. Dunk, A.S. 1993. 'The effect of budget emphasis and information asymmetry on the relation between budgetary participation and slack.' The Accounting Review, 68, 400-10.

29. Dwyer, F.R. & Oh, S. 1987. 'Output sector munificence effects on the internal political economy of marketing channels.' Journal of Marketing Research, 24, 347-58.

30. Echambadi, R., Campbell, B. & Agerwal, R. 2006. 'Encouraging best practice in quantitative management research: an incomplete list of opportunities.' Journal of Management Studies, 43, 1801-20.

31. Echambadi, R. & Hess, J.D. 2007. 'Mean-centering does not alleviate collinearity problems in moderated multiple regression models.' Marketing Science, 26, 438-45.

32. Eisenhardt, K.M. 1989. 'Agency theory: an assessment and review.' The Academy of Management Review, 14, 57-74.

33. Etzioni, A. 1961. 'A comparative analysis of complex organizations.' New York: Free Press.

34. Farh, J.-L., Earley P. C. & S.-C., Lin. 1997. 'Impetus for action: a cultural analysis of justice and organizational citizenship behavior in chinese society.' Administrative Science Quarterly, 42, 421-44.

35. Fisher, J.G., Frederickson, J.R. & Peffer, S.A. 2002. 'The effect of information asymmetry on negotiated budgets: an empirical investigation.' Accounting, Organizations and Society, 27, 27-43.

36. Fisher, J.G., Maines, L.A., Peffer, S.A. & Sprinkle, G.B. 2002. 'Using budgets for performance evaluation: effects of resource allocation and horizontal information asymmetry on budget proposals, budget slack, and performance.' The Accounting Review, 77, 847-65.

37. Frederickson, J.R. 1992. 'Relative performance information: the effects of common uncertainty and contract type on agent effort.' The Accounting Review, 67, 647-69.

38. Garvey, G. & Milbourn, T. 2003. 'Incentive compensation when executives can hedge the market: evidence of relative performance evaluation in the cross section.' The Journal of Finance, 58, 1557-85

39. Garvey, G. & Milbourn, T. 2006. 'Asymmetric benchmarking in compensation: executives are rewarded for good luck but not penalized for bad.' Journal of Financial Economics, 82, 197-225

40. Gerdin, J. & Greve, J. 2004. 'Forms of contingency fit in management accounting research - a critical review.' Accounting, Organizations and Society, 29, 303-26.

41. Gerdin, J. & Greve, J. 2008. 'The appropriateness of statistical methods for testing contingency hypotheses in management accounting research.' Accounting, Organizations and Society, 33, 995-1009.

42. Gibbons, R. 1998. 'Incentives in organizations.' Journal of Economic Perspectives, 12:4, 115-32.

43. Gibbons, R. & Murphy, K.J. 1990. 'Relative performance evaluation for chief executive officers.' Industrial & Labor Relations Review, 43, 30-S-51-S.

44. Gong, G., Li, L.Y. & Shin, J.Y. 2011. 'Relative performance evaluation and related peer groups in executive compensation contracts.' The Accounting Review, 86, 1007-43.

45. Gordon, L. A. & Narayanan, V.K. 1984. 'Management accounting systems, perceived environmental uncertainty and organization structure: an empirical investigation.' Accounting, Organizations and Society, 9, 33-47.

46. Gundlach, G.T., Achrol, R.S. & Mentzer, J.T. 1995. 'The structure of commitment in exchange.' Journal of Marketing, 59:1, 78-92.

47. Hak, T., Veer, C.G. van der & Jansen, H. 2008. 'The three-step test-interview (TSTI): an observational instrument for pretesting self-completion questionnaires.' Survey Research Methods, 2, 143-50.

48. Hair, J.F., Black, W.C., Babin, B.J. & Anderson R.E., 2010. 'Multivariate data analysis: a global perspective.' (7th edition). New Jersey: Pearson Prentice Hall.

49. Hannan, R.L., Rankin, F.W. & Towry, K.L. 2006. 'The effect of information systems on honesty in managerial reporting: a behavioral perspective.' Contemporary Accounting Research, 23, 885-918.

50. Hartmann, F.G.H. 1997. 'Accounting for performance evaluation: effects of uncertainty on the appropriateness of accounting performance measures.' Maastricht: Maastricht University.

51. Hartmann, F.G.H. & Moers, F. 1999. 'Testing contingency hypotheses in budgetary research: an evaluation of the use of moderated regression analysis.' Accounting, Organizations and Society, 24, 291-315.

52. Himmelberg, C.P. & Hubbard R.G. 2000. 'Incentive pay and the market for CEOs: an analysis of pay-for-performance sensitivity.' Working paper.

53. Hofstede, G. 1981. 'Management control of public and not-for-profit activities.' Accounting, Organizations and Society, 6, 193-211.

54. Holmstrom, B. 1982. 'Moral hazard in teams.' Bell Journal of Economics, 13, 324-40.

55. Indjejikian, R.J. & Nanda, D. 2002. 'Executive target bonuses and what they imply about performance standards.' The Accounting Review, 77, 793-819.

56. Ittner, C.D. & Larcker, D.F. 2001. 'Assessing empirical research in managerial accounting: a value-based management perspective.' Journal of Accounting and Economics, 32, 349-410.

57. Ittner, C.D., Larcker, D.F. & Rajan, M.V. 1997. 'The choice of performance measures in annual bonus contracts.' The Accounting Review, 72, 231-56.

58. Janakiraman, S.N., Lambert, R.A. & Larcker, D.F. 1992. 'An empirical investigation of the relative performance evaluation hypothesis.' Journal of Accounting Research, 30, 53-69.

59. Jarvis, C.B., MacKenzie, S.B. & Podsakoff, P.M. 2003. 'A critical review of construct indicators and measurement model misspecification in marketing and consumer research.' Journal of Consumer Research, 30, 199-218.

60. Jensen, M.C. & Murphy, K.J. 1990. 'Performance pay and top-management incentives.' Journal of Political Economy, 98, 225-64.

61. John, G. 1984. 'An empirical investigation of some antecedents of opportunism in a marketing channel.' Journal of Marketing Research, 21, 278-89.

62. Kalleberg, A.L., Knoke, D., Marsden, P.V., & Spaeth J.L. 1996. 'Organizations in america: analyzing their structures and human resource practices based on the National Organizations Study.' Thousand Oaks, CA Sage Publications, Inc.

63. Kline, R.B. 2005. 'Principles and practice of structural equation modeling.' (2nd edition) New York: The Guilford Press.

64. Kren, L. 2002. 'Common uncertainty effects on the use of relative performance evaluation for corporate chief executives.' Advances in Accounting, 19, 119-38.

65. Kruis, A. 2008. 'Management control systems design and effectiveness.' Breukelen: Nyenrode Business University.

66. Kunz, A.H. & Pfaff, D. 2002. 'Agency theory, performance evaluation, and the hypothetical construct of intrinsic motivation.' Accounting, Organizations and Society, 27, 275-95.

67. Langfield-Smith, K. & Smith, D. 2003. 'Management control systems and trust in outsourcing relationships.' Management Accounting Research, 14, 281-307.

68. Larcker, D.F. & Rusticus, T.O. 2010. 'On the use of instrumental variables in accounting research.' Journal of Accounting and Economics, 49, 186-205.

69. Leone, A.J. & Rock, S. 2002. 'Empirical tests of budget ratcheting and its effect on managers' discretionary accrual choices.' Journal of Accounting and Economics, 33, 43-67.

70. Liu, L.S. & Stark, A.W. 2008. 'Relative performance evaluation in board cash compensation: UK empirical evidence.' British Accounting Review, *forthcoming*.

71. Luft, J. & Shields, M.D. 2003. 'Mapping management accounting: graphics and guidelines for theory-consistent empirical research.' Accounting, Organizations and Society, 28, 169-249.

72. Maas, V.S. & Van Rinsum, M. 2011. 'How control system design influences performance misreporting.' Working Paper.

73. Macher, J.T. & Richman, B.D. 2008. 'Transaction cost economics: an assessment of empirical research in the social sciences.' Business and Politics, 10, article 1.

74. Matsumura, E.M. & Shin, J.Y. 2006. 'An empirical analysis of an incentive plan with relative performance measures: evidence from a postal service.' The Accounting Review, 81, 533-66.

75. Merchant, K.A. & Manzoni, J.-F. 1989. 'The achievability of budget targets in profit centers: a field study.' The Accounting Review, 64, 539-58.

76. Murphy, K.J. 2001. 'Performance standards in incentive contracts.' Journal of Accounting and Economics, 30, 245-78.

77. Otley, D.T. & Berry, A.J. 1980. 'Control, organization and accounting.' Accounting, Organizations and Society, 5, 231-44.

78. Ouchi, W.G. 1979. 'A conceptual framework for the design of organizational control mechanisms.' Management Science, 25, 833-48.

79. Rainey, H.G. 1983. 'Public agencies and private firms: incentive structures, goals, and individual roles.' Administration Society, 15, 207-42.

80. Rogers, E.M. 1995. 'Diffusion of innovations.' New York: Free Press.

81. Speklé, R.F. 2001. 'Explaining management control strucuture variety: a transaction cost economics perspective.' Accounting, Organizations & Society, 26, 419-41.

82. Speklé, R.F. & Verbeeten, F.H.M. 2008. 'The use of performance measurement systems in the public sector: effects on performance.' Working Paper.

83. Towry, K. 2003. 'Control in a teamwork environment - the impact of social ties on the effectiveness of mutual monitoring contracts.' The Accounting Review 78, 1069-95.

84. Van der Stede, W.A. 2001. 'The effect of corporate diversification and business unit strategy on the presence of slack in business unit budgets.' Accounting, Auditing & Accountability Journal, 14, 30-52.

85. Van der Ven, A.H.& Ferry, D.L. 1980. 'Measuring and assessing organizations.' New York: John Wiley & Sons, Inc.

86. Veall, M.R. & Zimmermann, K.F. 1994. 'Goodness of fit measures in the Tobit model.' Oxford Bulletin of Economics and Statistics, 56, 485-99.

87. Wathne, K.H. & Heide, J.B. 2000. 'Opportunism in interfirm relationships: forms, outcomes, and solutions.' Journal of Marketing, 64:4, 36-51.

88. Widener, S.K. 2007. 'An empirical analysis of the levers of control framework.' Accounting, Organizations and Society, 32, 757-88.

89. Williamson, O.E. 1975. 'Markets and hierarchies: analysis and antitrust implications.' New York: The Free Press.

90. Williamson, O.E. 1985. 'The economic institutions of capitalism.' New York: The Free Press.

91. Zhang, Y. 2008. 'The effects of perceived fairness and communication on honesty and collusion in a multi-agent setting.' The Accounting Review, 83, 1125-46.

Samenvatting

Dit proefschrift bestudeert het gebruik van Relatieve Prestatie Evaluatie (RPE) voor de prestatiebeoordeling van business unit managers. RPE is een methode om de prestaties van een medewerker te beoordelen door zijn prestaties te vergelijken met een referentiegroep. Een dergelijke referentiegroep bestaat uit *peers* die opereren onder vergelijkbare omstandigheden en verantwoordelijk zijn voor vergelijkbare taken als de geëvalueerde medewerker. De prestaties van de referentiegroep vormen een prestatienorm, waarmee de werkelijke prestaties van de medewerker worden vergeleken.

RPE staat conceptueel tegenover het formuleren van prestatiestandaarden als absolute norm, bijvoorbeeld aan de hand van een *ex ante* vastgesteld budget. Het voordeel van een relatieve prestatiestandaard ten opzichte van een absolute norm is de gevoeligheid van de relatieve standaard voor externe omstandigheden, voor zover deze zowel de prestaties van de medewerker als de prestaties van de referentiegroep beïnvloeden. In het geval van een absolute, *ex ante* bepaalde norm kunnen externe omstandigheden de haalbaarheid van de norm sterk beïnvloeden. Tegenvallende omstandigheden maken het behalen van de norm onwaarschijnlijk of zelfs onmogelijk. Indien het behalen van de prestatiestandaard is gekoppeld aan een beloning, dan verlagen negatieve omstandigheden ook de compensatie van de medewerker. In dit geval draagt de medewerker het risico van factoren die buiten zijn bereik liggen. Dit komt de motivatie van de medewerker niet ten goede. *Vice versa* is het ook mogelijk dat de omstandigheden in het voordeel van de medewerker uitvallen. Dit vergemakkelijkt de realisatie van de prestatienorm. Ook in deze situatie motiveert de prestatienorm de medewerker suboptimaal, omdat eenvoudig realiseerbare doelen geen maximale inspanning vereisen. In beide gevallen beperken externe omstandigheden de realisatie van de organisatiedoelen door verminderde motivatie en inspanning van de medewerker.

Het gebruik van een relatieve prestatienorm kan echter uitkomst bieden. Als externe omstandigheden niet alleen de eigen prestaties beïnvloeden, maar ook die van de referentiegroep, dan beweegt de prestatienorm mee met de omstandigheden. Dit verkleint de kans op een te gemakkelijke of juist te moeilijke haalbaarheid van de norm. Externe omstandigheden die de haalbaarheid van een *ex ante* bepaalde, absolute prestatienorm

beïnvloeden, veroorzaken ruis in de prestatiebeoordeling. Het bereiken van de prestatienorm is dan immers niet langer alleen het resultaat van de inspanningen van de medewerker, maar ook van omstandigheden waarop de medewerker geen invloed heeft. Op theoretische gronden wordt verwacht, dat RPE deze ruis beperkt door de incorporatie van externe omstandigheden in de norm door middel van vergelijking met een referentiegroep.

Volgens de literatuur is de beperking van ruis in de prestatiebeoordeling echter niet het enige argument voor het gebruik van RPE. RPE kan ook dienen om de ruimte te beperken die medewerkers hebben om opportunistisch te handelen. *Ex ante* bepaalde prestatienormen worden veelal vastgesteld op basis van informatie die voor de geëvalueerde medewerker relatief eenvoudig is te manipuleren. Voorbeelden van dergelijke methoden zijn het baseren van de norm op de prestaties van de voorgaande periode of naar aanleiding van interne budgetonderhandelingen. In deze gevallen heeft de medewerker een prikkel om geen prestaties te leveren die beter zijn dan gebudgetteerd of om onjuiste informatie in te brengen in de onderhandelingen. Dit resulteert namelijk in een gemakkelijker budget voor de komende periode, waarbij hij minder inspanning hoeft te leveren om de prestatienorm te realiseren. De medewerker heeft echter geen (of: minder) invloed op de prestaties van de referentiegroep, en kan derhalve de daarvan afgeleide norm niet opportunistisch manipuleren. Dit argument suggereert dat RPE een geschikt middel kan zijn om de ruimte tot opportunistisch handelen onder medewerkers te verkleinen.

Bovenstaande argumenten doen vermoeden dat RPE op grote schaal wordt toegepast in de praktijk om zo ruis en opportunisme te beteugelen. Empirisch is hier echter weinig over bekend. Als eerste studie naar het gebruik van RPE op het business unit management niveau, bestudeert mijn onderzoek in hoeverre en waarom business unit managers worden afgerekend op relatieve prestatienormen. Daartoe zijn data verzameld onder 325 business unit managers aan de hand van een zeer uitvoerige vragenlijst. Deze vragenlijst levert veel gedetailleerde informatie op over de toepassing van RPE, die niet is verkregen bij eerder onderzoek.

Mijn onderzoek wijst uit dat RPE inderdaad op grote schaal wordt toegepast onder business unit managers; bij meer dan de helft van de managers in mijn steekproef speelt de vergelijking met een referentiegroep een (zeer) grote rol bij de beoordeling van de prestaties. Slechts 8% van de door mij ondervraagde managers geeft aan in het geheel niet te worden geëvalueerd aan de hand van vergelijking met een referentiegroep.

Vervolgens onderzoekt deze studie in hoeverre het gebruik van RPE kan worden verklaard door haar bijdrage aan het terugdringen van ruis en opportunisme. In hoofdstukken 2 en 3 van mijn proefschrift wordt de relatie tussen RPE en ruis in de prestatiebeoordeling geanalyseerd. Beide hoofdstukken bevestigen de verwachting dat RPE effectief is in het

beperken van de ruis die ontstaat door externe omstandigheden. Door het gebruik van een relatieve prestatienorm worden de externe omstandigheden uit de prestatiebeoordeling gefilterd, voor zover de geëvalueerde manager en de referentiegroep hier beiden aan zijn blootgesteld.

In hoofdstukken 2 en 4 onderzoek ik of en in hoeverre RPE bijdraagt aan het verminderen van de ruimte die managers hebben om opportunistisch te handelen. Tegengesteld aan mijn verwachtingen vind ik dat RPE de ruimte voor opportunisme vergroot in plaats van verkleint. Deze bevinding kan worden verklaard door het feit dat circa 92% van de RPE gebruikers in mijn steekproef de referentiegroep (gedeeltelijk) baseren op peers binnen de eigen organisatie (bijvoorbeeld: andere business unit managers). Hoewel een dergelijke referentiegroep potentieel de vergelijkbaarheid van de prestaties vergroot en daarmee de vergelijking beter toepasbaar maakt voor het reduceren van ruis in de prestatiebeoordeling, is het goed denkbaar dat de prestatienorm manipuleerbaarder wordt. Het is mogelijk dat de prestaties van directe collega's gemakkelijker zijn te beïnvloeden door de business unit manager dan een absolute, *ex ante* gedefinieerde norm. Een dergelijke situatie zou samenzwering makkelijker maken teneinde gezamenlijk de prestatienormen te versoepelen. De literatuur bevestigt de opvatting dat om de manipuleerbaarheid van de prestatienorm te beperken, het beter is om externe peers te kiezen (zoals managers van vergelijkbare, concurrerende business units) dan een interne referentiegroep.

Deze bevinding diskwalificeert RPE als zodanig echter niet als methode om opportunistisch gedrag onder business unit managers tegen te gaan. RPE, zoals het momenteel overwegend wordt toegepast in de praktijk, lijkt weliswaar niet geschikt om opportunisme te mitigeren. Echter, RPE beteugelt opportunistisch gedrag mogelijk wel indien de prestatienorm wordt vastgesteld op basis van een externe referentiegroep. Externe referentiegroepen worden echter weinig gebruikt onder mijn respondenten. Mijn onderzoek biedt geen aanleiding om de theoretische verwachtingen omtrent opportunisme-beperkende eigenschappen van RPE in het algemeen te verwerpen. Een eerste suggestie aan de praktijk luidt derhalve om, indien een organisatie RPE wil gebruiken om opportunisme onder haar business unit managers te beperken, een referentiegroep samen te stellen met externe peers.

Biography

Hilco van Elten (1979) holds an MSc degree from Erasmus University (2004; Economics and Business, major Accounting and Controlling), and a Bachelor degree from Utrecht Polytechnic (2002, HEAO Accountancy RA). After graduating, he decided to stay at Erasmus University to work for the Accounting section of the Erasmus School of Economics. Also, Hilco joined the Erasmus Research Institute of Management (ERIM).

In January 2006, Hilco left Erasmus University to join the Centre of Accounting, Auditing & Control at Nyenrode Business Universiteit. At Nyenrode, he wrote his dissertation and got involved in various research and teaching activities. Currently, he is employed as Assistant Professor at Nyenrode, where he dedicates most of his time to teaching and research. His teaching responsibilities include Management Accounting and Control and methodological courses, which he enjoys greatly. Concerning his research activities, his main interest lies in incentive system design.

Hilco is married to Maria, with whom he has two daughters Frederique (2009) and Josephine (2012). Outside of work, Hilco enjoys spending time with family and friends. The few remaining hours, he spends reading, playing the guitar, or in the gym.

NYENRODE
BUSINESS UNIVERSITEIT

ISBN 978-90-8980-038-1