

JAN NOEVERMAN

Management Control Systems, Evaluative Style, and Behaviour

Exploring the Concept and Behavioural Consequences of Evaluative Style



Management Control Systems, Evaluative Style, and Behaviour:

Exploring the concept and behavioural consequences

of evaluative style

Management Control Systems, Evaluative Style, and Behaviour:

Exploring the concept and behavioural consequences

of evaluative style

Management control systemen, beoordelingsstijl, en gedrag:

een verkenning van het concept en de gedragsmatige gevolgen

van beoordelingsstijl

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam op gezag van de
rector magnificus

Prof.dr. S.W.J. Lamberts
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 21 december 2007 om 09.00 uur

door

Jan Noeverman
geboren te Haarlemmermeer



Promotiecommissie

Promotoren: Prof.dr. E.G.J. Vosselman
Prof.dr. A.R.T. Williams

Overige leden: Prof.dr. F.G.H. Hartmann RC
Prof.dr. D.T. Otley
Prof.dr. J. Paauwe

Copromotor: Dr. B.A.S. Koene

Erasmus Research Institute of Management (ERIM)
RSM Erasmus University / Erasmus School of Economics
Erasmus University Rotterdam

Internet: <http://www.erim.eur.nl>

ERIM Electronic Series Portal: <http://hdl.handle.net/1765/1>

ERIM Ph.D. Series Research in Management, 120

ISBN 90 - 5892 - 151 - 2

Design: B&T Ontwerp en advies www.b-en-t.nl / Print: Haveka www.haveka.nl

© 2007, Jan Noeverman

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the author.

Preface

When I started working on this thesis, it was not my intention to make it a lifetime project. It was just a Ph.D. thesis. Now, with the benefit of hindsight, it is hard to deny that this thesis at least appears to have become a lifetime project, if not for myself, than certainly for my children. Since they were born, Marlijn, Gerjan and Hanne have only known a dad who had to work on 'his book', often also or even especially during holidays and on Saturdays. Until now, my wife Anneke has only known a husband who always felt the need and the pressure to finish his Ph.D. And my friends and relatives, who were lucky enough to get to know me before I started working on my Ph.D. thesis, stopped asking about the project a long time ago, probably also because they noted that I did not really want to talk about it. In this sense, over the years this project has become much more than just a thesis. It has had a great impact on my life and that of people surrounding me. But now that it is finished, it is good to realise that in the end it is still just a Ph.D. thesis.

I owe thanks to many people. Without them, this Ph.D. would not have existed. First of all my promotors, Roger Wiliams and Ed Vosselman. Roger, you are the only person who has been involved during the whole process, from start to finish. You never stopped believing in me and this project, and your continuing support motivated me to continue. I was often amazed by the way you were able to summarise what I intended to say in a very comprehensive but also very compact way, and really thinking through the consequences of what I had written. Thank you for always being enthusiastic about the topic of my thesis! Ed, thanks for pushing me to just finish the Ph.D., although it often must have seemed as if it did not help much! Your support also helped me to believe that this project was worth working on. Thanks also for making me a 'real academic', because as the head of our department, you offered me a job as assistant professor.

Second, a special thanks to my co-promotor, Bas Koene. Bas, you appear to be even more critical and perfectionistic than I am! When I finally thought the thesis was good enough, you still pointed out some issues that in your opinion needed to be resolved. And even though I did not like the idea of working on it any further, after I made the changes, I knew you were right - as always. When you became involved to introduce me to VDBN, I am sure you did not expect that your involvement would last for years. I am glad that your efforts have been rewarded by being appointed as co-promotor. We wrote a number of papers together, got two of them published, and I hope more joint publications will follow in the future.

Third, I like to thank the many people who have contributed to (parts of) my Ph.D. thesis at various stages: Frits Krens, my first promotor, and Jan Aukes, who were involved in the beginning of the project, and helped me sharpen my ideas and think about how to collect the data. David Otley, who has really helped me shape the topic of this thesis when I discussed

the first drafts of my thesis proposal with him at the Free University of Amsterdam many years ago. David pointed out to me a lot of shortcomings in the existing RAPM literature and gave me some of his co-authored working papers years before they got published. The approach I take in this Ph.D. thesis would have been much harder to defend without the ideas from these papers. Harry Commandeur helped me with the process of writing the Ph.D. when I had a lot of unstructured written pieces. He encouraged me to make a preliminary table of contents, and forced me to think about the conclusions and main contribution of the thesis before writing any chapter. Since then, although the thesis still had to be written, the main ideas were clear and have remained the same. Arie de Bruin en Rob Potharst, colleagues I met through my teaching activities, voluntarily offered their help to try to overcome some problems I ran into when I got to the data analysis phase. They helped me oversee what was needed to analyse the data. Patrick Groenen provided me with useful suggestions on statistical analysis techniques I could use. Roland Speklé read a draft of the manuscript when it was near completion and discussed with me some possibilities to get the thesis done. Of course I thank the respondents at Van den Bergh Netherlands for their cooperation, and Lennard Boogaard, Human Resource Director of Unilever, for giving me permission to collect data and to write about VDBN and their performance evaluation system. My present and former colleagues of the Accounting section of the Erasmus School of Economics helped me by commenting on papers or presentations derived from the thesis, but also by teaching classes so that I could have time to work on my Ph.D.

I already mentioned my friends and relatives. The 'Heemskerkvrienden' are very special friends, and their friendship means a lot to me. I surely hope our friendship is a friendship for life. My parents were always willing to either let me or the kids stay at their house, so that I could work a few days or weeks on my Ph.D. without being disturbed. Thanks mom and dad!

The last words that I wrote in this book are these words for Anneke and the kids. You are my life. Words cannot express how much I love you. Anneke, without you I could not have handled the stress and pressure that come along with writing a Ph.D. I realise this is only true because you probably carried most of the stress and pressure. We made a lifetime commitment, and it was the best decision I ever made. I love you. Thanks for being my wife and for loving me. I can only imagine how hard that must be at times!

Finally, I thank God, my Creator and Redemeer, Father, Son and Holy Ghost, for His great gift of life and love. Lord, You are the Resurrection and the Life. You gave Your life for me, so that I shall not perish but have everlasting life. There is no greater love than this! Knowing You has always helped me to put everything else in perspective. It is my prayer that Christ will live through me and fill me with life more abundantly. That is all that really matters.

December 2007

Table of Contents

1. Introduction	1
1.1 Introduction	1
1.2 Research relevance	3
1.3 Research purpose and method	3
1.4 Overview and structure of dissertation	4
2. RAPM-research: an overview and evaluation	8
2.1 Introduction	8
2.2 The traditional view of management control, management accounting and managerial behaviour	8
2.2.1 Introduction	8
2.2.2 Anthony's definition of management control	8
2.2.3 Management control and management accounting	9
2.3 An overview of RAPM-literature	12
2.4 An evaluation of RAPM-literature: a review of criticisms	16
2.4.1 Introduction	16
2.4.2 Criticisms on conceptualisation and measurement of evaluative style	20
2.4.3 Criticisms relating to research method	21
2.4.4 Criticism relating to theoretical development	22
2.5 Implications of criticism for further research	23
2.5.1 Introduction	23
2.5.2 Construct measurement of evaluative style	23
2.5.3 Theoretical development	27
2.6 Conclusion	27
3 Construct measurement of evaluative style: a review and proposal	29
3.1 Introduction	29
3.2 Review of existing measures of evaluative style	30
3.2.1 Introduction	30
3.2.2 Cell A1: Hopwood (1972a; 1973)	33
3.2.3 Cell A2: Otley (1978) and Govindarajan (1988)	35
3.2.4 Cell A3: Brownell (1982)	37
3.2.5 Cell A4: Brownell (1985)	38
3.2.6 Cell A6: Harrison (1992; 1993)	40
3.2.7 Cell B4: Hirst (1983) and Van der Stede (2000; 2001)	42
3.2.8 Cell B6: Govindarajan (1984)	43

3.3	Evaluation	44
3.3.1	Conceptualisation of evaluative style	44
3.3.2	Validity	45
3.3.3	Reliability	45
3.3.4	Conclusion of review	45
3.4	Implications for future construct measurement	46
3.4.1	Allow a broader perspective on type of information	46
3.4.2	Include the manner in which superiors use information	47
3.4.3	Study differences in evaluative style in context rather than across contexts	48
3.4.4	Control for the organisational context	49
3.5	The concept of evaluative style in practice: the case of Service Group	51
3.5.1	Introduction	51
3.5.2	Description of organisation and performance evaluation system	52
3.5.3	Performance evaluation in practice	53
3.5.4	Discussion	61
3.6	Conclusion	63
4	Antecedents and consequences of evaluative style: a theoretical framework	65
4.1	Introduction	65
4.2	The antecedents of perceived evaluative style	67
4.2.1	Introduction	67
4.2.2	Factors at the organisational level	68
4.2.3	Factors at the level of groups	72
4.2.4	Factors at the individual level: subordinate characteristics	77
4.3	Superior characteristics: Leadership style	79
4.4	Consequences of perceived evaluative style: fairness of evaluation, trust in superior, and subordinate managerial behaviour	81
4.4.1	Introduction	81
4.4.2	The role of fairness of evaluation	84
4.4.3	Agreement with evaluative criteria: the role of task uncertainty	86
4.5	Conclusion	93
5	In-depth field study: research design	96
5.1	Introduction	96
5.2	Purpose and method of the field study	96
5.3	Data collection	97
5.3.1	Sample	97
5.3.2	Interview data	100

5.3.3	Written questionnaire	102
5.3.4	Internal documents	104
5.4	Variable measurement in the written questionnaire	104
5.4.1	Introduction	104
5.4.2	The measurement of perceived evaluative style	104
5.4.3	The measurement of leadership style	107
5.4.4	The measurement of contextual variables	110
5.4.5	The measurement of fairness of evaluation	111
5.4.6	The measurement of trust in superior	113
5.4.7	Variables relating to behavioural outcomes	113
5.4.8	Descriptive statistics	115
5.5	Methods of data analysis	116
5.5.1	Method of analysis of interview data	116
5.5.2	Method of analysis of questionnaire data	117
5.6	Summary and conclusion	117
6.	Description of the organisational context	119
6.1	Introduction	119
6.2	Description of organisation (VDBN)	119
6.3	Performance evaluation at VDBN: description of the Integrated Approach	122
6.3.1	Purpose of the Integrated Approach	122
6.3.2	Elements of the Integrated Approach	123
6.4	General observations	125
6.5	Sample characteristics: the use of categorical principal components analysis	127
6.6	Conclusion	135
7	The construct of perceived evaluative style: empirical evidence	137
7.1	Introduction	137
7.2	Differences in perceived evaluative style: a qualitative analysis	138
7.2.1	Introduction	138
7.2.2	Underlying dimensions of evaluative style emerging from the interview data	138
7.2.3	Conclusion	144
7.3	Relevance of measured aspects of evaluative style	144
7.3.1	Introduction	144
7.3.2	Data analysis technique	145
7.3.3	Results	145
7.4	Underlying dimensions of perceived evaluative style	147

7.4.1	Introduction	147
7.4.2	Data analysis technique	149
7.4.3	Results	150
7.5	Evaluative style as a group level phenomenon	155
7.5.1	Introduction	155
7.5.2	Data analysis technique	156
7.5.3	Results	157
7.6	Discussion and conclusion	163
8	Antecedents and consequences of perceived evaluative style: empirical evidence	169
8.1	Introduction	169
8.2	Perceived evaluative style and context	169
8.2.1	Introduction	169
8.2.2	Data analysis technique	170
8.2.3	Results	172
8.3	Perceived evaluative style and leadership style	175
8.3.1	Introduction	175
8.3.2	Data analysis technique	176
8.3.3	Results for leadership style and 17 perceived evaluative style items	176
8.3.4	Results: leadership style and three perceived evaluative style subscales	178
8.4	Perceived evaluative style and subordinate managers' behaviour: a quantitative analysis	181
8.4.1	Introduction	181
8.4.2	Data analysis technique	183
8.4.3	Results	186
8.5	Perceived evaluative style and subordinate manager's behaviour: a qualitative analysis	190
8.5.1	Introduction	190
8.5.2	Subordinates' perceptions of the PDP-system	190
8.5.3	Overall fairness of evaluation	194
8.5.4	Dysfunctional behaviours	195
8.5.5	Functional behaviours	196
8.6	Summary and conclusions	197
8.6.1	Contextual factors and perceived evaluative style (1.3)	197
8.6.2	Perceived leadership style and perceived evaluative style (1.4)	201
8.6.3	Perceived evaluative style and subordinate managers' behaviour	203

9 Conclusion	206
9.1 Introduction	206
9.2 Summary of main findings	206
9.3 Discussion	212
9.4 Unresolved issues and limitations	215
9.5 Possibilities for further research/limitations	217
References	220
Appendices	231
Appendix A: Interview protocols used within Service Group (see § 3.5)	231
Appendix B. Organisational structure of business unit of Service Group	234
Appendix C. Interview protocols used at VDBN	235
Appendix D. Overview of variable measurement in the questionnaire	241
Appendix E. Output of CATPCA analysis reported in § 6.5.2	247
Appendix F. Calculation of within-group agreement coefficients (rwg)	251
Appendix G. The calculation and coding of within-group differences	252
Samenvatting (in Dutch)	253
Biography	261

1 Introduction

1.1 Introduction

Many large organisations have staff departments where people are being paid to develop, implement and maintain all kinds of systems that are intended to influence and motivate participants to behave in the best interest of the organisation. High investments are made in sophisticated management control systems, performance evaluation systems, accounting information systems, etc. But do these systems really affect people in the way that they are intended to?

Some early studies (Argyris, 1952; Hopwood, 1972a) found that using accounting information for evaluating the performance of subordinate managers could enhance feelings of tension and evoke dysfunctional behaviours such as manipulating accounting information and inferior relationships with peers and superiors. These studies also suggest that participants' behaviour is not affected by the adequacy of the system only, but as much by how managers use these systems. In the past few decades, many studies have tried to relate performance evaluation style to behavioural and attitudinal outcomes at the subordinate level. Performance evaluation style, or, in short, evaluative style, refers to the manner in which managers evaluate the performance of their subordinate managers. In the accounting literature, research addressing evaluative style has based the distinction between different evaluative styles on the extent to which a superior relies on accounting performance measures when evaluating subordinates' performance. The literature on this concept of evaluative style has become known as RAPM (reliance on accounting performance measures). Although this literature provides consistent support for the existence of a relationship between evaluative style and job-related tension, performance and dysfunctional behaviours (e.g., Hopwood 1973; Otley, 1978; Brownell & Hirst, 1986; Ross, 1995), there is also much support for the hypothesis that this relationship is conditional on organizational (e.g. Otley, 1978; Hirst, 1981; 1983; 1987), environmental (e.g. Brownell, 1985; 1987), and personal factors (e.g. Hopwood, 1973; Brownell, 1981). However, a clear understanding of the nature of this relationship under different conditions is still lacking. There are several problems with the RAPM-literature that may have limited a good understanding of the concept of evaluative style and, subsequently, of its relationship with subordinates' managerial behaviour.

First, RAPM studies have been severely criticised in terms of variable measurement (Otley & Fakiolas, 2000; Otley & Pollanen, 2000; Vagneur & Peiperl, 2000), data collection (Hartmann, 2000), data analysis (Hartmann & Moers, 1999) and theoretical development (Briers & Hirst, 1990; Hartmann, 2000; Otley & Fakiolas, 2000). Many of these criticisms refer to the conceptualisation and measurement of evaluative style. The construct of evaluative style has been modified in later studies without showing the necessity and validity of these

modifications, thereby inhibiting comparisons between studies. Other studies use an earlier measure of evaluative style without making any modifications, but have not investigated whether this measure still captures the same phenomena in their study as in earlier studies. This can not be automatically assumed, given the many changes in the management accounting and management control field over the past two decades. Some of these changes, such as the attention given to the Balanced Scorecard, directly concern the role of financial performance measures and of budgets. Therefore, future conceptualisations and measures of evaluative style should be based on distinctions that are relevant in the particular organisation(s) studied. Greater effort should be devoted to assessing and reporting the relevance, reliability, and validity of measures of evaluative style.

Second, important issues that have been raised in earlier studies have been neglected in later studies. For example, in his pioneering study on the relationship between evaluative style and managerial attitudes, Hopwood distinguished three evaluative styles. One of these styles, which Hopwood labelled "the Non-accounting style", was the most reported style in his study (44% of all respondents), but was treated as a residual category. Hopwood (1973, p.20-21) concluded: "With the benefit of hindsight, the failure to investigate the full nature and scope of the Non-accounting style was most likely a mistake. It may well include the most interesting approaches to the use of accounting information and some of the more promising means for improving practice." Despite this conclusion, evaluative styles in which accounting information is relatively unimportant have been neglected in later research. As a second example, Otley's (1978) sample consisted of 41 operating unit managers selected from three geographically different groups within one single company. Each group was headed by one group manager. His results indicated that 'despite considerable differences in evaluative style perceived by managers in each group, there were significant differences between groups'. Hopwood (1973) reported similar results. Thus, the limited evidence available (Hopwood, 1972a; Otley, 1978) suggests that differences in perceived evaluative style partly result from differences at the group level (between-group differences), and partly from differences at the level of individual subordinates (within- group differences). This raises level-of-analysis issues that have been largely ignored in subsequent research.

Third and finally, existing studies have failed to contribute to an understanding of either the process by which evaluative style is related to subordinates' managerial behaviour or the process by which evaluative style emerges (Briers & Hirst, 1990). This can be largely explained by the limited research methods that have been used in the RAPM literature. Many findings are based solely on cross-sectional survey data. Briers and Hirst (1990, p. 394) however note, *'while survey methods have their advantages, our limited knowledge about the processes involved in both the emergence of supervisory style and their associated effects, coupled with the sensitive nature of the phenomena in focus, suggests that descriptive case studies are likely to be beneficial.'*

1.2 Research relevance

Despite the numerous problems with the RAPM-literature, research on behavioural aspects of performance evaluation is “no less important today” than it has been since the 1950s, as Otley & Pollanen (2000, p. 495) argue:

“Budgetary control systems still form a major plank of the control apparatus of many organisations, although there are also indications that their role is changing. The stream of work on the impact of budget use on managerial behaviour and performance that has emanated from Hopwood’s seminal (1972a) study has made an important contribution to understanding the roles that budgetary information can play in performance evaluation. .. [...] The major contribution made by the behavioural accounting literature since its inception in the 1950s has been to draw attention to the impact of accounting systems on human behaviour, and vice versa. This is no less important today, and the new control mechanisms being adopted by organisations will provide a rich variety of practice which accounting researchers can both learn from and contribute to.”

To better understand the relationship between evaluative style of superiors and subordinate managers’ behaviour, however, future research on this relationship has to deal with the problems that are apparent in existing RAPM-studies.

1.3 Research purpose and method

This thesis addresses the relationship between evaluative style of superiors and subordinate managers’ behaviour. The first purpose of the research reported here is to contribute to a better understanding of the concept of evaluative style, taking into account the shortcomings of, criticisms on and concerns raised in the existing RAPM-literature. A good understanding of the concept of evaluative style is a necessary condition to achieve the second aim of this study: to contribute to a better (theoretical) understanding of how and why evaluative style is related to subordinate managers’ behaviour. This study consists of a theoretical and an empirical part. The theoretical part of this research entails an extensive review and analysis of the RAPM-literature in the light of contemporary developments within the management control field. This review results in a number of recommendations for future research, and the development of a preliminary theoretical framework. The empirical part of this research consists of an exploratory field study that has been conducted within twelve units of a contemporary Dutch food-processing company. Data in the field study have been obtained from documents, personal observations, interviews with all twelve superiors and two or three of their subordinate managers (27 in total), and a written questionnaire that was sent to all subordinates (n=57) who reported to one of the twelve superiors.

1.4 Overview and structure of dissertation

As indicated above, this thesis aims at improving the understanding of how and why evaluative style is related to subordinates' managerial behaviour. The lack of understanding, despite a mass of literature on this topic, originates from the many problems with existing RAPM-research. Chapter 2 provides a synopsis of the RAPM-literature, describes the problems and unresolved issues within this research field, and discusses the implications for future research on the relationship between evaluative style and subordinates' managerial behaviour. First, however, readers are shortly introduced to the concepts of management control, management accounting and behavioural accounting, as these concepts are important to understand the emergence and relevance of the RAPM-literature within the management control and accounting literature. The overview of the RAPM-literature indicates the need for further research on evaluative style that, first, devotes greater effort to assess and report the meaning, relevance and validity of measured differences in evaluative style in contemporary organisations (first purpose) and, second, aims at further development of theory on the relationship between such differences in evaluative style and subordinates' managerial behaviour (second purpose). However, the second theme (improvement of theory on the relationship between evaluative style and subordinates' managerial behaviour) cannot be addressed without a thorough understanding of the construct of evaluative style and the problems associated with its measurement. Therefore, I will elaborate the first theme, the construct measurement of evaluative style, in chapter 3 and the beginning of chapter 4 before I investigate the relationship between evaluative style and subordinates' managerial behaviour in the remainder of chapter 4.

To gain an understanding of differences and similarities in conceptualisations of evaluative style in the RAPM-literature, and the implications for future research, in chapter 3 I will review the existing measures of evaluative style. I will give particular attention to (a) the underlying concept of evaluative style operationalised by each measure, (b) the measurement tests carried out to validate each measure, and (c) validity problems with each measure. This review will result in a number of implications for the measurement and conceptualisation of evaluative style in future research. To assess the relevance of the conclusions and the applicability in practice I conducted a pilot field study. The findings will be reported in chapter 3 too.

After I have addressed the construct of evaluative style, in chapter 4 I will develop a theoretical framework that links evaluative style to other variables. First, I will focus on the relationship between evaluative style and variables that may explain the occurrence of differences in perceived evaluative style, i.e. antecedents of evaluative style. As indicated in the introduction, in addition to problems with the relevance, reliability and validity of measures of evaluative style, later RAPM-studies have ignored level-of-analysis issues that were raised in the earlier studies. An important topic, which also emerged from the pilot

study in chapter 3, is the level of agreement on the perceived evaluation style of their superior among subordinates that are evaluated by the same superior. This raises questions such as whether perceived evaluative style is a group level or individual level phenomenon, and why differences in perceived evaluative style between and within groups emerge. Based on a review of RAPM-literature, I will identify contextual factors that, at different levels of analysis, could explain why subordinates perceive differences in evaluative style. I will also give attention to the relationship between leadership style and evaluative style. Next, I will review the RAPM literature to identify variables that may help explain the relationship between evaluative style and subordinates' managerial behaviour, i.e. consequences of perceived evaluative style¹. Particular attention will be given to the role of fairness of evaluation and trust in superior as intervening variables between perceived evaluative style and outcome variables such as job satisfaction and job related tension.

While chapter 2, 3, and 4 lay the theoretical² foundation for an in-depth field study on evaluative style, the remainder of this study (Chapter 5 through Chapter 8) describes the empirical part of this research project: an in-depth field study conducted within a large Dutch food-processing organisation.

In chapter 5, I will describe and explain the research design of the field study, including the interview protocol, variable measurement through the use of a written questionnaire and the sampling of respondents. While chapter 5 is mainly descriptive, chapter 6 through chapter 8 present the findings from the field study.

Chapter 6 provides a general description of the organisation and a more detailed description of the evaluative system that is used within this particular organisation, to provide an organisational background against which the findings in chapter 7 and 8 should be interpreted.

Next, in chapter 7, I will explore the concept of evaluative style in the context of the particular organisation (purpose 1), as a clear understanding of how differences in evaluative style are measured and conceptualised in this field study will be necessary before it is possible to explore how and why such differences in evaluative style relate to (differences in) subordinates' managerial behaviour (purpose 2) in chapter 8. The exploration of the concept of evaluative style in chapter 7 is based on evidence from both interview data and written questionnaire data. In the written questionnaire, subordinates were asked to indicate the

¹ The terminology of antecedents and consequences is used mainly out of convenience. The terms suggest causal relationships. It should be noted, however, that I will not actually test the causality between two variables, but will focus on associations only.

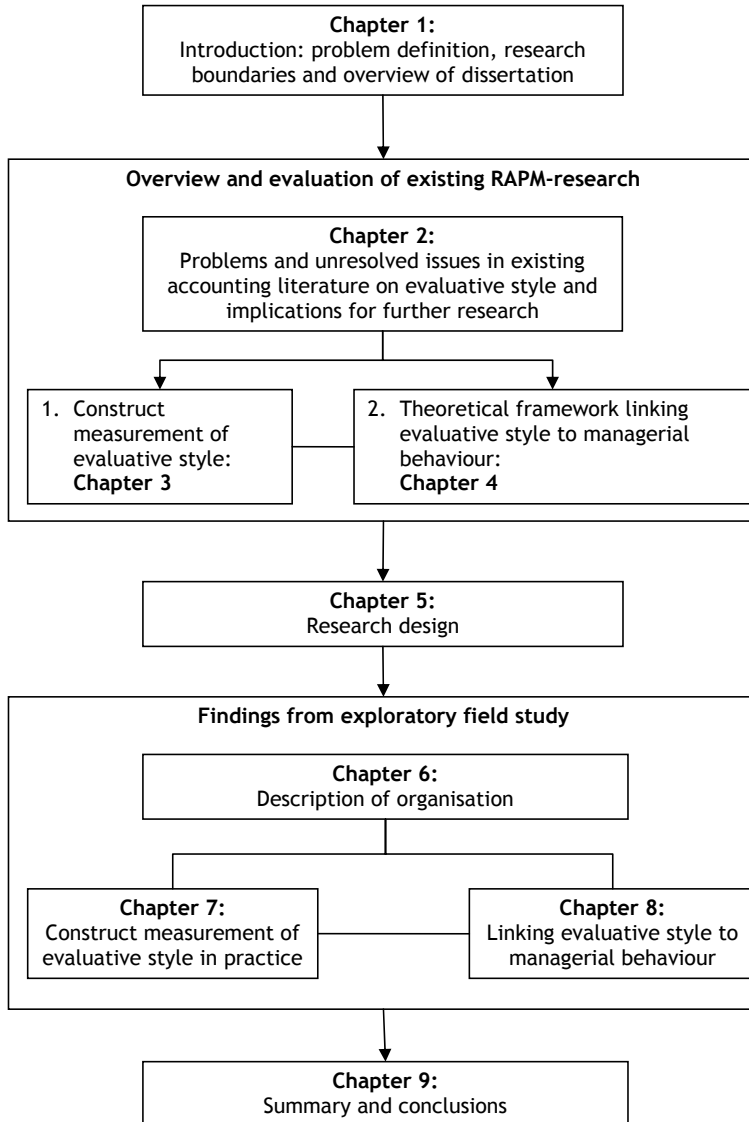
² Although these chapters are mainly theoretical, the theory is complemented with some empirical data from a pilot study, in which eight interviews were held in two units of a Dutch Service organization.

importance that their superior attaches to 17 different aspects of the performance evaluation process. I will explore whether these 17 aspects capture relevant differences in evaluative style in four different ways. First, I will address the relevance of the 17 aspects to the respondents. Second, I will analyse the underlying structure of the 17 aspects of perceived evaluative style to explore whether there are underlying dimensions of evaluative style. Third, I will explore whether there are significant differences in evaluative style between groups. Fourth, I will explore the level of within-group agreement in perceived evaluative style.

In chapter 8 I will explore antecedents of consequences of perceived evaluative style within the context of the particular organisation. This exploration will be based on the propositions that I have developed in chapter 4, although not all propositions from chapter 4 will be explored empirically. Where necessary, the propositions will be tailored to the empirical dimensions of evaluative style as they emerged from the analysis in chapter 7. In chapter 8 I will thus address the same three topics as in chapter 4. The first topic is the relationship between contextual factors and perceived evaluative style. This exploration will be based on data from the written questionnaire (provided by subordinates), and data from interviews with subordinates and superiors. The second topic is the relationship between leadership style and perceived evaluative style. I will examine this topic using data from the written questionnaire. The third topic is the relationship between perceived evaluative style and subordinate managers' behaviour. I will investigate this relationship first by analysing data from the written questionnaire, and next by analysing the interviews data, particularly the data obtained from the subordinates.

Chapter 9 summarises the main findings from this study. Chapter 9 also discusses the limitations and shortcomings of this research project, as well as the major contributions to existing literature. Finally, several opportunities for further research are stated.

Figure 1.1 Structure of dissertation



2 RAPM-research: an overview and evaluation³

2.1 Introduction

This chapter will provide a brief overview of the RAPM-literature (§ 2.3), describe the main problems with and criticisms vented on existing literature (§ 2.4), and discuss the implications for future research on the relationship between evaluative style and managerial behaviour (§2.5). Finally, the main conclusions from the chapter as to the issues that will be explored in the remainder of this thesis are summarised in § 2.6. First, however, in § 2.2, readers are shortly introduced to the concepts of management control, management accounting and behavioural accounting, as these concepts are important to understand the emergence and relevance of the RAPM-literature within the management control and accounting literature.

2.2 The traditional view of management control, management accounting and managerial behaviour

2.2.1 Introduction

In this section, I will describe and define management control, management accounting, and behavioural accounting in order to provide the necessary background for the critical review of the literature on evaluative style in the remainder of this chapter. Both the description of management control and management accounting will be based on the "traditional" view, since this view has been dominant in the RAPM-tradition, as I will prove in following sections. The reader should be aware beforehand that this traditional view has been challenged during the last twenty years in many ways.

2.2.2 Anthony's definition of management control

One of the most influential authors on management control over the years has been Robert Anthony. For the description of the traditional view on management control, therefore, consideration should be given to his framework on management control. This framework, which was developed in 1965 and was updated in 1988, has dominated much of the research on management control and has been mentioned in many books and articles, as Anthony (1988) himself notices in the foreword to the 1988 edition of his book. The classical definition of management control, originating with Anthony, is: *Management control is the process by which managers assure that resources are obtained and used effectively and*

³ Parts of this chapter have been published as: Noeverman, J. (1996), The Behavioural Impact of Budgets on People: an Analysis of the Theoretical Development in the Accounting Literature, In: C.M.T. Boneco, A. de Bos, C.D. Knoops, *FMA-Kroniek 1996* (pp. 485-502), Samsom.

efficiently in the accomplishment of the organization's goals' (Anthony, 1965; Otley & Berry, 1980; Machin, 1983). In 1988, however, Anthony (1988, p.10) defined management control slightly different as *'the process by which managers influence other members of the organization to implement the organization's strategies'*. This definition has two important implications.

First, management control is primarily seen as a behavioural process: managers use management control to influence other members of the organization. This influencing is not done arbitrarily, but with the purpose to implement the organization's strategies. This brings me to the second implication: management control is aimed at attaining the organizational strategies and goals. Ideally, then, these two implications are intertwined: managers succeed in influencing other members in the organization in such a way that organization's strategies are attained. Yet, it is possible that managers influence their members in such a way that it induces them to take actions and decisions that are detrimental to the organization. Thus, one of the main concerns about designing and using management control systems is *how* organization members can or should be influenced in such a way that they indeed act and make decisions in the best interest of the organization and its stakeholders. Before giving specific attention to this behavioural and motivational aspect, first some consideration will be given to the role of management accounting within the traditional management control system.

2.2.3 Management control and management accounting

Traditionally, management accounting and management control are closely related. Yet, they are not identical. This will become clear by looking at some of the definitions of management accounting. The National Association of Accountants (1990, p.4) for example, defines management accounting in SMA 1A as *'the process of identification, measurement, accumulation, analysis, preparation, interpretation, and communication of financial information used by management to plan, evaluate, and control within an organization and to assure appropriate use of and accountability for its resources'*. Similarly, Horngren, Foster and Datar (1994, p.4) describe management accounting as follows: *'Management accounting, focusing on internal customers, measures and reports financial and other information⁴ that assists managers in fulfilling goals of the organization.'* From these definitions we can conclude, that, while management control focuses on attaining the organization's goals and strategies, through planning, evaluating, and controlling activities, management accounting is focused on preparing, measuring, and reporting the financial information that is needed to practise these functions efficiently and effectively. Thus, management accounting provides tools for management control. This explains why many accounting textbooks contain a

⁴ That 'other information' is included in the definition is an important difference with the definition of the NAA. However, the issue whether management accounting is or should be limited to financial information, will be implicitly discussed later in this chapter.

section on management control. In such sections, attention is usually given to topics like responsibility accounting, budgeting, standard costing, decentralization and transfer pricing (see for example Horngren & Sundem, 1993; Kaplan & Atkinson, 1989; Horngren, Foster & Datar, 1994; Hansen & Mowen, 1994).

That management accounting and management control are closely related does not become clear solely by looking at the definition of both concepts, but also by examining how management control has been described by Anthony (1988), and, subsequently, in many accounting textbooks. In describing the traditional view on management control, a distinction can be made between the management control structure and the management control process; the structure is what the system is, the process is what it does (Young, 1994).

Traditionally, the main characteristic of management control *structure* is responsibility accounting: the organization is organized in different sub-units, which are responsible for either costs, revenues, profit, or profit/investment. Depending on their responsibilities, the sub-units are respectively referred to as cost center, revenue center, profit center, or investment center^{5,6}. Other parts of the management control structure are rules and procedures, and organizational culture (Anthony, 1988).

The management control *process* is usually divided into four phases: programming, budgeting, execution and evaluation. Programming deals with formulating a long-range plan, a program. This program is often referred to as the strategic plan. Yet, the programming process is not identical with the strategic planning process. In practice, it is often impossible to separate these two processes. Conceptually, however, programming should be distinguished from strategic planning, because strategic planning is unsystematic and programming is systematic. While strategic planning is concerned with formulating goals and strategies, programming takes these goals and strategies as given and deals with the effective implementation of the strategies (Anthony, Dearden & Govindarajan, 1992). The program serves as a point of departure for the formulation of annual budgets. Budgets are plans for the coming period (usually a year), expressed in quantitative terms (not per se financial) (cf. NAA, SMA 2, 1990, p.11; Horngren, Foster & Datar, 1994). During the year, these plans are being executed: the most important functions of budgets in this stage are coordination, communication and motivation. During the period and at the end of it, actual performance has to be evaluated. Usually actual performances are evaluated against the pre-set plans. Comparing actual and budgeted results could result in variances which should be

⁵ For criteria when to use which responsibility center, see Kaplan & Atkinson, 1989, pp.529-533.

⁶ A recent study in the Netherlands has shown that most organizations in the Netherlands use a responsibility structure consisting of profit centers (De Wit & Van der Woerd, 1994).

explored and in corrective actions being taken, and in new plans or adjusted plans to be used in the next period. This control cycle, in which control is attained through setting budget-goals (plans), execution of these plans, measurement and evaluation of actual performance, and taking corrective actions if necessary, eventually resulting in new or adjusted plans, is referred to as budgetary control.

Although the control cycle is similar for different types of responsibility centers, the kind of accounting information used will differ, depending on the type of responsibility center. In cost centers, for example, managers will use cost information to monitor and assess operations. Thus, an important aspect of the evaluation of the performance of cost centers and their employees is comparing actual costs with target or standard cost levels for the actual output. In investment centers, managers and other employees are held responsible for revenues, costs, and the level of investment, as if the unit is an independent business. An important performance measure for evaluating performance in investment centres is return on investment (ROI), relative to some target, or a profit measure that adjusts profit for the level of investment made and the cost of capital (such as residual income or EVA).

For many years, in the accounting literature budgetary control has been the dominant control mechanism. A management control system in this view is essentially seen as a rather simple, cybernetic system, involving a single feedback loop, much like a thermostat (Anthony, 1965; 1988).⁷ Yet, Anthony (1965; 1988) recognizes that, unlike a thermostat, management control cannot be applied mechanically, because organisations involve humans. As a management control mechanism, budgetary control is part of *"the process by which managers influence other members of the organization to implement the organization's strategies"*⁸. The involvement and judgement of humans, especially in the administration of budgets, is an essential part of budgeting. For example, managers have to decide whether and which performance measures could be used to motivate others to implement the organization's strategies. Managers face decisions on what constitutes an appropriate standard for the chosen performance measures. When exploring variances, managers are faced with multiple possible causes, some within and some outside the responsibility and controllability of their subordinates. Furthermore, managers have to decide whether meeting or not meeting budget-standards affects the realisation of overall goals and strategies of the organisation, and if so how and to what extent. The RAPM-literature, although mainly

⁷ Over the years, this traditional view of management control has evolved to a broader concept. In this broader view, management control includes all the devices or systems managers use to encourage, enable or, sometimes, force employees to act in the organization's best interest (Merchant & Van der Stede, 2003). In this view, management control systems do not focus on measured performance only, although performance measurement and management still forms a major part of the management control literature. Management control in the modern view includes other mechanisms or procedures such as a code of conduct, direct supervision, human resource practices etc. Furthermore, in some views, management control is not involved with strategy implementation only, but with strategy formulation and adaptation as well (Simons, 1996).

operating within the performance-oriented, cybernetic view of management control, has focused on this human aspect of budgeting, by studying the consequences of different evaluative styles on subordinate managers' behaviour.

2.3 An overview of RAPM-literature

A major contribution to the literature on the consequences of superior's use of accounting information in the performance evaluation of subordinates, has been made by Hopwood (1973). Hopwood (1973, pp.53-54) identified and described three different evaluation styles: a Budget-Constrained style, a Profit-Conscious style and a Non-Accounting style. The distinction between these styles was based on two broad factors. The first factor was the extent to which accounting information was used at all for the purpose of performance evaluation. In one of the three styles, accounting information played a relatively unimportant role in performance evaluation, at least in the perception of subordinates, which was designated as a Non-accounting evaluative style accordingly. The second factor distinguished between an emphasis on the underlying organisational purposes served by the information as against an emphasis on the information itself. Hopwood used this factor to distinguish between the other two styles, which were really different from one another, even though they were both accounting-oriented evaluative styles. Hopwood referred to the evaluative style emphasising accounting information itself as a Budget-Constrained style, i.e. budgetary information was taken at face value and rigidly used in evaluating the performance of subordinate managers. The other style, with a more flexible orientation towards budgetary outcomes, interpreting these outcomes and subordinate performance against the organisational objectives and context, was called a Profit-Conscious style. To measure a supervisor's style of evaluation, Hopwood asked cost centre heads to indicate three items from a list of eight that they thought to count most in how their departmental supervisor evaluated their performance. They were asked to put the number 1 next to the most important item, the number 2 to the next important item, and, finally, the number 3 next to the item that counts third most (Hopwood, 1973, p.208). The superior's evaluation style was considered to be a Budget-Constrained style when meeting the budget ranked among the top three criteria, but concern with costs was not. When both meeting the budget and concern with costs ranked among the top three criteria, the evaluation style was specified as a Budget-Profit style. Later in his research, Hopwood combined the Budget-Constrained style and the Budget-Profit style into one style, because results for these two styles did not differ significantly. A style was designated as a Profit-Conscious style when concern with costs ranked among the top three criteria, but meeting the budget not. Finally, when neither meeting the budget nor concern with costs ranked among the top three criteria, the superior's style was classified as a Non-Accounting style.

⁸ This is Anthony's (1988) definition of management control, quoted on p. 9 of this thesis.

Limiting his research to one large US company, Hopwood investigated and confirmed the hypotheses that a Budget-Constrained style leads to higher job-related tension, higher data manipulation and lower social relations with peers and superiors than either a Profit-Conscious or a Non-Accounting style. From these results Hopwood hypothesised that a Budget-Constrained style would lead to lower performance than the other two styles, although he did not empirically test this hypothesis.

For a good understanding of his results, it is important that Hopwood (1973, p. 15) begins his pioneering study with pointing out several shortcomings of accounting measures.

Firstly, accounting measures are unable to capture all social behaviours that are essential to organisations. This shortcoming reflects that accounting measures can only measure economic performance or the economic aspects of human behaviour.

Secondly, accounting reports represent outcomes, while managerial activity is concerned with the processes that ultimately result in financial outcomes. The relation between effort and outcome often is influenced by environmental organisation technological factors, or by the behaviours of other managers, implying that in many cases the accounting reports do not reflect managerial performance adequately. For this reason, many accounting textbooks argue that for a fair evaluation, the controllable component of the reports should be isolated. However, it may be impossible to determine exactly whether outcomes are controllable by manager's efforts or not, which is a third shortcoming. If we knew, Hopwood argues, would it not imply a certainty that would radically change the role of the manager? Fourthly, accounting measures can only approximately represent the full complexity of any organisation's economic structure.

Fifthly, accounting measures can only approximately reflect performance in departments, which have to deal with complex interrelated activities.

Sixthly, accounting measures emphasise short-term results while the evaluation of managerial performance is often concerned with more long-term considerations.

Given these characteristics, Hopwood argues, subordinates may be inclined, and have the possibility, to manipulate accounting numbers in such a way that they will reflect a more favourable performance, without increasing the effectiveness of the organisation. Yet, such behaviour is only of value when performance is evaluated on the basis of a rather rigid interpretation of the accounting information, i.e. when the superior is using a Budget-constrained style. When working under a Profit Conscious supervisor, favourable short-term indices do not necessarily imply favourable evaluation, since the indices are interpreted in terms of their effects on organisational purposes, in the long run. Instead, the behaviours may lead to a rather unfavourable evaluation under a Profit Conscious style. In addition, since in a Non-accounting evaluation accounting information is not of primary importance, improving short-term accounting reports does not have direct value for the subordinate.

Therefore, Hopwood concludes that 'manipulation of the accounting information is seen as a

response to a particular use of the information rather than to the characteristics of the information itself.' (p. 24-25) This conclusion however is based on Hopwood's assumption that accounting information is inevitably inaccurate (Hirst, 1981; 1983). Otley (1978) questions this assumption in a later replication of Hopwood's study.

According to Otley (1978), distortion of information may occur for two reasons. First, because inadequacies exist within the system. In this case, the information provided by the accounting system does not adequately reflect the complexity of underlying organisational and economic events. Secondly, because personal goals are not aligned with organisational goals. This may provide a reason for individuals to manipulate reports so that they reflect their contribution to organisational performance more favourably. The main reason for Otley's (1978, p. 122) study is that 'it is important to know whether distortion of accounting information is inevitable and can therefore be limited only by stricter methods of audit and control, or whether it depends upon precisely how accounting information is used within an organization.' In Hopwood's study on this issue, unfortunately, technical characteristics of the accounting system on the one hand and the way in which managers use the accounting information on the other hand have been confounded. Otley's study therefore 'was designed to eliminate technical failings in the accounting system, as far as possible, by observing the operation of a well-designed system in a type of organization that was well suited for the application of budgetary control.' (p. 123) Therefore, Otley focused primarily on differences in outcome that occur due to the different use that managers make of accounting information. He extended the analysis to performance. However, Otley's findings not only failed to support Hopwood's findings that a Budget-Constrained style was associated with higher job-related tension and lower social relations than the other two evaluative styles, but contrary to Hopwood's hypothesis Otley also found that a Budget-Constrained style was associated with higher managerial performance than the other evaluative styles.

It is highly unlikely that these contradictory findings are caused by differences in the measurement of evaluative style, since Otley's instrument to measure evaluative style was a carefully revised and slightly adapted version of Hopwood's (1973) instrument described above. The revisions and adaptations were made to suit the particular organisation studied, while at the same time making sure that the same construct of performance evaluation style was measured as in Hopwood's study. There are no signs that there were important differences between the two studies in the measurement of evaluative style. Therefore, in either study, the difference between Budget-constrained and the Profit-conscious style was most likely the same. This difference existed in a difference in how the superior used accounting information when evaluating performance. A supervisor with a budget-constrained style used this information rigidly, emphasising short-term results, while a supervisor with a profit-conscious style used accounting information in a more flexible manner, considering long-term objectives. Since the styles in both studies were essentially

the same, differences in the findings from both studies can be explained only by taking into account systematic differences in the organisations studied by Hopwood (1972a) and Otley (1978). Otley's failure to replicate Hopwood's finding that a budget-constrained style was associated with higher tension and with dysfunctional behaviour in a different context thus suggests that Hopwood's findings are conditional on the organisational context, as Otley concluded.

Several studies have attempted to reconcile the contradictory findings based on a contingency approach (cf. Ezzammel & Hart, 1987; Ross, 1995). Some context factors studied within the RAPM-literature which may have caused the contradictory findings, followed by the hypotheses associated with these factors, are:

- task uncertainty: the higher the task uncertainty, the more likely it will be that use of a Budget-Constrained style and rigid performance measures has detrimental effects (Hirst, 1981; 1983; 1987).
- environmental uncertainty: the higher the environmental uncertainty, the more likely it will be that use of a Budget-Constrained style and rigid performance measures has detrimental effects (Govindarajan, 1984; Brownell, 1985; 1987; Kren, 1992).
- interdependence between responsibility centers: the higher the interdependence between responsibility centers, the more likely it will be that use of a Budget-Constrained style and rigid performance measures has detrimental effects (Otley, 1978).
- culture: the lower power distance and the higher individualism⁹, the more likely it will be that use of a Budget-Constrained style and rigid performance measures has detrimental effects (Harrison, 1992; 1993; Lau, Low & Eggleton, 1995; Otley & Pierce, 1995).
- locus of control: the less subordinates feel that they can influence outcomes (have control over their environment), the more likely it will be that use of a Budget-Constrained style and rigid performance measures has detrimental effects (Brownell, 1981; 1982b; Frucot & Shearon, 1991).
- budgetary participation: the role of budgetary participation in understanding the relation between evaluative styles and subordinates' behaviour is still unclear, despite the many studies undertaken (Brownell, 1982a; Brownell & Hirst, 1986; Brownell & McInnes, 1986; Brownell & Dunk, 1991; Dunk, 1990; 1993a; 1993b; for a review of this literature see Greenberg, Greenberg & Nouri, 1994 and Van Dijk, 1992). One hypothesis is that the higher participation of subordinates in setting (budgetary) targets or goals, the less likely it will be that use of a Budget-Constrained style and rigid performance measures has detrimental effects (e.g. Brownell, 1982a; Brownell & McInnes, 1986; Dunk, 1993a). However, another hypothesis is that the higher budgetary participation the higher subordinates' propensity and possibility to create slack in their budgets will be, especially

⁹ Power distance and individualism are two of Hofstede's (1980) "dimensions" of culture. Hofstede distinguished two other dimensions, uncertainty avoidance and masculinity.

when a Budget-Constrained style is used in a highly uncertain environment (e.g. Brownell & Hirst, 1986; Brownell & Dunk, 1991). It is not surprising, therefore, that findings with regard to the consequences of budgetary participation have been ambiguous and contradictory.

Through all these studies, taking into account many different factors and/or many different combinations of factors, it is now widely accepted among researchers that negative consequences do not arise because of the use of accounting information in performance evaluation per se, but rather that the manner and the circumstances in which superiors use accounting information to evaluate performance largely affects performance and may cause tension and dysfunctional behaviours to occur (Hopwood, 1974; Macintosh, 1985; Ross, 1995, p. 9). So, as a whole, the existing literature suggests that although the manner and extent in which the superior uses accounting information in the performance evaluation of subordinates affects job-related tension, performance and dysfunctional behaviours, there is much support for the hypothesis that the effects of particular evaluative styles are conditional on organizational, environmental and personal factors. However, these studies fail to make clear how and when accounting information can be efficiently and effectively used for control purposes, while at the same time minimizing detrimental effects. Despite almost three decades of research, a conceptual framework linking evaluative style to attitudinal variables such as job related tension, performance and dysfunctional behaviour is still missing. This is not surprising, since there are numerous problems with the accounting literature on evaluative style, as will become clear by a review of the many criticisms on the RAPM-literature.

2.4 An evaluation of RAPM-literature: a review of criticisms

2.4.1 Introduction

The last few years the number of criticisms on the RAPM-literature has grown rapidly. In 1990, Briers and Hirst published a comprehensive review of the RAPM literature in which they raised a number of questions regarding the relevance and reliability of the survey-based research on the behavioural implications of evaluative style following Hopwood's (1973) and Otley's (1978) seminal studies. In particular, Briers and Hirst (1990) criticise the confusing conceptualisation of evaluative style, the lack of theoretical justification of including additional variables, and the uncritical use of survey research. However, little attention has been paid to the critical remarks that Briers & Hirst made. Instead, researchers carried on along the same lines as before. Yet, in 2000, in a single issue of the leading accounting journal in the RAPM field, *Accounting, Organisations and Society*, four critical reviews appeared, showing the problematic state of the stream of literature following Hopwood (1973) and Otley (1978). Hartmann (2000), Otley & Fakiolas (2000), Otley & Pollanen (2000), and Vagneur & Peiperl (2000), cast detailed criticisms on the conceptualisation and measurement of evaluative style, on the research method, and on the theoretical

Table 2.1 Criticisms on RAPM-literature

1.	Measurement of evaluative style	
1.1	Concept of evaluative style	<p>Hartmann (2000)</p> <p>Despite the central role of the RAPM construct, RAPM studies show no unity in its empirical measurement. (p.466)</p> <p>Otley & Fakiolas (2000)</p> <p>The same terminology [RAPM] has been used for different concepts. (p.504)</p> <p>Different modifications represent not only differences in measurement, but also differences in concept.</p> <p>Otley & Pollanen (2000)</p> <p>The same words are sometimes used to describe quite distinct concepts. (p.495)</p> <p>Vagneur & Peiperl (2000)</p> <p>The more the variables were modified in later research, the further they appear to have moved from capturing the same dimension that Hopwood originally captured. (p.523)</p>
1.2	Content validity of evaluative style	<p>Otley & Fakiolas (2000)</p> <p>Relevance of the questionnaire items to the managers to whom they were administered was not explicitly reported upon in some cases, and there is therefore no evidence that the items had the same meanings attributed to them by the managers surveyed. (p.503)</p> <p>Validation of the scales to reflect the terminology within the organisation(s) studied, as both Hopwood and Otley did, has become neglected.</p> <p>Assessment of the validity of instruments lacks any systematic attention. (p.507)</p> <p>Hartmann (2000)</p> <p>In none of the studies that modified the instrument to measure evaluative style, formal proof is provided for the equality or validity of the instruments used. (p.466)</p> <p>Vagneur & Peiperl (2000)</p> <p>The content validity of the modified response sets used in studies following Hopwood and Otley is a matter of speculation. It will have to be assumed that these modifications were intended to provide improved content validity. However, whether this was accomplished by inductive or deductive means is unclear. [...] Unless appropriateness of a content set has been empirically determined before data collection, ex-post, its validity must be a matter of speculation. (p.516)</p>

Table 2.1 (continued)

2. Research method		
2.1	Variable measurement and validity	Otley & Pollanen (2000) the same variable (not just evaluative style, JN) is measured in importantly different ways across studies (p.483);
2.2	Sampling	Otley & Pollanen (2000) different samples of managers (from different types of organisations; from different cultures). (p.483) small sample sizes. (p.483)
		Hartmann (2000) The typical small sample sizes below 50 provide very unstable results when used to test contingency hypotheses through statistically demanding interaction models. (p.469)
2.3	Data analysis	Hartmann (2000) Lindsay (1995) has recently criticized the dominance of tests for statistical significance for accepting or rejecting research hypotheses. (p.452) The unquestioned use of moderated regression analysis to test RAPM contingency hypotheses although alternative formats for testing such hypotheses are available. (p.469)
		Otley & Pollanen (2000) Prior work has concentrated on statistical significance of interaction effects alone without paying attention to the size of the substantive effect (p.493);
2.4	Research design	Hartmann (2000) Young (1996) points to the often careless use of questionnaire survey methodology, which is the typical data collection method in RAPM studies (p. 452).

Table 2.1 (continued)

3. Theoretical development		
3.1	Contingency approach	<p>Briers & Hirst (1990)</p> <p>Theoretical development has been piecemeal and has taken a secondary role to statistical analysis (...)¹⁰.</p> <p>Hartmann (2000)</p> <p>The uncritical use of contingency theory in RAPM research (p. 452).</p>
3.2	Lack of replication	<p>Hartmann (2000)</p> <p>Lindsay & Ehrenberg (1993, p. 224) even mention the RAPM paradigm as an example of an area of unsuccessful replications, and note: "Taken as a whole, this body of research, although typically 'interesting' in seeking to explain discrepancies, does not add up to a coherent body of knowledge or understanding. (p. 465)</p> <p>Young (1996, p. 55) makes a similar remark, as he notes that: "... 25 years of survey research has yet to yield a cohesive body of knowledge about management accounting and control practices within and across industries and nations. (p. 465)</p> <p>Otley & Pollanen (2000)</p> <p>no single study has been replicated (p.484, p.494); different studies use different subsets of variables (p.483);</p>

¹⁰ Quoted by both Otley & Fakiolas (2000) and Hartmann (2000).

development of RAPM-research. These and other criticisms are summarised in Table 2.1. In this section I do not intend to completely reproduce all criticisms on RAPM, since at this point the sole purpose of this review is to show that there are many problems associated with existing RAPM-literature. Therefore, I will review the latest criticisms only.

2.4.2 Criticisms on conceptualisation and measurement of evaluative style

Recently, Otley & Fakiolas (2000) and Vagneur & Peiperl (2000) have critically reviewed the conceptualisation and measurement of evaluative style in the RAPM-literature. Otley & Fakiolas divide RAPM studies into four groups, based on the similarity of the method used to measure evaluative style to that used by Hopwood (1973), and differences in conceptualisation and measurement of evaluative style are explored analytically. Vagneur & Peiperl distinguish five different approaches to measuring evaluative style in the RAPM-literature, based on what they call "calculation type". These calculation types range from the use of three categories based on ranking scores (e.g. Hopwood, 1973) to the use of a continuous variable based on the summed Likert importance scores on (the) two accounting items from Hopwood's (modified) list of eight items (e.g. Brownell, 1985; 1987) to the use of a continuous variable calculated by dividing the summed Likert importance scores on (the) two accounting items by the summed Likert importance scores on eight other items (e.g. Harrison, 1992; 1993). Since both studies discuss Hopwood's conceptualisation and measurement of evaluative style and later modifications or alternatives in the literature, there is no need to do the same here. For our purpose, i.e. to illustrate the problematic state of the RAPM-literature, it will suffice to limit the discussion to the conclusions from these two reviews. Firstly, many studies differ with respect to the number of items and/or the wording of items they use to measure evaluative style. Furthermore, even studies using the exact same items differ in whether they use ranking scores, absolute scores, contrasts or an aggregate score to identify different evaluative styles to be used in further analyses. Secondly, the differences in the measurement of evaluative style represent different underlying concepts of evaluative style. Indeed, based on inter-correlations of the five different calculation types in measuring evaluative style and the associations of measures of evaluative style with performance, Vagneur & Peiperl (2000, p.523) concluded that 'the more variables were modified for subsequent research, the further they appear to have moved from capturing the same dimension that Hopwood originally captured'. Thirdly, Hopwood's original instrument of eight items was carefully developed to measure three specific evaluative styles within a specific organisation. The distinction between and the detailed description of the three evaluative styles was based on interviews he held within the organisation prior to developing the list of items and the ranking procedure. All items matched the language used within the organisation, and the two items he used to differentiate between a Budget-Constrained, a Profit-Conscious and a Non-Accounting evaluative style (i.e., "meeting the budget" and "concern with costs") had distinctive meanings for the participants in the organisation. Hopwood used the ranking scores to

effectively categorise complex phenomena that he observed in the organisation. By using a relatively simple list of eight items in this way he successfully measured three evaluative styles with distinct characteristics and distinct behavioural consequences. However, the content and conceptualisation of the three evaluative styles were much more complex than the instrument to measure them. It is questionable whether the uncritical application of Hopwood's measurement of evaluative style in cross-sectional questionnaires in later studies is meaningful without prior investigation that any single item on the scale means the same thing in each organisation and that the terminology captures the same phenomena. Furthermore, the necessity and validity of later modifications, whether in the number or the wording of the items or in the method that is used to categorise evaluative styles, is lacking in these later studies¹¹.

2.4.3 Criticisms relating to research method

Besides the many problems and criticisms relating to the measurement of evaluative style, several researches have severely criticised different aspects of the research method used in most RAPM-studies. Three papers are of particular importance in this respect: Hartmann & Moers (1999), Hartmann (2000) and Otley & Pollanen(2000).

Hartmann and Moers (1999) argue that the statistical testing in many studies in this field is inappropriate, sometimes even incorrect, making interpretation and comparison of the findings from different RAPM-studies difficult, if not impossible. More specifically, from a review of 28 RAPM-studies, published in three high-quality accounting journals, Hartmann & Moers conclude that in 27 of these studies (!) at least one of six types of errors occur in the application of Moderated Regression Analysis (MRA) to test contingency hypotheses. These errors range from incorrect specification of the MRA-equation, statistically testing different hypotheses than the hypotheses specified, incorrect use of higher-order interactions, and incorrect testing of the nature of interactions to incorrect interpretation of main effects, and incorrect conclusions about effect sizes.

Otley and Pollanen (2000) replicate parts of five previous studies in a different context. They argue, differences between previous studies are not surprising because of the small and diverse samples, and the differences in variable measurement. Furthermore, they argue that universal findings should not be expected, since control practices differ across organisations, cultures and time. For that reason, they deliberately chose to replicate the five studies by measuring the exact same variables exactly the same way in a different organisation. Since their results differ significantly from previous findings, they suggest that this can validly be explained by the different organisation involved in their study.

Finally, although primarily concerned with the theoretical development of the RAPM-literature, Hartmann (2000: 452) refers to criticisms directed at data collection and data analysis: "Regarding data collection, Young (1996) points to the often careless use of

¹¹ I will amplify these issues in chapter 3.

questionnaire survey methodology, which is the typical data collection method in RAPM studies. Regarding data analysis, Lindsay (1995) has recently criticised the dominance of tests for statistical significance for accepting or rejecting research hypotheses."

Taken as a whole, these criticisms on the research method cast severe doubt on the interpretability, validity, and reliability of the research findings reported in many RAPM-studies.

2.4.4 Criticism relating to theoretical development

A third criticism, narrowly connected with the former two criticisms, has been directed to the lack of theoretical development in the RAPM-literature. Briers & Hirst (1990) were the first to point this out. However, little has changed since then, for recently Hartmann & Moers (1999), Otley & Fakiolas (2000) and Hartmann (2000) all refer to this criticism by Briers & Hirst. The lack of theoretical development is in part due to the lack of replication in RAPM (Otley & Pollanen, 2000), and the few studies that did aim at replicating previous findings have not been very successful (Hartmann, 2000).

Hartmann (2000) in particular deals with the theoretical flaws of the RAPM-literature. He argues that there is a need and an opportunity to enhance RAPM theory based on the concept of uncertainty. In many RAPM studies, the typical hypothesis predicts that uncertainty negatively affects the appropriateness of the reliance on accounting performance measures when evaluating subordinates. These negative effects are hypothesised because RAPM theory is largely based on role conflict and on the "controllability principle". When uncertainty increases, accounting performance measures are hypothesised to become less controllable, less complete and less relevant. However, Hartmann shows that the empirical evidence supports both a negative and a positive effect of uncertainty on the appropriateness of (reliance on) accounting performance measures. He proposes several possible ways to address what he calls "the uncertainty paradox" of RAPM research, one of which is that RAPM research becomes more explicit about the meaning of appropriateness. To serve this purpose, Hartmann suggests, possible frameworks could be based on either goal theory or 'equity theory'. Goal theory states that clear and specific goals and objectives lead to higher performance, which may explain the positive effects of the reliance on accounting performance measures when uncertainty is high, since accounting performance measures are relatively clear and specific. Equity theory is concerned with peoples' perceptions of "fairness" and "justice" in social relationships. "Fairness" has often been mentioned as an important reason for subordinates' resistance against 'uncontrollable' performance measures. Thus, equity theory could help explain the negative effects of reliance on accounting performance measures when uncertainty is high.

2.5 Implications of criticism for further research

2.5.1 Introduction

Most of the critical reviews of RAPM studies have lead to several research recommendations (e.g., Hartmann, 2000; Otley & Fakiolas, 2000; Vagueur & Peiperl, 2000) that are summarised in Table 2.2. In this section, I will describe the research implications of the criticisms on RAPM literature for the conceptualisation and measurement of evaluative style and for the theoretical development. These implications are stated in general terms, identifying the topics that need further exploration in the remainder of this thesis, both empirically and theoretically.

2.5.2 Construct measurement of evaluative style

To overcome the many problems with the existing RAPM-literature, first, future research should validate the measure of evaluative style that is used. Only if the validity and reliability of evaluative style measures is systematically evaluated, research findings on the relationship between evaluative style and managerial behaviours and attitudes can be (better) interpreted. The flaws in research on evaluative style are remarkably similar to the flaws in organisational strategy research identified by Venkatraman & Grant (1986). Venkatraman & Grant (1986) critically review the state of construct measurement in organisational strategy research, and some of their critiques, such as the inadequacy of validity assessment and the lack of operational definitions of the key concepts, seem to apply directly to existing RAPM-research as well. To illustrate the necessity for any research to pay close attention to construct measurement, they quote Peter (1979, p.6):

“Valid measurement is the *sine qua non* of science. In a general sense, validity refers to the degree to which instruments truly measure the constructs which they are intended to measure. If the measures used in a discipline have not been demonstrated to have a high degree of validity, that discipline is not a science” (Venkatraman & Grant, 1986, p.71)

From the recommendations in Table 2.2 we can conclude that to increase the validity of the construct measurement of evaluative style, the measures that are used to categorise evaluative styles should reflect the organisational language (Otley & Fakiolas, 2000) and should capture the most important elements in the performance evaluation environment (Vagueur & Peiperl, 2000). This may imply that new measures of evaluative style would be needed, although some of the existing measures may capture relevant underlying dimensions of evaluative style (cf. Otley & Fakiolas, 2000; Vagueur & Peiperl, 2000). An analysis of the concepts of evaluative style that underlie existing measures will be necessary, however, to explore whether existing measures could be relevant and valid in contemporary

Table 2.2 Research recommendations evolving from criticisms

1. Measurement of evaluative style		
1.1	Concept of evaluative style	<p>Otley & Fakiolas (2000)</p> <p>Otley & Pollanen (2000)</p> <p>Vagneur & Peiperl (2000)</p>
		<p>A number of dimensions to evaluative style are becoming apparent, and effort should be devoted to constructing better measurement instruments to detect these. (p.508)</p> <p>Further investigation of the original distinction made by Hopwood between short-term devices intended to improve reported budgetary results and a longer-term emphasis on real economic performance. (p.494)</p> <p>Further, this requires an analysis of the subtle differences in the conceptual bases of the different approaches. These include differences in the conceptualization of constructs, method, and research strategy and design, in particular the idea of performance evaluative style as a continuum rather than as a set of discrete categories. (p. 523); Their findings suggest that ratings-based methods reflect an increased complexity relative to the nominations-based methods. They argue that such complexity may be capturing more of the underlying essence of performance evaluation style. The larger number of significant associations between these variables and the performance measures suggests that this is in fact the case. If it is, then the divergence of the more recent research may be necessary and appropriate. (p. 523)</p> <p>Scales should reflect the terminology in use within the particular companies studied, comparable to how both Hopwood (1972a) and Otley (1978) designed their instruments (p. 506)</p> <p>Support the development of measurement instruments with field study and extensive pilot testing of the instruments. (p.507)</p> <p>The content validity of the instrument to measure evaluative style depends on the extent to which the question includes the most important elements in the performance evaluative environment. If important elements are missing, responses may reflect variation in less important alternatives, and thus may not adequately measure within-sample variation in evaluative style. (p. 516)</p> <p>Because ex-post the validity of content can only be speculated upon, researchers should report their method of validating the ability of the variable content to reflect the environment being classified. (p.523)</p>

Table 2.2 (continued)

2.	Research method	
2.1	Variable measurement and validity	
2.2	Sampling	Otley & Pollanen (2000) Restrict consideration to a single organisation, in order to be able to control for the fact that accounting controls are only a small part of a wider set of organizational controls that differ significantly across organizations. This enables the researcher to elucidate the impact of different accounting controls within their wider context. (p.494)
2.3	Data analysis	Otley & Pollanen (2000) More careful specification of two-way analyses conducted in a variety of settings, instead developing ever increasing number of interactions (p.494)
2.4	Research design	Otley & Pollanen (2000) Aim should be the development of understandings about the operation of organizational control processes, which requires a greater emphasis on case-based and longitudinal work (p.494) Take account of recent changes in management accounting practices within organizations, such as the replacement of large, divisionalized organizations by smaller, more focused organizations and the increasing evidence on the changing role of budgeting as a tool of managerial control. (p.495)
		Otley & Fakiolas (2000) Several changes suggest that the roles of budget-based control techniques are likely to be undergoing substantial change in response to these pressures. At the very least, the design of studies of the impact of accounting techniques in the performance measurement and management function should ensure that they are able to pick up such changing emphases rather than expecting the world to behave in the same way as it did in the early 1970s. (p.509)

Table 2.2 (continued)

2.4	Research design	Vagneur & Peiperl (2000)	Most Anglo-American companies use budget-centered performance measurements as a central feature of their management control systems. Since budgets are explicitly intended to influence individual decision making, evidence linking budgets and negative behavior or performance suggests that, in practice, the management of management controls may be important. Mainstream management theorists rarely consider control systems, yet variation in evaluative style may be an indicator of system factors which are weak or have failed, rather than a causal effect itself (Otley, 1978; see Vagneur, 1996). This is a potential research area that should be explored. (p. 523)
			It is also well worth exploring individual psychological responses to performance assessment and the nature of the systemic effects created by budgets and other formal and informal management control processes (e.g. reward, planning, training and information systems). This would require synthesis of two levels of analysis (individual and system), as well as consideration of psychology, organizational behavior, behavioral accounting and systems theory research. (p. 524)

3.	Theoretical development		
3.1	Contingency approach	Otley & Pollanen (2000)	More careful specification and measurement of variables when progressing along the line of the contingency-based approach. (p.495)
3.2	Broaden the use of theories	Otley & Pollanen (2000) Vagneur & Peiperl (2000)	Draw on more recent developments in social-psychological literature to improve the quality of research. (p.495) Reconciliation of differences in the conceptual base of scales to measure evaluative style would require consideration of both individual psychological responses to performance assessment and the nature of systemic effects created by budgets and other formal and informal management control processes (e.g. reward, planning, training and information systems). This presents a significant opportunity for further research drawing on psychology, organizational behavior and behavioral accounting research. (p.514)

organisations. Therefore, similar to Venkatraman & Grant (1986), in Chapter 3 I will review the construct measurement of evaluative style in the existing RAPM-literature and describe the implications for better construct measurement of evaluative style in future research.

2.5.3 Theoretical development

Despite the lack of consistent findings and a consistent underlying framework, in my opinion, the existing RAPM literature provides useful insights on factors that should be considered when exploring how and why evaluative style is associated with subordinate behaviour. To improve our understanding of the control processes underlying evaluative style and the relationship with other variables, field studies would be more appropriate than the use of survey research that has dominated the RAPM-literature. Field studies facilitate exploration of evaluative style and its relationship with subordinates' managerial behaviour in the broader context of organisational controls rather than studying accounting controls as isolated control tools as is usual in the RAPM-literature (Otley & Pollanen, 2000). Similar to Hopwood (1972; 1973) and Otley (1978), such a field study could combine interviews with written questionnaires. This enables triangulation of findings, to increase the validity of findings. Furthermore, by conducting interviews unresolved issues from the literature can be explored, such as the relevant underlying dimensions of evaluative style, while the use of questionnaires enables the collection of data on variables that have been consistently shown to be relevant in understanding the relationship of evaluative style to subordinates' managerial behaviour. According to Otley & Pollanen (2000), the inclusion of variables in the questionnaire should be more carefully specified and measured than in previous research. The use of theories from other fields, such as psychology and organisational behaviour, may be helpful in identifying the relevant variables and its measurement (Vagnuer & Peiperl, 2000; Otley & Fakiolas, 2000). In accordance with Hartmann's (2000) suggestion, relevant theoretical frameworks could include 'goal theory' and 'equity theory'. The development of a theoretical framework that links evaluative style to other variables will be the subject matter of Chapter 4.

2.6 Conclusion

In this chapter I have reviewed and evaluated the many criticisms on and problems within existing RAPM research to identify the issues that need to be addressed when studying evaluative style and its relationship with subordinates' managerial behaviour. For despite almost three decades of continuous research on this topic, there is still a lack of understanding of behavioural aspects of evaluative style. From this analysis two major issues have been identified that I will concentrate on in the remainder of this thesis:

1. the conceptualisation and measurement of evaluative style; and
2. the development of a theoretical framework to understand how and why evaluative style relates to subordinates' managerial behaviour.

A focus on the conceptualisation and measurement of evaluative style is necessary since many studies have used the instrument to measure evaluative style originally developed by Hopwood (1973), some (e.g. Ross, 1994; 1995) in an uncritical manner, others (e.g. Brownell (1985), Brownell & Hirst (1986)) in confusing ways by making such modifications -minor and major- to the instrument that it becomes uncertain to what extent this modified instrument measures the same or different constructs as those measured by the instrument in earlier studies. Furthermore, since existing RAPM-studies have uncritically used survey research as the main research method, not only do different measures of evaluative style most likely represent different concepts, the validity and meaning of each concept is unclear.

Therefore, future research should be more explicit in assessing and reporting the meaning, relevance and validity of the construct measurement of evaluative style in contemporary organisations. This is even more important because concepts of evaluative style in the RAPM-literature have been based on a rather narrow view of management control. This view is based on the assumption "that the budgetary control system is the main integrative control system in most business organisations." (Otley, 2001) There is increasing evidence that the usefulness of budgeting as a management control tool may be limited in contemporary organisations (see, for example, Bunce et al. (1995) and Marginson, 1999). This may imply that new measures of evaluative style would be needed. An analysis of the concepts of evaluative style that underlie existing measures will be necessary to explore whether existing measures could be relevant and valid in contemporary organisations. Such a review and the implications for better construct measurement of evaluative style in future research will be the theme of chapter 3.

Focusing on the further development of theory on the relationship between evaluative style and subordinates' managerial behaviour is necessary as the theoretical justification of many research models and hypotheses in RAPM studies and the further development of theory from research findings has been neglected. Future research should broaden the use of theory in order to specify which variables to include in the research project and to interpret the results. I will develop a theoretical framework that links evaluative style to other variables in chapter 4.

3 Construct measurement of evaluative style: a review and proposal¹²

3.1 Introduction

The last few years the number of criticisms on the RAPM-literature has grown rapidly. Hartmann (2000), Otley & Fakiolas (2000), Otley & Pollanen (2000), and Vagneur & Peiperl (2000) cast detailed criticisms on the conceptualisation and measurement of evaluative style, on the research method, and on the theoretical development of RAPM-research. Hartmann & Moers (1999; 2003) have argued that the statistical testing in many studies in this field is inappropriate, sometimes even incorrect, making interpretation and comparison of the findings from different RAPM-studies difficult, if not impossible. Recently, Dunk (2003) has argued that there are “far greater issues” to deal with than flaws in statistical testing. One of these “greater issues” I address here is the conceptualisation and measurement of evaluative style. In this chapter I review existing measures of evaluative style from studies in which the researcher(s) explicitly argue that they have measured a construct that is closely related or similar to that of Hopwood (1972a; 1973) and/ or Otley (1978). This decision rule was adopted in order to limit the review to measures of evaluative style that can be seen as central to the RAPM-field. Earlier reviews of measures of evaluative style in this area (see Otley & Fakiolas (2000) and Vagneur & Peiperl (2000)) have mainly focused on identifying the (dis)similarity in concepts, in particular as compared to Hopwood’s (1973) original conceptualisation. This review takes a different perspective. The main purpose of this chapter is to assess how future research can conceptualise and measure evaluative style in a relevant, valid and reliable way, and whether any of the existing measures can be used in future research, regardless of their (dis)similarity to Hopwood’s conceptualisation.

The chapter is organised as follows. First, existing measures of evaluative style in the RAPM-literature will be reviewed. This review focuses on the underlying concept, validity and reliability of these measures. Subsequently, I will evaluate the outcomes of the literature review, and discuss four important implications for future research on evaluative style. Finally, the conclusion of this chapter summarises my suggestions for the conceptualisation and measurement of evaluative style in future research.

¹² Parts of this chapter (§ 3.1 through § 3.4) have been published as Noeverman, J., B.A.S. Koene & A.R.T. Williams (2005), Construct Measurement of Evaluative Style: A Review and Proposal, *Qualitative Research in Management and Accounting*, 2 (1), pp. 77-107.

3.2 Review of existing measures of evaluative style

3.2.1 Introduction

In table 3.1 I have categorised existing studies, showing how each study has identified and measured evaluative style. The categorisation is adapted from Otley & Fakiolas (2000) and Vagneur & Peiperl (2000). To categorise studies, Otley & Fakiolas use the (dis)similarity to Hopwood's study as the criterion, operationalised as both the (dis)similarity in content and wording of the items used *and* (dis)similarity in calculation type¹³, while Vagneur & Peiperl use differences in calculation type as the sole criterion. Combining the categorisations from both reviews and extending these reviews with an evaluation of validity and reliability of existing measures presents a more detailed synopsis of differences and similarities in conceptualisation and measurement of evaluative style in the RAPM-literature. This provides a good basis to assess the relevance of existing operationalisations of evaluative style for use in future research. Studies that have used Hopwood's items -with or without modifications- have been categorised in column A, studies that use different items have been categorised in column B. Studies within each of the columns A and B have been further classified according to differences in calculation type, as defined by Vagneur & Peiperl (2000). As can be seen from table 3.1, ranking procedures are typically used in conjunction with the use of Hopwood's items, as none of the studies in column B have used a ranking procedure.

With respect to the measurement of variables in general, researchers need to address two questions: first, how accurate or precise are our measures¹⁴, and, second, how accurately are we measuring what we think we are measuring? The first question relates to the reliability of measures. A reliable measure is (relatively) free from measurement errors, yields consistent results when used repeatedly, and reflects only one underlying construct (uni-dimensional). The second question concerns the validity of measures. Considering this, the review of different measures of evaluative style from table 3.1, specifically addresses (a) the underlying concept of evaluative style operationalised by each measure, (b) the measurement tests carried out to assess the validity and reliability of each measure, and (c) (validity and reliability) problems with each measure. To limit the review, not all measures that are listed in Table 3.1 will be extensively reviewed. In most cases, a single, representative study from each cell has been chosen to discuss the underlying concept and (validity) problems.

¹³ "Calculation type" refers to the manner in which the items are used to calculate different values or to identify distinctive patterns of evaluative style. Each row in table 3.1 represents a different calculation type.

¹⁴ I will use the term "measure" in this chapter for all the activities that are carried out to measure a variable. These activities include the selection of items, and the manner in which these items are used to calculate different values or to identify distinct categories for the variable.

Table 3.1 *Categorisation of RAPM-studies by conceptualisation and measurement of evaluative style. Adapted and extended¹⁵ from Otley & Fakiolas (2000) and Vagneur & Peiperl (2000).*

	A. Use of Hopwood's items	B. Use of different items
1. Nomination to ranking (categories): categorisation on top three criteria	<u>Hopwood (1972a):</u> 1973) 8 items Ross (1994; 1995) 8 items from Hopwood (1972a) Hirst (1987): 10 items from Brownell & Hirst (1986) Pope & Otley (1996): 10 items, amended version of Brownell & Hirst (1986)	
2. Nomination to relative ranking (categories): categorisation on relative position within top three criteria	<u>Otley (1978):</u> 'How well I co-operate with colleagues' omitted 'How efficiently I run my unit' replaced 'My concern with costs' 'How much profit I make' added Govindarajan (1988): 8 items from Hopwood (1972a)	
3. Nomination to ranking (binary): BP and BC styles merged into "high budget emphasis" and NA and PC styles into "low budget emphasis"	<u>Brownell (1982):</u> 8 items from Hopwood (1972a) Categorisation on top three criteria <u>Brownell & Hirst (1986), Emsley (2001):</u> 10 items 'My attitude toward my work and company' split into two items 'Cooperation with individuals outside the firm' added. 'My long-run concern with costs and revenues' replaced 'My concern with costs' 'My ability to meet budgeted targets in the short-run' replaced 'Meeting the budget' Categorisation on top four criteria	

¹⁵ Studies included in this table that have not been categorized by Otley & Fakiolas (2000) or Vagneur & Peiperl (2000) appear in bold.

Table 3.1 (continued)

A. Use of Hopwood's items	
4. Summed Likert importance scores (continuous)	<u>Brownell (1985, 1987):</u> 8 items, amended version of Hopwood (1972a) "My concern with profit" replaced "My concern with costs" Summed Likert importance scores on two accounting items <u>Brownell & Dunk (1991), Lau & Tan (1998):</u> 8 items from Hopwood (1972a) Summed Likert importance scores on two accounting items <u>Dunk (1992):</u> Modified items of Brownell (1985) Summed Likert importance scores on two accounting items <u>Lau, Low & Eggleston (1995, 1997):</u> 10 items from Brownell & Hirst (1986) Summed Likert importance scores on two accounting items
	<u>Dunk (1989, 1990):</u> 8 items from Hopwood (1972a) Summed Likert importance scores on two accounting items Then dichotomized this score at its mean to indicate "accounting-based" versus "non-accounting" based styles.
	<u>Harrison (1992, 1993):</u> 10 items from Brownell & Hirst (1986) Summed Likert importance scores on two accounting items divided by summed importance scores on other eight items <u>Imojsili (1989):</u> 8 items from Hopwood (1972a) "Average of the raw scores to determine the styles" <u>O'Connor (1995):</u> 10 items, 8 modified from Hopwood (1972a) (or other study)? "My ability to meet budgeted non-financial quantitative targets: (i) in the short run (less than 1 year); (ii) in the long run (more than 2-3 years)" added. summed score of four budget items (quantitative measures) divided by the summed score of four qualitative items.
	<u>Govindarajan (1984), Govindarajan & Gupta (1985), Gupta (1987):</u> 1 item Percentage of bonus and performance appraisal that is objective formula-based vs. subjective based
B. Use of different items	
5. Summed Likert importance scores (dichotomy)	<u>Hirst (1983), Hirst & Yetton (1984):</u> 5 items, developed for this study Summed Likert scores on five items Degree to which quantitative data are used to assess performance. Therefore, related to use of budgets but not directly comparable. <u>Van der Stede (2000, 2001):</u> • 7 items developed for this study
6. Algebraic use of Likert importance scores (continuous)	

3.2.2 Cell A1: Hopwood (1972a; 1973)

Underlying concept. Given the many, mostly anecdotal, examples in the accounting literature of dysfunctional behaviours in relation to accounting systems, Hopwood's (1972a; 1973) central research question was:

Is dysfunctional behaviour a necessary consequence of using the accounting data in performance evaluation or at least of the imperfections in accounting systems or, rather, is it dependent upon the precise manner in which the accounting data are used?

In the process of evaluating the performance of managers, accounting data is interpreted and given meaning and can thus be used in a variety of ways. Hopwood developed a perspective to distinguish the plurality of ways in which accounting data can be used to evaluate performance. His distinction between a Budget-Constrained style, a Profit-Conscious style and a Non-accounting style was therefore a distinction between different uses of accounting data in performance evaluation. Hopwood (1973, p.18) used two broad dimensions to define these three styles of evaluation: first, the extent to which accounting performance measures are used at all for performance evaluation purposes, and, second, the emphasis on the underlying organisational purposes served by the information as against an emphasis on the information itself.

Validity and reliability. Hopwood (1972a; 1973) describes a number of procedures that support the validity of his measure. First, the distinction between the three styles was based on observations made during a series of exploratory interviews on the research site (Hopwood, 1972a, p.165). This enabled him to base the operational definition of the three styles on two phrases that had distinct meanings and were known by all cost centre heads that Hopwood interviewed. The operational definition of the three styles therefore attempted to 'gauge the correspondence between the language and meanings prevailing in the company and the operational procedures' (Hopwood, 1973, p.55). Hopwood then pre-tested the measure on groups of managers who were not included in the subsequent study. Second, in subsequent analyses Hopwood provides evidence that the meaning and description of evaluative styles from subsequent interview data corresponds with the measurement of evaluative style from the questionnaire. Third, Hopwood (1973; 1974) gives particular consideration to the relationship between wider managerial behaviours and the use which is made of accounting data in performance evaluation. He relates evaluative style to (a) leadership style (Initiating Structure and Consideration) and (b) the way in which the superiors themselves are evaluated¹⁶. The results of these analyses are in line with the

¹⁶ Hopwood refers to this as the "contagion effect", a term he adopted from Bonini, C.P. (1963), *Simulation of Information and Decision Systems in the Firm*, Englewood Cliffs, NJ: Prentice Hall.

concepts underlying the different styles¹⁷. Finally, although he used the respondents' ranking of the top three criteria to distinguish the different styles of evaluation, Hopwood also asked respondents to rate the absolute importance of all eight criteria of performance evaluation. Hopwood (1972a) reports that an analysis of the mean absolute importance of the evaluative criteria indicated different patterns of evaluative behaviours associated with different styles. Differences in the absolute importance scores of the eight evaluative criteria were all consistent with the definition of the styles. Overall, despite the fact that it was impossible to effectively discriminate between the different criteria on the basis of the absolute scores, these analyses provided (weak) evidence that the two methods (ranking and absolute scores) to measure evaluative style converged.

Problems. Kahn (1972, p.183) points out a number of problems with the categorical concepts that Hopwood (1972a; 1973) identified. First, life is not so purely typical. Second, the distinction between a budget orientation (in a Budget-Constrained style) and a profit orientation (in a Profit-Conscious style) coincides with a short-term approach and a long-term approach. This matching raises the empirical question whether a budget-orientation is intrinsically short-run and a profit-orientation intrinsically longer-run. Kahn (1972, p. 186) suggests to make the dimensions underlying the categorisation of evaluative style more explicit, analysing short-term versus long-term, budget-orientation versus profit-orientation, and positive versus negative reinforcement separately as well as in combination. Third, Kahn points out the unsatisfactory treatment of the non-accounting category as a residual type, and notes that this style is not clearly defined. In his response to these comments, Hopwood (1972b, p.190) agrees that the treatment of the non-accounting category as a residual category because of his focus on the difference between the Budget Constrained and the Profit Conscious style,

'was most likely a mistake since the non-accounting category may include some of the most interesting approaches to the use of accounting data'.

He adds that, despite the simplicity of his categorisations, a broad set of behaviours underlies the different styles:

'... while the present categorical styles are overly simple, the behaviours and attitudes which they represent include more than the accounting related aspects in isolation. The study shows that the final impact of an accounting system as a means of influencing behaviour depends on its interaction with other means of influence and control. Accounting controls are actualised through the forms of social and self-controls which are reflected within the various managerial

¹⁷ Hopwood provides additional evidence based on other relationships between the three styles and behaviours and attitudes than just the two relationships described. See, for example, Hopwood, 1972a, p.167 (especially footnote 28).

climates, and the personal motives and defences which they activate. Any moves towards a more dimensional perspective must take account of these wider issues.’ (Hopwood 1972b, p. 192)

3.2.3 Cell A2: Otley (1978) and Govindarajan (1988)

Underlying concept. The studies in cell A2 seem to measure the same underlying construct of evaluative style as Hopwood. For example, Otley (1978, p.127) describes style of evaluation as “the way in which a unit manager perceives the budget to be used in evaluating his performance, in relation to other relevant information”. Some studies¹⁸ (e.g. Otley, 1978; Brownell & Hirst, 1986) have modified the wording of some items that Hopwood used. Such modifications could be necessary to adjust the language of the items to the language of the organization in which the measure is used, and do not imply differences in construct measurement. Contrary, using the same items as earlier measures without adjusting them to the prevailing language and meaning in the organization(s) studied could imply differences in construct measurement. Therefore, the validity of making modifications or using unchanged items has to be evaluated in order to be able to assess whether studies measure the same or different constructs.

Govindarajan (1988) uses the same items as Hopwood (1972a), with only slight modifications in wording. He describes evaluative style as:

“a continuous variable whose values depend on the amount of emphasis that meeting budgetary goals receives during the ongoing process of evaluating the performance of an SBU’s general manager. At one end of the spectrum are situations in which budgetary information plays a relatively unimportant part in a superior’s evaluation of an SBU general manager’s performance. At the other end are situations in which an SBU general manager’s performance is primarily evaluated on the basis of ability to continually meet the budget and in which the manager receives an unfavorable performance evaluation if unfavorable budget variances occur, regardless of any mitigating circumstances.”

Although Govindarajan (1988) in essence uses the same underlying construct of evaluative style as Hopwood (1972a) and Otley (1978), a closer look reveals a slight change of emphasis. Instead of comparing categories of evaluative style with different underlying patterns of behaviour, evaluative style is now defined in terms of only one dimension, which is the emphasis on meeting budgetary targets. Govindarajan therefore measures the degree to which “styles” approach a Budget-Constrained style, i.e. the level of “Budget-Constrainedness”.

¹⁸ This does not only apply to studies in cell A2, but to studies from cells A3, A4, A5, and A6 as well.

Validity and reliability. Although the underlying construct in these studies is similar to that measured by Hopwood (1972a; 1973), there are important differences in how the styles are categorised. Hopwood merged the Budget-Profit style and the Budget-Constrained style into one style when he presented his findings, because of the small numbers in the Budget-Profit style and because there were no significant differences in findings between these two styles. The profit-conscious style differed from these other two styles mostly on the time-horizon taken into account: the PC style took a long-term perspective, as against the short-term concern with targets in the BC style. In contrast, Otley (1978) found that in the organization he studied a Budget-Profit style and a Profit-Budget style were styles that occurred more frequently than in Hopwood's study and were meaningful categories. The styles described superiors who focused on both short and long term objectives, e.g. trying to balance short term interests with long term organizational objectives. Based on the pattern in his categorical measurement of evaluative styles, Otley (1978, p.128) concludes that there appears to be a continuum of style. Otley validates and supports this conclusion through an analysis of the absolute importance ratings given to the items that were used to categorise evaluative style. Furthermore, Otley's modifications of Hopwood's items are based on a pilot study investigating the criteria that unit managers considered potentially important. Finally, Otley based his findings on both interviews and questionnaire data.

Although Otley adapts the operationalisation of the underlying construct, the procedures he followed provide convincing evidence of the validity of his measure, and of the comparability of his results to Hopwood's earlier findings (cf. Vagneur & Peiperl, 2000).

Govindarajan (1988) discussed the preliminary drafts of his questionnaire with scholars, and pretested the questionnaire with ten SBU managers for clarity and relevance. For validation purposes, Govindarajan (1988) mailed a questionnaire not only to the SBU managers, but also to their superiors. Govindarajan presents a significant correlation ($r = .41$, $p < .001$) between the responses of the SBU managers and their superiors as evidence for the convergent validity of the measure. As further evidence for the construct validity of his measure of evaluative style, the superior of each SBU manager was asked to indicate his agreement with an additional statement, namely "Your subordinate's salary increases are closely tied to budget performance". According to expectation the responses on this statement correlated negatively with the variable measuring evaluative style ($r = -.27$, $p < .01$).

Problems. Theoretically, Otley's continuum of styles ranges from a Budget-Constrained style to a Non-accounting style, but the Non-accounting style is virtually absent in his study (only one respondent reported a Non-accounting style). The continuum thus implies that on the one extreme we find an evaluative style focusing on short term targets and budgets, on the other extreme an evaluative style which evaluates short-term performance against long term organisational objectives. In the middle of this dimension we find a style that considers both performance on short term targets, but also evaluates performance against long term objectives. This mixed evaluative style has not received much attention in the literature

after Otley's study, although Otley's findings suggest that the behavioural associations of the mixed style could be important for the further development of theory in the relationship of evaluative style with managerial behaviour.

Govindarajan (1988) follows the exact same procedure as Otley (1978). However, Govindarajan does not report the consistency of ranking scores with absolute importance ratings of the two items that were used to categorise evaluative style. This is an important omission since Govindarajan (1988) claims to assess the degree of emphasis on budgetary targets, which assumes a difference in the absolute importance scores of the two items both within and between styles¹⁹. Furthermore, Govindarajan has not assessed the exact meaning of the items that he used to measure evaluative style. These meanings are important, since Govindarajan (1988) uses the measure in a survey, with respondents from 24 different organisations. Without an indication that the meaning of the different items is the same for all respondents, the reliability of this measure is doubtful²⁰. Additionally, at the very least, it should be established that the items used contain the most important elements of performance evaluation in all organisations studied (Vagneur & Peiperl, 2000).

Otley's suggestion of a continuum of styles has contributed to the development of continuous measures of evaluative style in later research rather than the categorical measurement applied by Hopwood and Otley. However, such continuous measures of evaluative style are not necessarily measuring the same continuum of evaluative styles suggested by Otley (1978), as I will show below.

3.2.4 Cell A3: Brownell (1982)

Underlying concept. Although Brownell (1982) in his hypotheses refers to a budget-constrained style, he defines evaluative style as "the primary emphasis in the evaluation of subordinates on budget achievement", which is a different concept (Otley & Fakiolas, 2000). Brownell uses the same items and procedure reported by Hopwood to distinguish between four different evaluative styles, but then collapses these styles into two groups: a high budget-emphasis class (containing BC-style and BP-style) and a low budget-emphasis-class (PC-style and NA-style). Otley & Fakiolas (2000, p.502) point out that the concept of budget emphasis that Brownell develops implies that the categorisation of styles is reduced to

¹⁹ Otley (1978) provides a good report of the differences between the absolute importance scores on the items "Meeting the budget" and "How efficiently I run my unit" for each style (within style) and in comparison to other styles (between styles). See Table 2 on p.129 of his study.

²⁰ Especially the requirement that a measure yields consistent results when used repeatedly is problematic, if the same wordings are used without assessing the meaning of these wordings. This is important since Hopwood clearly indicated that he developed a relatively simple measure to grasp distinct complex patterns of evaluative style.

"whether the item "meeting the budget" appears in the top three items ranked (high budget emphasis) or not (low budget emphasis). The responses to other questions, including "concern with costs" are irrelevant."

Validity and reliability. Brownell (1982) and Brownell & Hirst (1986) do not report any validity test of the measure used, nor do they justify the categorisation of styles into two groups.²¹

Problems. Although the respondents in both studies in this cell were sampled from a single manufacturing company²², in contrast to Govindarajan's sample, the validity problems with Govindarajan's (1988) measure outlined above apply to the measures in this cell as well, because, again, no formal assessment was made of the influence of the specific organisational context on the interpretation of the items.

3.2.5 Cell A4: Brownell (1985)

Underlying construct. Brownell (1985) used the eight-item measure from Hopwood (1973), but 'with some situational changes in the wording and some changes to the style of its administration'. I have argued above that changes in wording do not necessarily imply a difference in the concept that is measured. As to the 'style of administration', to overcome the problem Brownell encountered in an earlier study (Brownell 1982), namely that many of his respondents failed to complete the ranking of items, Brownell (1985) used a five-point Likert-type rating scale for each of the eight items. As a measure of evaluative style, the ratings on the two items addressing the use of accounting information - "My concern with costs" and "Meeting the budget"- were summed. This change in 'style of administration' does imply an important difference in concept (Otley & Fakiolas, 2000). Brownell (1985) now measures evaluative style as the absolute importance that a superior attaches to accounting performance measures when evaluating his subordinates (Otley & Fakiolas, 2000), rather than the relative importance of accounting criteria in relation to non-accounting criteria, as implied by the ranking procedures followed by Hopwood (1972a), Otley (1978) and Brownell (1982). Furthermore, the contrast between the two items "Meeting the budget" and "Concern with costs" that underlies the original distinction between a BC-style and a PC-style disappears in this concept. Instead, the two items are assumed to measure the same concept, i.e. the degree of emphasis on accounting performance measures. Brownell (1985; 1987) refers to his underlying concept as "Reliance on accounting information".

²¹ Both studies do report a comparison of distribution of respondents over the different styles with previous studies. This comparison, using χ^2 -tests, gives interesting information but is not a test of the validity of the measure.

²² Brownell (1982) sent a questionnaire to 48 managers of a large manufacturing company in the San Francisco Bay Area, while Brownell & Hirst's (1986) sample consisted of 80 line managers from a large manufacturing company headquartered in Sydney, Australia.

Validity and reliability. At the very least, the legitimacy of summing the absolute importance scores on the two accounting items depends on a significant positive correlation between the two items (Brownell, 1985; 1987). Table 3.2 presents the correlations reported in studies using this measure of reliance on accounting performance measures. With two exceptions (Lau & Buckland, 2001; Vagneur & Peiperl, 2000), the reported correlations between the two items are relatively high, ranging from .52 to .75. In one study (Vagneur & Peiperl, 2000), the two items have no significant correlation. No further tests of validity have been conducted.

Table 3.2. *Reported correlation between two accounting items in studies in which the scores on these items is summed*

Study	Correlation	Significance	Range
Brownell (1985)	$r = .73$	$p < .01$	2-10
Brownell (1987)	$r = .75$	$p < .01$	not reported
Brownell & Dunk (1991)	not reported	-	6-14
Dunk (1989)	not reported	-	8-10
Dunk (1992)	not reported	-	8-10
Lau, Low & Eggleton (1995)	$r = .56$	$p < .01$	2-14
Lau, Low & Eggleton (1997)	$r = .52$	$p < .001$	2-14
Lau & Tan (1998)	$r = .59$	$p < .001$	5-14
Lau & Buckland (2001)	$r = .30$	$p < .01$	5-14
Otley & Pollanen (2000)	$r = .59$	not reported	2-10
Vagneur & Peiperl (2000)	$r = .24$	n.s.	5-15 ²³

Problems. There are at least three problems with the validity of this measure of 'reliance on accounting information.' First, although most studies in this cell report the correlation between the two accounting-based items (see Table 3.2), the correlation with other items is not considered. Only Lau et al.(1995) report that the correlation between the two accounting items is higher than any of the correlations between one of the two accounting items with the other six items. This provides additional evidence to the legitimacy of the procedure to sum the ratings on the two items, but it is not a test of the validity of the measure. To assess the validity of the measure, factor analysis would have been more appropriate, because factor analysis helps to assess whether the two items belong to a single dimension that is different from the dimensions to which all other items belong. Second, many researchers (Lau et al. 1995; 1997; Lau & Tan, 1998) justify the use of importance ratings rather than rankings because it avoids the problem that respondents have with the ranking procedure²⁴,

²³ The actual range of 5-15 indicated by Vagneur & Peiperl (2000, p. 521) is incorrect. The theoretical range for the variable measuring evaluative style by summing the two accounting items, when a seven-point Likert scale is used as Vagneur & Peiperl claim to have used, is 2-14 rather than 1-15 as they report. Similarly, the actual range probably must be 5-14 instead of 5-15.

²⁴ This problem was encountered by Brownell (1982), and has resulted in the modifications made by Brownell (1985). See the section "Underlying construct" earlier in this chapter, where Brownell's (1985) measure is

and permits evaluative style to be measured as a continuum²⁵ (cf. Harrison, 1993). However, it should be noted that both Otley (1978) and Govindarajan (1988) used a continuum of (categorised!) styles with ranking scores. Furthermore, none of the studies in this cell asked respondents to rank the top three of the most important criteria. As the problems with the ranking procedure have only been reported by Brownell (1982), it would have been more appropriate to use both rating and ranking, to enable assessment of the consistency between the two procedures. Third, it remains unclear why eight or ten items are included in the questionnaire, when the response on only two items is used in subsequent analyses. Apparently, the content of all other items are not considered relevant for the construct of evaluative style that is measured. None of the studies addresses the question why the other items would be needed. This casts serious doubt on the attention that is given to the generation of items that capture the domain of interest²⁶ in the existing RAPM-literature.

In line with the studies presented in cell A4, Dunk (1989; 1990, cell A5) summed the ratings on the two accounting items, dichotomising the final score around its mean into an accounting and a non-accounting style. Because of the similarity in measurement as the cells in A4, I will not discuss the validity and reliability of Dunk's (1989; 1990) measure in cell A5 separately. The validity and reliability issues of the studies in cell A4 apply to this measure too.

3.2.6 Cell A6: Harrison (1992; 1993)

Underlying concept. Harrison's (1992; 1993) work exemplifies yet another approach. Harrison (1992; 1993) modified Brownell's (1985) measure to maintain its continuous nature, while also capturing the *relative* rather than *absolute* emphasis on accounting criteria. He divided the summed scores on the two accounting items by the summed scores for the other eight items. Despite Harrison's (1993, p.330) own statement that his measure is 'broadly similar to Hopwood', there are significant differences in the underlying concept with Hopwood (1973) and Otley (1978). Like Brownell (1985), Harrison adds the scores of the two accounting items, before dividing the summed score by the summed score of the other non-accounting items, thus assuming that the two accounting items measure the same concept, i.e. the degree of emphasis on accounting performance measures. This is in sharp contrast

discussed.

²⁵ Both Otley (1978, p.146) and Govindarajan (1984, p. 130) argue that performance evaluation style is better conceptualised as a continuum rather than to classify it into distinct categories.

²⁶ After specifying what is included and excluded in the construct definition, items have to be generated that assess the domain of interest as specified. The adequacy with which empirical measurement captures the specified domain of interest, is referred to as 'content validity' (cf. Hinkin, 1995; cf. Venkatraman & Grant, 1986).

with Hopwood's and Otley's categorisation which was grounded in the observation that these two items had very distinct meanings.

Validity and reliability. Although Harrison discusses the different measures of evaluative style at length to justify his measurement of evaluative style and to relate it to other measures, he provides no further tests of the validity of his measure.

Problems. There are several validity problems with Harrison's (1992; 1993) measure. First, Harrison used his measure in a survey study, with respondents from 28 Australian and Singaporean department and retail stores. This raises the same questions as with Govindarajan's (1988) measure with regard to the meaning of the items to different respondents, and the inclusion of all important aspects of performance evaluation in different settings. Second, summing the scores of items without validating that they do represent the same concept is questionable. Surprisingly, although the use of Brownell's (1985) measure has been criticised on this point, such criticism has not been vented on Harrison's (1992; 1993) measure. In fact, Otley & Fakiolas (2000, p. 504) state that the most interesting use of Brownell's (1985) procedure is that used by Harrison, which "may have merit in its own right". Similarly, Otley & Pollanen (2000, p. 492) argue that Harrison's approach "represents a potential improvement on the Brownell absolute measure." However, Harrison's (1992, 1993) measure has the same underlying assumptions regarding validity and reliability as any other measure that is calculated as the summed score of importance ratings on multiple items. This implies that a number of tests should be conducted to assess the validity and reliability of summing the scores of the two accounting items, and of summing the scores of the eight non-accounting items. First, the two accounting items need to be highly correlated, as with Brownell's (1985) measure. Second, in a factor analysis all eight non-accounting factors should load on a single factor, showing the uni-dimensionality of the construct of "emphasis on non-accounting measures". Third, the Cronbach's α of the summed score of the eight non-accounting items preferably needs to be higher than .70, but at least .60 (Nunnally, 1978). Finally, all item-total correlations need to be higher than .30. Even if the validity of summing the items has been assessed, it should be noted that after summing the items in this way, Harrison's (1992, 1993) measure then combines the two concepts (i.e. emphasis on accounting versus emphasis on non-accounting performance measures) into a single concept, i.e. relative emphasis on accounting performance measures. Although this appears to be sensible, the validity of combining the two concepts in this way is hard to assess²⁷. Overall, we can conclude that Harrison's calculation is valid under much

²⁷ The difficulty is that dividing the two scores first assumes that the two scores represent different concepts, and then assumes that they in fact measure a similar concept. If the two scores do indeed represent the same concept, i.e. relative emphasis on accounting performance measures, rather than two different concepts, it would have been easier to calculate the score as the summed score of the two accounting items plus the reversed scores of all non-accounting items. Conceptually, such a measure is much easier to understand (if

stricter conditions than Brownell's measure, none of which has been assessed by Harrison. Similar validity problems also apply to the measure that O'Connor (1995) and Imoisili (1989) used.

3.2.7 Cell B4: Hirst (1983) and Van der Stede (2000; 2001)

Underlying concept. Studies in this cell have developed different instruments to capture a concept of evaluative style that, according to the authors, is based on Hopwood's categorisation of evaluative styles. The reasons for developing a different instrument vary between the studies.

Hirst (1983), for example, uses five items to measure what he calls 'reliance on accounting performance measures'. The five items aim

'to detect the extent to which the receipt of rewards was contingent on a participant performing satisfactorily in terms of quantitative performance criteria and whether the circumstances surrounding a participant's evaluative situation were conducive to a high reliance on APM.' (Hirst, 1983, p. 598)

In a footnote Hirst adds that the categorical measure of budget usage developed by Hopwood (1972a) and used by Otley (1978) and Brownell (1982) is specifically tailored to manufacturing settings and as such it is not suitable for his study in non-manufacturing settings.

Van der Stede (2000, 2001) has developed a new instrument that comprises seven items to capture 'the emphasis on meeting the budget'. According to Van der Stede (2001, p. 129), this variable 'captures the extent to which top management considers meeting the budget essential on a short-term basis', thus measuring a 'rigid budgetary control style'. Although this concept of evaluative style is similar to Hopwood's (1973) concept, Van der Stede (2001, p. 129) argues that a new instrument was needed in his study, because Hopwood's

'measure contains items that are relevant at the operational level and maybe at higher managerial levels within business units—e.g. functional managers (Brownell, 1985)—but less at the junction between corporate management and its business unit general managers, as in this study'.

Validity and reliability. Based on theory, Van der Stede (2001) developed 40 items to capture five components of a broader construct, i.e. tight budgetary control. He conducted a confirmatory factor analysis to determine how well each of these 40 items represented the theoretically expected component. Next, items with factor loadings less than 0.40 on their expected component were omitted, to increase validity of the components. Based on these

valid and reliable) than Harrison's (1992, 1993) measure.

two steps, seven items to measure budget emphasis were retained. Cronbach's α for the seven items-scale was 0.83. Overall, these results provide evidence for the validity and reliability of Van der Stede's measure of budget emphasis. In contrast, Hirst (1983) does not report any validity test of his newly developed instrument. He does provide some evidence for the reliability of his instrument: Cronbach's α for the five items-scale was .76, with inter-item correlations that ranged from .23 to .62.

Problems. Hirst (1983) argues that he is measuring the same concept as Hopwood (1973) and Otley (1978). His instrument however measures the emphasis on quantitative information, which may or may not include non-financial quantitative information. A bigger validity problem is that the measure includes two items which are not necessarily relating to how a superior evaluates subordinate managers. These two items ask whether pay prospects and promotion prospects are closely related to how actual performance (expressed in quantitative terms) compare to expected performance (expressed in quantitative terms). These questions could relate to company rules and procedures, rather than to differences in evaluative style. Finally, although Cronbach's α is satisfactorily, an inter-item correlation of .23 for one of the items is rather low. This cast some doubts on the uni-dimensionality of this measure²⁸.

Van der Stede's (2001) measure of "emphasis on meeting the budget" seems to be free of validity and reliability problems. The validity and reliability of this measure needs be confirmed in other independent samples though.

3.2.8 Cell B6: Govindarajan (1984)

Underlying concept. Govindarajan (1984, p. 127) conceptualizes evaluative style as

"the degree of reliance superiors place on formula vs. subjective (non-formula) approaches towards the evaluation of the subordinate's performance and in deciding the subordinate's rewards (such as incentive bonus)."

He argues this conceptualization is similar to Hopwood's typology. A BC-style would be similar to an evaluative style which is strictly formula-based, a NA-style similar to an evaluative style in which the superior disregards financial data and relies totally on his subjective judgment, and a PC-style would be an evaluative style that is partly formula-based and partly subjective.

²⁸ Hirst (1983) did not report which item had the lowest inter-item correlation. If this item is one of the two items relating to pay or promotion prospects, the uni-dimensionality would be even more questionable.

Validity and reliability. Govindarajan does not report any validity test for his measure, except that he deductively argues that his measure of formula-based vs. subjective performance evaluation is similar to Hopwood's typology.

Problems. Govindarajan (1984) has used only one item, which blends performance evaluation and bonus determination. The manner in which bonuses are determined, may not depend (solely) on someone's superior, but could be determined by company policy. In a footnote, Govindarajan (1984, p. 130) himself notes that "the measure may be culture-bound as it depends upon distribution of annual bonuses to the managers under study." Overall, the validity and reliability of this measure is problematic.

3.3 Evaluation

3.3.1 Conceptualisation of evaluative style

The review of existing measures of evaluative style above shows that there are many different conceptualisations of evaluative style. It is apparent, however, that all conceptualisations of evaluative style have been based on the conventional view of management control, with a strong focus on budgetary control and responsibility accounting. In their pioneering studies, Hopwood (1972a; 1973) and Otley (1978) conceptualised evaluative style as a more complex pattern of behaviours and attitudes. Their measures were developed to grasp the plurality of ways in which accounting information was used to evaluate the performance of subordinate managers. This plurality was based on two dimensions: (1) the extent to which accounting information was used at all, and (2), if used, the manner in which this accounting information was used. The emphasis in later RAPM-research has shifted from this broad conceptualisation of evaluative style, to a more narrow concept of evaluative style focusing on the extent to which a superior uses budgetary data (or other quantitative data) to evaluate the performance of subordinate managers only. Brownell (1985, p. 511), for example, concludes:

"As to the role of accounting information in performance evaluation, future studies might consider refining its measurement to capture differences not only in the extent of use but also in the manner in which accounting information is used in different functional areas of the organization. Failure to acknowledge the latter might explain why I could not confirm the expectation that reliance on accounting information would show a stronger positive association with managerial performance in marketing than in R & D."

While Brownell acknowledges that he excluded the 'manner dimension' by measuring the reliance on accounting information, other researchers have done the same without clearly recognising it. The neglect of the distinction and interaction between the extent and the manner in which accounting information is used for evaluating subordinate's performance,

has added to the confusion around the concept and measurement of evaluative style (cf. Briers & Hirst, 1991).

3.3.2 Validity

Besides the many different conceptualisations of evaluative style, the review indicates that researchers have hardly paid any attention to the validity of their measures. Notable exceptions are Hopwood (1973), Otley (1978), and Van der Stede (2001). Most of the existing measures have validity problems that inhibit the use of these measures for future research. Furthermore, the validity of the measures of evaluative style that Hopwood (1972a; 1973), Otley (1978), and Van der Stede (2001) used appears to be strongly context-dependent. The items in Hopwood's (1973) and Otley's (1978) measure of evaluative style were developed within the context of a specific organisation. The items matched the organisational language, and had distinctive meanings in the organisations studied. Similarly, Van der Stede (2001) developed his measure specifically to study the emphasis on meeting the budget at the level of general business unit managers reporting to corporate managers.

3.3.3 Reliability

The reliability of many existing measures is problematic as well, especially for the measures in column A. These measures are either based on ranking procedures (cell A1-A3), have only used two items, without always assessing the correlation between them (cell A4-A5), or have used methods requiring more rigorous validity and reliability testing than were reported in the study (cell A6). Similarly, Govindarajan's (1984) measure (cell B6) is based on only one item, which means that no reliability tests can or have been conducted. The only measures of which the reliability has been reported are Hirst's (1983) and Van der Stede's (2001) measures (cell B4). However, the reliability of these two measures has not been replicated thus far in other, independent samples.

3.3.4 Conclusion of review

Our review thus shows recurring weaknesses in the operationalisation of the measures of evaluative style. Although researchers could overcome problems of validity and reliability of existing measures, the narrow conceptualisation of evaluative styles in terms of reliance on financial, budget-oriented accounting information casts serious doubts on the relevance of existing measures for future research. To increase the relevance of future conceptualisations of evaluative style, grasping the complex patterns of behaviour and attitudes of superiors when evaluating subordinate managers' performance in the context of the evaluative system that is in place in the organisation(s) studied, the conceptualisation of evaluative style has to be broadened in two ways.

First, future concepts of evaluative style should allow a broader perspective on the type of information that is used to evaluate the performance of subordinate managers, instead of

limiting the concept a priori to financial budgets or even quantitative targets, or even to diagnostic, results-oriented types of control.

Second, future concepts of evaluative style should include the manner in which superiors use this information. This is necessary to distinguish the behavioural impact of the **design** of management control tools such as budgeting from the impact of the **use** of such tools.

These two requirements imply that none of the existing measures of evaluative style is ideal for future research, and that future studies may have to conceptualise evaluative style in the context of a specific organisation. I will discuss these implications in the next section.

3.4 Implications for future construct measurement

3.4.1 Allow a broader perspective on type of information

In the conventional view, prevailing in most RAPM research, management control is primarily seen as a monitoring device, with a strong emphasis on responsibility accounting and based on a cybernetic, closed model of control. In modern accounting literature, however, management control is seen as a device to enhance continuous improvement, learning, innovation, creativity and empowerment. Instead of limiting management control to the implementation of strategies only, some authors argue that management control should be concerned with strategy formulation and revision too (see, for example, Simons, 1987; 1990; 1991; 1994; 1995; Langfield-Smith, 1997). Chenhall (2003, p. 129) summarises the development of the perspective on management control nicely as follows:

"The definition of MCS has evolved over the years from one focusing on the provision of more formal financially quantifiable information to assist managerial decision making to one that embraces a much broader scope of information. This includes external information related to markets, customers, competitors, non-financial information related to production processes, predictive information and a broad array of decision support mechanisms, and informal personal and social controls."

As a result of this development, non-financial and qualitative performance measures are receiving more and more attention, both in practice and in the accounting literature. Performance measurement frameworks like the (Strategic) Balanced Scorecard (Kaplan & Norton, 1992; 1993, 1996a; 1996b; 2001a; 2001b) and the Performance Pyramid (Lynch & Cross, 1992) are only a few examples of the development in the accounting field towards the inclusion of more non-financial measures in performance measurement systems. It is likely to assume that organisations that have included non-financial items in their measurement system will use some of these measures for performance evaluation purposes too. And even when superiors do not *use* such measures in performance evaluation, the *mere existence* of

these measures may affect the perceived style of evaluation (cf. Ridgway, 1956)²⁹. Otley (1978) showed a similar phenomenon for financial performance measures: although managers' intended evaluative style corresponded reasonably with the style as perceived by subordinates, subordinates significantly overrated the importance of quantitative financial measures in the manager's evaluative style, as soon as these measures were available. Furthermore, there is growing evidence that the role of financial budgeting and budgetary targets for motivational and appraisal purposes is limited in contemporary organisations (see for example, Marginson, 1999). The changing role of budgets in contemporary organisations (Bunce et al., 1995; Hansen et al., 2003) may indicate that evaluative styles in which budgets are relatively unimportant could be the most important or interesting styles to investigate. In 1973, Hopwood (1973, p. 20-21) already concluded that, as 44% of his respondents reported a non-accounting style, he had not given this style enough attention. However, later research aiming to reconcile the contradictory results between Hopwood and Otley with respect to the role of budgets in performance evaluation, has largely neglected the non-accounting style. Future research should not make the same mistake. To investigate evaluative styles in which budgets are relatively unimportant, the focus should be on measures that *are* emphasised, i.e. non-financial measures, or on other distinctive features of these styles, not on performance measures that are *not* emphasised, i.e. budgetary measures.

3.4.2 Include the manner in which superiors use information

Besides extending the types of information included in the analysis (the issue of the *extent* to which certain types of information (e.g. financial or non-financial, accounting or non-accounting) are used), it is important to include the *manner* in which this information is used for evaluating the performance of subordinates. Focusing on the extent dimension is valuable to better understand when the use of accounting information for performance evaluation purposes will be (in)appropriate to stimulate behaviour that helps to achieve the organisational objectives. However, the rationale for Hopwood's study and his conceptualisation of evaluative style was the anecdotal evidence in the accounting literature of dysfunctional behaviours that were associated with the use of accounting information in general, i.e. the extent to which accounting information was used. In contrast, Hopwood (1973, p. 195) hypothesised and concluded:

"Accounting information does not in and of itself pose a threat to the members of an organisation. Rather, the perception of threat and the consequent defensive behaviours which may be dysfunctional to the organisation as a whole are dependent upon the manner in which the information is seen as being used."

²⁹ Ridgway (1956) reviews the knowledge of dysfunctional consequences of implementing a system of performance measures. One of his conclusions is: "Even where performance measures are instituted purely for purposes of information, they are probably interpreted as definitions of the important aspects of that job or activity and hence have important implications for the motivation of behavior."

Thus, the occurrence of dysfunctional behaviours associated with the use of accounting information could be reinforced or weakened by the manner in which superiors use this information when evaluating the performance of subordinates.

In my opinion, future research on evaluative style would be more relevant for understanding the effectiveness of management control structures, when the impact of how managers actually use management control information would be studied in complement to the impact of control systems design (i.e. its technical availability). Instead of using a construct of evaluative style that is limited to the extent dimension only (such as the level of budget emphasis), evaluative styles should also refer to certain behavioural or attitudinal characteristics of the superior relating to the use of control mechanisms (cf. Brownell, 1985; Govindarajan, 1984). If researchers exclude this behavioural dimension from the measurement of evaluative style, the distinction between a management control tool (such as a budget) and evaluative style (how humans use a management control tool) becomes blurred. Ross (1995) offers a clear example of a confusing use of the term "evaluative style", in a study in which he tries to replicate the original Hopwood and Otley studies. His findings do not concur with Hopwood's findings that subordinates who report to be evaluated under a Budget-Constrained style experience more tension than those under a Profit-Conscious style. To explain this finding, Ross (1995, p.9) suggests that perhaps 'a more flexible use of variances for performance evaluation has developed.'

Making this remark, Ross implies that the difference between a Budget-Constrained style and a Profit-Conscious style is based on the extent to which "meeting the budget" is emphasised, without considering the manner in which the budget is used. If the Budget-Constrained style in his study does indeed imply a flexible use of variances, why did he not include this kind of behaviour in the measurement of evaluative style? Furthermore, how would the Budget-Constrained style be different from a Profit-Conscious style?

3.4.3 Study differences in evaluative style in context rather than across contexts

The distinction between the extent and the manner is, in my opinion, as important today as it was when Hopwood (1972a) created it in his pioneering study, but even in the latest criticisms on the RAPM-literature the importance of this distinction is not always recognised. For example, while Hartmann (2000) recognises Hopwood's contribution to the literature by adding the distinction between the manner in which accounting information is used and the *budget system* itself, many of his recommendations seem to be aimed at improving our understanding of the *contexts* in which reliance on accounting performance measures is more or less appropriate. Underlying these recommendations, again, is the notion that budgeting tools motivate functional or dysfunctional behaviours depending on the context in which they are used, without recognising the importance of the manner in which these tools are used. However, there is also a need for further research that addresses the question how differences in the use of the same control tools strengthen or limit the motivational aspects

of these tools within a certain (organisational) context. Therefore, there is a need for research on evaluative style that addresses the appropriateness of using certain type of performance measures *in distinct ways* within the same organisational context. Such research would need to control for the impact of the **technical design** of control tools and focus on the behavioural impact of different evaluative styles used in conjunction with the technical design. Conceptualisations of evaluative style in future research have to identify the relevant underlying dimensions that capture the plurality of ways in which superiors may use a certain performance evaluation system and the information from this system. Identification of these relevant dimensions partly depends on the type of information that is available for performance evaluation as determined by the technical design of control tools. To understand the relationship between evaluative style and managerial behaviour, the design of the evaluative system is one of the most important contextual variables to consider and to control for.

3.4.4 Control for the organisational context

The importance of the organisational context for understanding evaluative style and the relationship with subordinates' managerial behaviour has been widely recognized within the RAPM literature. Many studies have applied a 'contextual approach' to the study of evaluative style that, however, has generally been understood as an equivalent for applying a 'contingency approach' (Hirst, 1981; 1983; Govindarajan, 1984; Imoisili, 1989). Several contingency variables³⁰ have been added to research frameworks to explain differences in findings on the relationship between evaluative style and managerial behaviour, arguing that some earlier results will only hold in certain "contexts"³¹ and not in others. A typical hypothesis in RAPM-research is that (dys)functional behaviours are the consequence of the interaction between RAPM and a contextual factor (Hartmann, 2000). Thus, RAPM will only lead to dysfunctional consequences in certain contexts, but not in others. In terms of contingency theory, this is referred to as an interaction fit approach (cf. Drazin & Van de Ven, 1985), or a Cartesian Contingency approach (Gerdin & Greve, 2004).

Contextual variables have not only been used to explain differences in the behavioural consequences of evaluative styles across different settings. The RAPM-literature and many other studies (e.g., Merchant, 1981; Mia & Chenhall, 1994; Chenhall & Langfield-Smith, 1998) support the hypothesis that accounting-based controls, such as budgeting, are more likely to be used in some contexts than in other contexts, i.e. the design of control systems will depend on the context. In terms of contingency theory, this is referred to as a selection fit approach (Drazin & Van de Ven, 1985) or as a Cartesian Congruence approach (Gerdin & Greve, 2004). In this view, contextual factors are treated as antecedents of accounting-based

³⁰ These variables include environmental uncertainty, task uncertainty, national culture, and size.

³¹ Differences in contexts are defined by the contingency variable. For example, Hirst (1981; 1983) distinguishes high task uncertainty situations from low task uncertainty situations.

controls, i.e. factors that may help explain the importance of accounting-based controls in different settings.

Although these two approaches of contingency theory differ significantly, they both aim at identifying general relationships between contextual variables, evaluative style, and subordinate behaviour. This is one of the main reasons for the use of cross-sectional surveys as the dominant research method in the existing RAPM-literature. Taken as a whole, these studies do provide consistent and convincing evidence that the extent to which accounting-based controls are used in a particular organisation will depend on contextual factors, and that using such controls in the particular setting will motivate either functional or dysfunctional behaviours. However, because cross-sectional survey research samples managers from different organisations, it is impossible to assess whether differences in subordinate behaviour across settings are due to differences in the manner in which budgets are used (i.e., evaluative style), or due to differences in the extent to which budgets are used (i.e., design). This could only be assessed³², of course, if all or at least the most important contextual variables that could influence the results have either been controlled or accounted for. This is highly unlikely in existing cross-sectional survey-based RAPM-studies since these consider only one or a limited number of contextual variables from a long list of potentially important contextual variables³³. To control and/or account for the influence of the technical design of controls, and to limit the amount of contextual variation, the best approach to study evaluative style and its impact on subordinates' behaviour seems to conduct research within single organisations. Rather than applying a contingency approach, such research could be described as applying a situation-specific approach. According to Fisher (1995, p.29), "the situation-specific approach is similar to the contingency approach, but [*can be applied when*] the number of possible combinations of contingent factors is so large that attempting to find broad classes of contingent variables is seen as futile". Contextual variables are important in this approach as they can help to understand the findings. But, in contrast to the contingency theory, no attempt is made to generalize the role of (individual) contextual variables, or even to identify the impact of individual contextual variables apart from all other variables, beyond the specific situation studied³⁴.

³² Apart from other problems, such as the differences in conceptualisation and measurement across studies as reviewed above, small sample sizes, and incorrect interpretation of results from statistical data analysis (see Hartmann & Moers, 1999).

³³ This list includes, for example: environmental uncertainty, interdependence among responsibility centres, national culture, corporate culture, size, strategy, and level of decentralization.

³⁴ This does not imply that findings of situation-specific research can never be generalized. Conducting a number of situation-specific case studies could result in the identification of similar patterns between cases that may lead to the development of general hypotheses or even theory. But generalizing the findings is not the main purpose of applying this approach.

3.5 The concept of evaluative style in practice: the case of Service Group

3.5.1 Introduction

To explore the validity of these conclusions and to assess the relevance and opportunity to study evaluative style in practice I conducted an exploratory case study. The study was carried out within a large business unit (> 10.000 people) of a Dutch organisation in the service industry, to which I will refer as Service Group. The organisation is a large branch-type organisation, which makes it relatively easy to interview managers at the same level in the organisation, i.e. branch managers, who report to and are evaluated by the same superior, i.e. an area manager. I got access to the organisation through the Head of Management Development.

Data were collected through different sources. I conducted two interviews with the Head of Management Development and collected a number of documents that described the rules and procedures of the performance evaluation system. The Head of Management Development gave me the names of two area managers that I could talk to. Both immediately agreed to participate in the study.

In one area (area X) I interviewed the area manager and three branch managers. In the other area (area Y) I interviewed the area manager (Y), two branch managers (Y1 and Y2) and the manager of Customer Service. Although the area manager evaluates the latter's performance too, it is important to realize that this function is a staff function, not a line management function. Therefore, it is not directly comparable to the interviews with branch managers. The branch managers were the lowest management level (of four levels) in the organisation.

An interview protocol was developed to guide the interviews and to collect data in a semi-structured format. The interview protocols that were used for the interviews with branch managers and with area managers are described in Appendix A. All interviews started with a short description of the research project and a short introduction of the researcher. After this introduction, I would ask the interviewee to introduce himself or herself. The first questions (1 through 7) in the interview protocol are mainly included for this purpose. Subsequently, the interview protocol focuses on the evaluation criteria that are used in the organisation to evaluate performance and on the flow and availability of (accounting) information within the organisation (question 8 through 20). Most of these questions are descriptive in nature, but some (i.e., questions 12,13,17,19, and 20) explicitly ask the opinion of subordinates. These latter questions are included to give the researcher an indication of the extent to which interviewees agree with evaluation criteria. Evaluative style however does not only refer to the kind of measures used to evaluate performance, i.e. evaluation criteria, but even more important is the manner in which superiors use these evaluation criteria. Therefore, the next section of the interview protocol contains a number of questions asking how the interviewee's superior uses evaluation criteria (question 21

through 25) and, again, the interviewee's opinion (question 26 through 28). Finally the interviews concluded with two questions on behavioural outcomes and its possible relationship with performance evaluation (question 29 and 30). Interview-transcriptions were sent back to interviewees to check the validity of statements.

3.5.2 Description of organisation and performance evaluation system

Before presenting the findings from the interviews with the two area managers and the branch managers, I will first give a short description of Service Group in general, and the design of the performance evaluation system. This description is based on the two interviews with the Head of Management Development, and on some internal documents provided by Service Group.

Description of organisation

At the time of my study the strategy of Service Group focuses on both cost efficiency and quality. Since Service Group is very labour intensive, with labour costs that are 70% of the total costs, it is trying to reduce costs by replacing human resources with other resources wherever possible. Yet, despite big changes in the past few years, Service Group has never fired anyone. Furthermore, Service Group has never been confronted with strikes. Both morale and commitment of the people working in this organisation has always been good. This is something I noticed in the interviews too. People seem to be proud to be part of this organisation, and many people have been with the organisation for many years. As to quality, beyond the use of the business model of the European Foundation for Quality Management (EFQM), Service Group's are active within the EFQM and the Netherlands Quality Institute in order to exchange information with other companies. The business unit in which the interviews were conducted is the first organisation within this particular industry that is ISO 9001 certified.

Description of (formal) performance evaluation system

The organisation has a detailed performance evaluation system that is in use throughout the whole organisation. Furthermore, most managers receive training with respect to the performance evaluation cycle. The performance evaluation cycle takes up one year, and it follows the following steps:

- a. making agreements (beginning of year)
- b. evaluation mark (year' end)
- c. consequences for salary (year end)
- d. making agreements for next year (year end/ beginning of next year)

The agreements should meet the following conditions:

- Measurable.
- Clear.
- Challenging.

In October, the area manager and the branch manager sit down for a final evaluation. The area manager writes a report and gives a final evaluation mark. The organisation has four possible marks a superior can choose from: poor, passable, good, and excellent. If a manager gets a mark "good", he will go to the next step in his salary scale. If his mark is "excellent", it is possible that the manager is skipping one step and goes up two steps in his salary scale. No other financial consequences (rewards, bonuses etc.) are involved.

As described above, both cost efficiency and quality are strongly emphasised throughout the organisation. To support continuous improvement, Service Group applies the following management principles:

- Focus on results.
- Base decisions and discussions on facts.
- Encourage personnel to improve and to seek involvement.
- Earn internal and external appreciation.
- Communicate.

The quality approach is an integral part of the management and assessment cycle. All managers have personal quality targets for process quality, customer satisfaction and employee motivation. External research is used to gather facts about all aspects of results in order to be able to check whether progress has been made. Both internal and external benchmarking play an important role in the performance evaluation cycle. This makes it harder for area managers and branch managers to hide behind remarks like "In our area/branch, things are different". As an example, in a particular area for years absenteeism was 10%. There seemed to be good reasons why it was high; it was thought and accepted that the high absenteeism had to do with specific characteristics of the area (the area was largely situated in a large city). Yet, when they found out that in an area with comparable characteristics absenteeism was only 2%, those reasons appeared to have been misleading. Soon after this discovery, absenteeism was reduced drastically.

3.5.3 Performance evaluation in practice

Benchmarking

As described above, internal benchmarking is an important aspect of performance evaluation at Service Group. This is observed in the interviews with the area managers and branch

managers too. Every area manager receives monthly figures on the performance of his own area, but also about the performance of all the other areas. This makes it easy for him to see how he is performing compared to others. The same applies to branch managers: they receive monthly reports that contain figures and graphs not only with regard to their own branch, but to all branches in their area. This implies several things. First, many of the agreements are standardised. Branch managers usually accept these agreements as fair, although they are tight agreements. Second, the area manager will not accept excuses for not meeting these agreements easily. If other branch managers succeed in achieving their targets, the branch manager has to prove that the reasons why he would not be able to meet particular agreements apply to his own branch only. If he can not prove that, he and his area manager know that he has to blame himself. Third, benchmarking seems to enhance feelings of competition, and it stimulates managers to do their best. Branch managers do not want to be last in their area on any item. In the words of one branch manager:

"Internal benchmarking is used; and of course it is not nice when you have the lowest score within the area on employee motivation for example."

Fourth, benchmarking seems to enhance learning from each other. Both area manager Y and the branch managers Y1 and Y2 indicated that benchmarking makes it easy to ask a colleague what he did to achieve such good results.

Evolving performance measures

Both area managers base the evaluation of performance for a large part on pre-set targets, called agreements. They are objectified, quantified measures of a number of items, of which costs and quality seem to be the most important. Furthermore, the managers are expected to use the performance evaluation cycle. It is part of management culture. They even receive training in using it.

The organisation is (still) changing, and so are the items on which performance evaluation is based. In the past, the only thing that counted was costs. In recent years, the organisation has become increasingly aware of the customer. Besides costs, quality became important. In addition, at the time the interviews were conducted, customer satisfaction and employee motivation had just recently become part of the performance evaluation cycle as well, and received increasing attention in the contracts. There is a tendency towards the inclusion of non-financial items. Costs were very important in the past and they still are at present, but the non-financial performance measures have changed the perspective of performance evaluation as is illustrated by the following quote from one of the branch managers:

"Of course, costs and quality are important, but I think that besides that a lot of attention has to be given to customers and customer satisfaction. That goes for employee motivation too.. I think the thought behind the contract in its present form is that there is more than just costs.

In the past everything revolved around costs. But I believe that only after <an important event>³⁵ we discovered that more things are important. There has to be a certain balance”.

Relation between design and style

Changes in the design of the performance evaluation cycle over the years seem to be intertwined with changes in subordinates’ perceptions of how their superiors evaluate their performance. This makes it more difficult to distinguish the role of the superiors (style) from the role of the formal architecture of the performance evaluation system. However, an extra opportunity to explore the role of superiors occurred in area X, because area manager X had only been employed in area X for a couple of months when I conducted the interviews. Thus, in the interviews with the branch managers in area X I asked branch managers whether they perceived any differences between the former area manager and the current area manager. Although subordinates did perceive some differences between the superiors, part of these differences seemed to reflect the changes in the design of the system. For example, branch manager X3 indicated that the former area manager emphasised costs more than anything else, while area manager X emphasises that more things are important than just costs. He added that this was a logical result of changes in the organisation as a whole: customer satisfaction and employee motivation have only been measured for two years. However, differences in “style” between the two area managers were perceived as well. Branch manager X1, for example, explained that, in contrast to the former area manager, area manager X is willing to listen to explanations of why targets have not been attained. The branch manager illustrated this difference in evaluative style by referring to a conflict he once had with the former area manager:

“I explained him: “I did everything I could, but it just doesn’t work, I can’t attain this goal. There are circumstances which make this goal unattainable.” But he would answer: “Yes, but still, you have to attain it; you signed for it”. That was with the former. With this one, you have to be specific about what you did of course. But if you can explain it, well...”

Differences in evaluative style between areas (between area managers)

Given the design of the performance evaluation cycle, it is not surprising that all branch managers indicate that costs and targets are an important part of their performance evaluation. However, even though costs and targets are important this does not necessarily mean the superior’s evaluative style is strictly financially-oriented or target-oriented. Some area managers may think costs are important, but other items are just as important. Area manager X seems to be an example of the latter, while the former area manager in area X seemed to have an evaluative style that was more rigidly financial-oriented. Furthermore, although area manager Y emphasises both costs and quality, and can be said to have a non-

³⁵ I chose this description because revealing the nature of this event would make it easier to identify the organization. Yet, this event could be important in understanding the changes that Service Group underwent.

financial oriented evaluative style, he seems to adhere more strictly to pre-set targets than area manager X does. Although branch managers Y1 and Y2 may be able to explain why they did not achieve their targets, this does not change the fact that area manager Y still thinks that they should have attained the targets: managers have either met agreements or they have not; explanations can not change that. Yet, when managers provide a reasonable explanation this does influence what the consequences of not meeting the agreements will be for the overall evaluation. One of the branch managers in area Y described the manner in which his performance would be evaluated as follows:

"Initially, the conversation will focus on the items in the contract, but he will ask for the story behind it. Often he does already know the story anyway. The items in the contract are rigorous, and will be checked item by item. All items will be considered. And when I do not meet certain items, then at first it is not good. And even if I have a good story, it remains wrong. But it does not mean that the evaluation will be negative. If you are able to show what you did, then it will be taken into account in the evaluation".

While area manager X also stresses that the agreements in the contract have to be met, the interviews suggest that area manager X puts a lot more emphasis on the process by which branch managers achieve the results than area manager Y, also in the daily practice between formal evaluation moments. The importance of how a branch manager operates is an ever-recurring theme during the interview with area manager X. First, when describing his own function, the area manager says: "I regard it as my most important responsibility to determine what has to be realised, but above all how it has to be realised." Second, concerning his own evaluation, the Area Manager says he slightly disagrees with how his boss evaluates him, because the evaluation is limited to the items in the contract, while he feels dedication is more important. "In evaluating my performance, my boss runs through almost every item on which an agreement has been made. I do not feel that's necessary." Third, the area manager has even introduced a special term (a verb) in his area to describe how he expects a (branch) manager to behave: "a manager has to ZIP". "ZIP" is an abbreviation of the Dutch words "Zelfstandig", "Initiatief nemend", and "Probleemoplossend", which can be translated as Autonomous, Taking initiative, and Problem solving (ATP). I will use the abbreviation of the English equivalents, ATP, instead of the Dutch abbreviation ZIP in the remainder of this chapter. ATP is explicitly included in the contract as part of the agreements. A branch manager's contract in area X therefore contains agreements on expected outcomes and agreements on how the manager should achieve these outcomes. Both types of agreement are part of branch managers' performance evaluation. The area manager stated:

"A branch manager who does not show ATP puts problems at my desk that do not belong there. A branch manager will come across that because I will point it out to him. He will also come across it at the end of the year during the performance evaluation meeting, because ATP is a part of the contract that we both signed at the beginning of the year... When I believe a

branch manager does not show ATP, his evaluation will be somewhat negative. If the branch manager disagrees, he has to illustrate that he did show ATP. For example, if he claims to have shown problem-solving behaviour, then he should name, for example, five problems he solved in the past year."

The purpose of introducing ATP, however, is not to provide branch managers with an excuse for not reaching expected outcomes or to decrease the importance of meeting the targets. Quite the contrary, as the following quote from the area manager illustrates.

"The contract has to be observed. All agreements in the contract are realistic. They are obtainable with reasonable effort. However, in the past you could always explain why you did not attain these targets. People began to focus on finding good excuses to allow them not to attain their targets. I was very annoyed by that. I thought: "I work hard, but I do not attain my targets. What are my managers doing to attain them?" It had to change. The story that a branch manager puts forward for not attaining his targets has to be realistic. There have to be good reasons and you must be able to show what you have done to solve problems, or to attain certain targets. If a manager can show that he really managed everything well, deviations from the contract are allowed. Initially it is necessary to head for the targets; there should not be any discussion about these targets."

The interviews with the branch managers in area X for the greater part supported the description of the (intended) evaluative style that area manager X provided. According to two of the branch managers, although the area manager attaches importance to outcomes, he has defined another important component of performance, which is how you attain these outcomes. As one branch manager said:

"The area manager calls it -and it becomes increasingly clear- showing ATP, he calls it moneymaking, and he calls it centralising the customer process. As to moneymaking, the area manager says: "you have a contract, but actually, you should be able to achieve even more than what's in it." Thus, you get rid off: "now I have to spend my whole budget, otherwise I will have less to spend next year." ATP stands for Autonomous, Taking initiative, and Problem solving, so "don't bother me with your problems". As to customer process: do you have the customer between the ears and can you bring it between the ears of your own subordinates".

Similarly, another branch manager said:

"Our contract includes not just targets, but also agreements on the manner of functioning: a number of behavioural characteristics that fit the management style that the area manager propagates. Then you talk about a number of concepts. What is essential to him is so-called ATP. That is autonomous, taking initiative, and problem solving... We have a culture -from the past, and a little exaggerated, and black and white- that we do what we are appointed to, and we try to do that well. However, do not ask me too much, too much own initiative, too many problems to solve, because then it becomes difficult. That applies especially to the shop floor. Therefore, the aim is that we start to do things, that we signalise things, that we embrace opportunities, and that we do not talk in terms of problems, but in terms of solutions. It is not

interesting whether something is a problem, but it is interesting whether you can settle it, and solve it."

Differences in evaluative style within areas (between branch managers)

The descriptions that branch managers X1 and X2 gave of area manager X's evaluative style seem to be consistent with each other, as well as with the description given by area manager X himself. However, the description of branch manager X3 shows some differences. According to branch manager X3, behaviour and effort are not part of performance evaluation. This branch manager states that performance evaluation focuses on figures only; performance evaluation is too straightforward. Yet, this seems to contradict the impression the other three interviews provide: effort seems to be important; explanations for not attaining targets seem to be possible, lest they are realistic. All three refer to ATP, which is clearly referring to effort and behaviour, several times during the interviews; however, branch manager X3 does not mention it once. A number of reasons may explain this apparent contradiction.

1. We should realise that there has been no formal performance evaluation with area manager X yet. With the former area manager, performance was evaluated very rigidly against the pre-set targets. Branch manager X3's perception of his area manager's evaluative style may be coloured by past experiences. A comment from branch manager X1 concerning the visit he got from area manager X supports this argument:

"As part of the control cycle, we have progresses conversations, and two weeks ago the area manager was here. He had said: "I will come to have a progresses conversation". With the other boss it was always like this: prepare, and demonstrate to decimal places. So, thus I did. He immediately said: "Hey, it goes really well. I do not need to hear that at all. I would really like to know: how do you manage employee motivation?" So I sat here stuttering. I actually knew that he would do it like that, but you are still used to the cycle, in this case progresses conversation."

2. Branch manager X3 experienced trouble in meeting this year's agreement. In particular, the branch manager explained that he was using more labour than he was supposed to. There is only a slight chance that he will meet his agreement on this item, and he knows it. He feels threatened by it, and he thinks it is not fair that he is going to get a negative evaluation at the end of the year. He is sure he will get a negative evaluation. In contrast, branch managers X1 and X2 are quite sure they are doing a good job. The current performance may thus influence the perception of evaluative style. Possibly, managers who experience trouble in achieving their targets may perceive their superior's evaluative style in a different way than managers who know they are performing well. In this particular case, this effect may even be stronger because of past experiences (see the first reason above).

3. While the two former reasons assumed the difference in description occurred because of a difference in the perception of the performance evaluation, it is possible that this difference in perception reflects an actual difference in evaluative style. In other words, area manager X may use a different evaluative style with branch manager X3 than with the other two branch managers. Whether evaluative style is a personality trait or behaviour is an unresolved issue in the literature. If evaluative style is a personality trait, I would expect that the actual evaluative style does not differ across subordinates. If performance evaluation is behaviour, it is possible that the superior acts differently towards different subordinates, somehow adjusting his style to the particular subordinate. In this particular case, it is possible that area manager X stresses the importance of meeting the targets precisely because branch manager X3 has trouble meeting them. If the area manager at this stage accepts that branch manager X3 will not meet his targets, branch manager X3 may not do the best he can to meet them. Since both branch managers X1 and X2 will probably meet their targets anyway, there is no need to stress those targets. Here the emphasis is on what they did to attain these targets; these managers have to show that they attained their targets because of the way they managed, and not just because they got lucky.

The descriptions that branch managers Y1 and Y2 gave of area manager Y's evaluative style seem to be consistent with one another, as well as with the descriptions given by the service centre manager and the area manager himself. All three managers evaluated by area manager Y seemed to agree with the way in which they were evaluated. In contrast to the managers in area X, all managers interviewed in area Y had been employed in their current function for the past three years at least. The area manager selects the branch managers working under his supervision himself. The area manager explicitly stated that if managers received a formal evaluation mark "passable", they would have to do better in the following year. If they fail to do better, they have to find themselves another job. However, in contrast to the ATP notion in area X, the area manager Y does not provide clear guidelines on how performance can be improved³⁶.

Impact on subordinate managers' behaviour

There is some evidence in the conversations with managers X1 and X3 that evaluative style affects the level of tension that managers experience and perhaps managerial performance also. Especially branch manager X1 expresses strong suggestions in this direction. He said that this year he found pleasure in his job again, which he had lost in the previous year. He put in a lot of effort for the area as a whole, but none of these activities were considered in his performance evaluation, while he felt they should have been part of his evaluation.

³⁶ As stated earlier, one way to improve performance is by asking colleagues based on benchmarking data.

The interviews also provide some evidence of dysfunctional behaviours associated with rigid styles, i.e. evaluative styles in which the outcome of one particular target (e.g. budget, a quality target such as delivery time) determines the evaluation no matter the circumstances. Branch manager X1 described two different ways in which people would manipulate the numbers in the past, just to avoid exceeding their budgets. Of course it was just something the branch manager had heard; he never did and would never do it himself. Area manager Y also stated that he knew some branch managers in the past who achieved their costs and quality targets, but had very poor employee motivation records. He further stated that other managers are very reluctant to take risks, which in his opinion could be marked as dysfunctional since 'managing is taking risks'.

An important observation from the interviews is that including the process through which results are obtained in addition to a strong emphasis on meeting the targets could be important from a behavioural perspective as well. Area manager X, for example, strongly believes that showing ATP will contribute to attain the agreed outcomes:

"If the contract is not observed, this simply means that you do not show ATP. Thus, it all comes down to management style. The question is not so much what they do as how they do it. If they do it well, in the right manner, the results will be accordingly. A manager who shows ATP will undoubtedly meet the agreements on the outcomes."

The importance of including the process is also stressed in the interviews with the branch managers, as some branch managers indicate that not all targets can be influenced by them (the controllability principle). In that case, they feel they should be able to explain what they did to solve the problems, and to show that not meeting the targets is not their fault. Furthermore, some branch managers also indicate that a strong emphasis on targets could result in a short-term orientation.

"Your own contribution, the manner of operating, behavioural characteristics, you all need it to be able to reach your tough targets. I have agreements on absenteeism and when I do not realise it in a certain way, then maybe I will make it for half a year, by applying some sort of macho-management, putting people under heavy pressure, but that does not get me anywhere. So, the manner in which you operate is also terribly important."

Individual characteristics of subordinates

Personality seems to be important in understanding how performance evaluation affects behaviour. While branch manager X1 could immediately answer my question about what the agreements for this year were, branch manager X2 had far more trouble answering that question. This may be explained by the fact that branch manager X2 did not at really care about his contract. He did not consider it a specific guideline. He knew what he was capable of and what he had to do and he would do it, with or without contract. By contrast, branch manager X1 described the contract as a day-to-day guide, to direct his priorities. The

contract contained all the things that his boss considered important. Therefore, he focused on achieving that. He also said that if certain things would not be part of the contract, he would not do them. The contract helped him in getting these things done too. A possible explanation for this difference between branch managers X1 and X2 is a difference in personality. Branch manager X1 seems to be someone who needs the approval of others to know whether he is doing a good job, while branch manager X2 is not dependent upon the judgement of others. The difference can also be explained by the fact that branch manager X1 had become used to a very rigid evaluative style, making him unsure whether his area manager would appreciate all his efforts. Branch manager X2 did not have these negative experiences in the past.

3.5.4 Discussion

The findings from the Service Group case substantiate and further clarify the theoretical arguments made in section 2 regarding the need to broaden the perspective of evaluative style research.

Broadening the type of information

The Service Group case data clearly indicates the need to broaden the concept of evaluative style to include non-financial type of performance measures and to include behavioural characteristics of superiors. In the Service Group organisation evaluative styles cannot be distinguished solely on the manner in which budgets are being used. Although costs and budgets are important aspects of performance evaluation, quality targets are emphasised as well. In addition, measures of customer satisfaction and employee motivation receive increasing attention. It should be noted that, in contrast to later measures of evaluative style in the RAPM-literature, Hopwood's (1972a; 1973) original categorization of styles in a Budget Constrained, Profit Conscious, and Non-Accounting category, seems to be able to capture these broader types of information. However, a close look at the differences in evaluative style within Service Group casts some doubt on the ability of Hopwood's categorization to grasp the subtle nuances in style between area manager X and Y. Both stress the importance of financial and non-financial targets, and will only accept deviations from these targets if subordinates can provide a realistic explanation. However, between formal evaluation moments area manager X seems to be more concerned with the process by which his branch managers can attain their targets than area manager Y. This is clearly illustrated by the emphasis of area manager X on managing and supporting autonomous, problem-solving and initiative taking behaviour on the part of his branch managers.

We should take into account that using accounting information (targets) to evaluate managerial performance is quite appropriate within Service Group (since benchmarking enables area managers to correct outcomes for uncontrollable results, and evaluate performance of managers relative to other managers) (Merchant & Van der Stede, 2003). Therefore, I believe that my findings of the growing importance of non-financial performance

measures are important and that the changing (broadening) performance evaluation basis can be generalised to other contemporary organisations.

Thus my findings support the need to broaden the concept of evaluative style beyond the manner in which superiors use budgetary or other quantitative targets to evaluate performance.

Contextual influences on evaluative style

The analysis of the interview data of Service Group shows the influence of the organisational context in general, and the control system design in particular, on interpretation and meaning of differences in evaluative style. First, the emphasis on costs and quality in performance evaluation is not a personal preference of superiors per se. Instead, these items are emphasised throughout the organisation, in line with the organisational strategy.

However, the data analysis does reveal that there are differences in how superiors use these measures, not just when actually evaluating the performance of subordinates, but during the whole period in particular. This supports the importance of distinguishing between the evaluative system and evaluative style. Secondly, internal and external benchmarking is an important part of performance evaluation within Service Group. This could explain the emphasis on meeting targets, and why subordinates accept these targets. Furthermore, if subordinates disagree with targets, discussions about the (level of) targets are based on facts and on the (level of) targets in similar branches. In understanding the role of benchmarking, it is important to realise that although branches within Service Group may differ in size, the processes in all branches are very similar, making comparisons and standardised target-setting much easier than in organisations with business units which differ from each other with respect to inputs, processes and outputs.

Additionally, a strong emphasis on targets does seem to be appropriate for Service Group, and subordinates agree with and accept the targets. Therefore, evaluative styles that are characterised by a strong emphasis on targets seem to be appropriate, inducing the right pressure to realise the targets. If the targets are not emphasised, it is relatively easy for branch managers to miss their targets and make up all kind of excuses. Yet, not all targets are fully controllable by the branch managers, and when they are unable to meet the targets while they know that is the only thing that counts, they will not be motivated anymore to do the best they can. Thus, even though a strong emphasis on meeting the targets seems appropriate, there is also some evidence that an emphasis on the processes through which branch managers attain their targets, especially when evaluating their progress during the period, keeps managers motivated to do the best they can, and will help them to attain their targets. Especially area manager X applies such a style, which can be described as a more coaching type of style than area manager Y. Furthermore, it may prevent a too narrow focus on short-term performance at the expense of long-term performance. Thus, the interview data also provide some evidence that the behavioural impact of particular evaluative styles

should be considered against the background of the 'fit' between the design of the performance evaluation system and the wider organisational context.

The character of the superior-subordinate relationship

Our findings suggest an interesting dyadic dynamic between superiors and subordinate characteristics. In area Y the subordinates seemed to agree in their perception of how their superior evaluated their performance, but in area X one of the branch managers had a very dissimilar perception of his area manager's style of evaluation than two of his colleagues. Possible explanations include circumstances influencing the subordinate's reading of his superior's evaluative style, but also actual differentiation in style by the superior due to differences in subordinate characteristics. These observations, however, need further investigation in future studies. Do subordinates' perceptions of their superior's evaluative style differ? And if they do, do these differences reflect differences in the actual style used by the superior, or are they caused by differences in the reading of evaluative style by the subordinates? And when superiors evaluate subordinates in different ways, why? Can differences in personality between subordinates or differences in performance level (partly) explain differences in perceived or intended evaluative style, as my findings may suggest? These are important questions, as they relate to the validity and reliability of research findings that have used subordinate managers' responses for data analysis, as in most of the RAPM-studies, but most of all because they provide new avenues for building the field of evaluative style research beyond RAPM increasing its relevance for management practice.

3.6 Conclusion

In short, in this chapter I showed theoretically and empirically that due to the recent developments in management control it is questionable whether existing measures of evaluative style, even if valid and reliable, capture all relevant distinctions in evaluative style in contemporary organisations. In order to improve our understanding of evaluative style and its impact on subordinate managers' behaviour, it is necessary to conceptualise and measure evaluative style in ways that are relevant, meaningful and valid in the particular organisation studied. Many studies indicate that the organisational context influences the design of accounting-based controls such as budgeting. As a result, conceptualising evaluative style in terms of financial budgets as is typical in existing RAPM-research will be more relevant in some contexts than in others. In contexts where financial budgets are relatively unimportant, other appropriate dimensions or combinations thereof have to be identified to capture relevant differences in evaluative style. Otley & Fakiolas (2000) have recently identified five general dimensions of evaluative style that could serve as a starting point, but, given the current state of the literature, it is still not possible to identify beforehand which dimensions would be relevant in a particular context.

Instead of using the same measure of evaluative style for different organisations, tailored measures of evaluative style may be needed in different organisations, to account for differences in the technical design of controls (and other contextual variables). This enables the development of concepts of evaluative style that are relevant and valid in the context studied, and 'accommodate changes in contemporary control systems' (Chenhall, 2003, p. 131). In addition, there is a need for future studies on evaluative style that study the behavioural impact of evaluative style as a complement to the impact of the 'fit' between design of components of performance evaluation and the wider organisational context. Finally, future studies need to investigate the influence of subordinate characteristics. This is necessary to understand the effect of individual characteristics on the perception of evaluative style, and to determine the need for superiors to differentiate evaluative styles between subordinates. Conducting case studies in single organisations may enable researchers to conceptualise and measure evaluative style in such a way that it captures relevant differences in evaluative style, given the particular design of the evaluative system. This would be the first step towards a deeper understanding of how and why such differences in evaluative style are related to subordinate managers' behaviour.

4 Antecedents and consequences of evaluative style: a theoretical framework

4.1 Introduction

There is little disagreement among researchers that differences in evaluative style will lead to differences in subordinate managers' behaviour. However, due to the problems outlined in the previous chapters, there is also little disagreement among researchers that the RAPM-literature has not provided a clear understanding of how and why evaluative style relates to subordinates' behaviour. The main purpose of this chapter is to identify variables that help to understand how and why evaluative style in a specific context, i.e. of a single organization, is related to subordinate managers' behaviour. This will result in propositions about these relationships that can be tested within the context of a single organisation. To measure differences in evaluative style, most studies have relied on the subordinate's perception of the manner in which his or her superior evaluates his or her performance. This indeed seems to be the relevant and most appropriate level-of-analysis when trying to explain a subordinate's attitudes and behaviours. Yet, the question why subordinates perceive differences in evaluative style has remained largely unexplored in existing RAPM-studies. This is an important issue, especially salient in single organisation field studies as the exploratory field study in the previous chapter has illustrated. An important question that was raised in the previous chapter is whether perceived differences in evaluative style relate to actual differences in behaviours and attitudes of superiors in evaluating the performance of subordinates. I concluded there that the context is likely to affect a subordinate's perception of actual evaluative behaviour of his superior and also his judgment as to its appropriateness. One of the most direct influences on a subordinate's perception of evaluative behaviours of his leader will, of course, be the actual evaluative behaviours of this leader. The actual evaluative behaviours of the leader are likely to be influenced by the context of evaluation as well, as leaders, especially effective leaders, may adapt their style of evaluation to the context. If actual behaviour of the leader changes, it is likely that the perception of that behaviour by subordinates changes too. Therefore, adding the superior level as a level of analysis could add to a better understanding of why differences in perceived evaluative style emerge. Several authors (Hartmann, 2000; Covalski et al., 2003) have argued for alternative levels of analysis than just the individual subordinate:

"Theory is needed too about the appropriateness of RAPM under uncertainty for the performance evaluator. In RAPM studies 'appropriateness' is examined in terms of subordinate's attitudes and responses. An alternative level of analysis would be the superior, and a related question would be whether superiors' evaluative behaviors can be explained in terms of contextual 'appropriateness'." (Hartmann, 2000, p. 475)

"Almost all the extant psychology-based budgeting research is at the individual level of analysis, because of its focus on how the effects of budgeting vary across individuals. Two caveats should be kept in mind, however. First, the focus is typically on a subordinate's budget-related mental state, behaviour, and performance in the context of a superior-subordinate dyad (for example, as they work together to develop a budget for the subordinate). Although the dyadic relation provides the budgeting context, this research usually does not investigate the causes or the effects of a superior's mental state, behaviour, or performance, instead focusing only on the subordinate. Second, only a few studies focus on budgeting at the subunit level with multiple subordinates (e.g., Daroca 1984)... more might be learned about budgeting through attention to the dyadic and organizational context of budgeting." (Covaleski et al., 2003, page 22 and page 29)

Adding the superior as the level of analysis raises important measurement issues. One approach to measure the evaluative style of the superior would be by asking the superior questions to identify his evaluative style, i.e. using a self-reporting measure. Another approach would be to ask different subordinates reporting to the same superior to rate the evaluative style of their superior, and then assess the average perceived evaluative style as a measure for the perceived evaluative style at the superior level, i.e. using a group-level measure. Yet, the results of the field study in the previous chapter indicated that in that particular organisation one of the two groups of subordinates reporting to the same superior had quite different perceptions of the evaluative style of their superior. This measurement issue gives rise to different type of questions than those that have appeared in most of the RAPM-studies. For example: Do differences in perceived evaluative style across subordinates reporting to different superiors occur at the level of individual subordinates or do they reflect differences at a higher level, i.e. between-group differences? Do subordinates who are evaluated by the same superior report differences in perceived evaluative style, or do they report a similar style of evaluation? And if there are within-group differences in perceived evaluative style, do these reflect differences that are intended by the superior? Exploring this type of questions enhances our understanding of the construct and nature of evaluative style. Understanding the construct and nature of evaluative style is necessary in order to understand the behavioural consequences of perceived evaluative style. Therefore, before identifying variables that may help explain why perceived evaluative style is related to individual subordinates' behaviour, I will review the literature to identify factors that may help explain why differences in perceived evaluative style at different levels of analysis occur. Although the focus will be on evaluative style as perceived by individual subordinates, it is important to be aware that this perceived style might be affected by actual behaviour of the superior (i.e. superior characteristics).

This chapter is organized as follows. In § 4.2, based on a review of RAPM-literature, I will identify contextual factors that could explain why subordinates perceive differences in evaluative style. In addition to contextual factors, leadership style is likely to influence evaluative style too. In fact, leadership style and evaluative style are very similar concepts.

The level of analysis issues and contextual factors that apply to evaluative style most likely apply to leadership style as well. Since the theoretical reasoning behind the expected relationship with evaluative style is different for leadership style than for the contextual factors, the relationship between leadership style and evaluative style will be addressed separately in § 4.3. Next (in § 4.4), I will identify variables from previous RAPM-research that may explain why differences in perceived evaluative style within a specific organization may relate to differences in subordinate managers' behaviour. The chapter will be concluded with a summary of suggestions on how future research could improve our understanding of the construct of evaluative style and of the relationship of evaluative style and subordinate managers' behaviour.

4.2 The antecedents of perceived evaluative style

4.2.1 Introduction

In this section I will review the literature to identify contextual factors that could help explain why differences in perceived evaluative style occur, i.e. antecedents of perceived evaluative style. The type of contextual factors differ, depending on the level at which differences in perceived evaluative style are studied. In this section, three different levels will be considered: the organizational level, the group level, and the level of an individual subordinate.

Antecedent factors at the organizational level help explain why subordinates from different organizations, on average, may report differences in perceived evaluative styles. These antecedent factors are likely to influence the extent to which, for example, accounting-based controls in general will be used. Thus, the factors at this level explain differences in the technical design of control tools, and, accordingly, may explain differences in the relevance of particular dimensions of evaluative style such as budgeting or the role of quantitative targets. For example, these factors may help explain why in some organizations the non-accounting style is the most often reported style, while it is virtually absent in others. It should be noted, however, that studies that are limited to single organizations cannot actually test the influence of these factors on perceived differences in evaluative style. This is impossible since the assumption is that these organisational level factors do not vary across subordinates within the same organization. In other words, by limiting the study to a single organization the influence of these factors on individuals or groups is controlled for. For this reason, I will not present the expected influence of factors at the organizational level as propositions.

Antecedent factors at the group level may help explain why, within the context of a single organization, subordinates reporting to different superiors, on average, may report differences in perceived evaluative style. So, these factors may explain differences in

perceived evaluative style across groups of subordinates, where a group is defined as subordinates reporting to the same evaluator.

Finally, antecedent factors at the individual level may help explain why within a single organisation, and within the same group of subordinates, subordinates may report differences in perceived evaluative style.

The contextual factors and their expected influence on perceived evaluative style at different levels of analysis that I will discuss in this section are outlined in figure 4.1. The differences in evaluative style at different levels are related. For example, if differences in perceived evaluative style can be totally explained by organizational level factors, I would expect all individual subordinates within a single organisation to report a similar style. For this reason, I connected the differences in evaluative style at different levels in figure 4.1. I used a dotted line because the relationship between differences in style at different levels in itself is not the focus of the review here.

4.2.2 Factors at the organisational level

Table 4.1 reports the distributions of evaluative styles in several RAPM-studies that were limited to single organisations. As table 4.1 shows, there are important differences in reported styles between the studies, indicating that contextual factors could have influenced the distribution of styles.

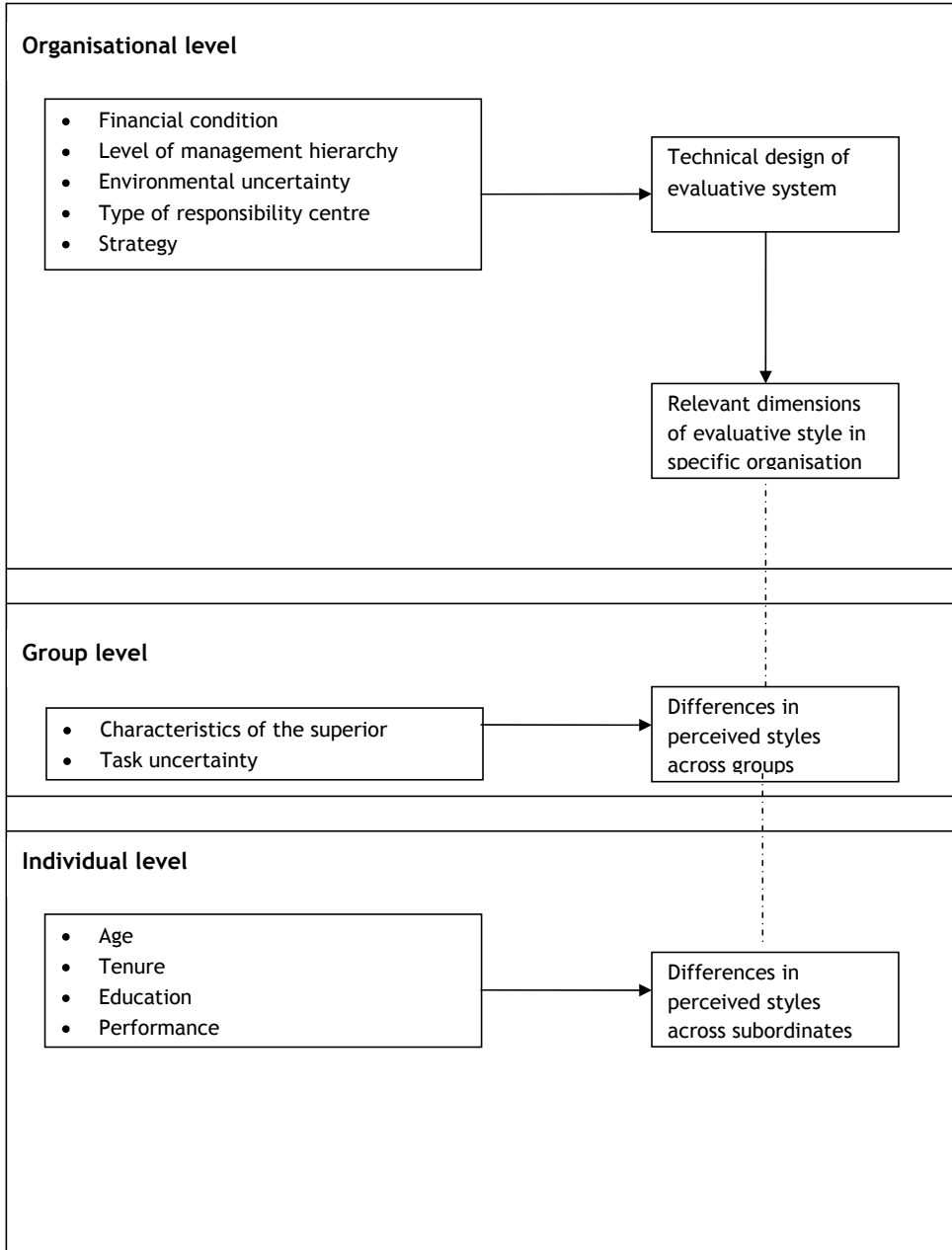
Table 4.1. *Distribution of different perceived evaluative styles in percentages of total sample across several RAPM-studies (Adapted from Brownell & Hirst, 1986, p. 245)*

Sample	Style of Evaluation			
	BC %	BP %	PC %	NA %
Hopwood (1973) (N = 167)	20	10	26	44
Otley (1978) (N = 39)	13	56	28	3
Brownell (1982) (N = 38)	21	24	24	31
Brownell & Hirst (1986) (N = 76)	17	41	7	35
Brownell & Hirst (1986) (production only) (N = 25)	12	44	8	36
Brownell & Hirst (1986) (nonproduction) (N = 51)	20	39	6	35
Imoisili (1989)	57	-	39	4

For example, in both Otley's (1978) and Imoisili's (1989) study, the Non-accounting style was virtually absent, while it was the most reported style in Hopwood's (1973) study. Imoisili (1989, p. 332) suggests that the relative absence of the non-accounting style in his study and in Otley's study is due to the poor **financial condition** of the organisations studied:

"It is likely that when companies are undergoing periods of declining profitability, there might be so much emphasis on financial data in performance evaluation that it may be difficult to truly classify managers into different styles on the basis of how much consideration is given to budget data in performance evaluation... Perhaps when organizations are passing

Figure 4.1 A multi-level model of contextual factors that may explain the emergence of evaluative style



through financially difficult times, there are no managers with budget responsibilities who are not evaluated on the basis of budget data. There is also the possibility that during periods of poor financial performance, there might be so much emphasis on budget information and reports on the interactions between a manager and his or her supervisor that most managers may have a tendency to think they are evaluated on the basis of budget data even though that might not be the case in circumstances where the company was profitable.”

Additionally, there is some evidence that the **type of responsibility centre** may influence evaluative styles chosen by superiors (Brownell, 1982, p.16; Otley, 1978, p.143). To explain the difference in distribution of perceived evaluative styles in Hopwood’s (1973) and Otley’s (1978) study, Brownell argues that apparently

“profit-centre managers (Otley’s sample) are evaluated much more on the basis of accounting information than are cost-centre managers (Hopwood’s sample and the sample in Brownell’s study).”

Differences in the type of responsibility centres studied may also imply a difference in the **level of management hierarchy** (Poe et al. 1991, p.171), where profit-centre managers are “generally considered to be higher in the management hierarchy”. In general, the higher the level of management hierarchy, the more managers may be evaluated on accounting performance measures.

Simons (1987) provides some evidence that **business strategy** may influence the “choice” of evaluative styles. His results suggest that defenders use budget-based compensation schemes to a greater extent than prospector firms (Briers & Hirst, 1990).

One of the most important factors that influence the type of measures a superior will use to evaluate performance seems to be **environmental uncertainty**. Environmental uncertainty refers to “uncertainty stemming from factors in the organization’s environment” (Hartmann, 1997, page 46). It can be defined as (Govindarajan, 1987, page 127):

“unpredictability in the actions of the customers, suppliers, competitors and regulatory groups that comprise the external environment.”

The empirical evidence regarding the use of accounting performance measures is mixed (Hartmann, 2000; Chenhall, 2003). While Merchant (1984) provide limited support for the hypothesis that the use of accounting performance measures will decrease when uncertainty increases, other studies provide evidence that the use of accounting performance measures is not affected by uncertainty (Govindarajan & Gupta, 1985) or that the use of accounting performance measures will increase when uncertainty increases (e.g. Ezzamel (1990) and Macintosh & Daft (1987)). Finally, Govindarajan (1984) hypothesised a difference in perceived evaluative style across business units with differences in perceived environmental

uncertainty. He found support for this hypothesis: the higher perceived environmental uncertainty, the more superiors used a subjective performance evaluation and reward system rather than a formula-based approach.

To explain these mixed findings, Hartmann (2000, p. 472) argues that “uncertainty provides us with a paradox for the design of management control systems”. To understand this paradox, we should recognise that management control theory distinguishes between the performance evaluation of organisational units, i.e. economic performance, and the performance evaluation of managers, i.e. managerial performance. Theoretically, evaluating managerial performance requires different (accounting) information than evaluating economic performance, because for motivational purposes, a manager should only be held responsible for (and evaluated in terms of) results that he is able to influence. This is known as the controllability principle. Using the same accounting information for evaluating organisational units and managers could result in behavioural displacement (Merchant, 1998), especially when uncertainty is high (Merchant, 1990). Hartmann (2000, p.473) therefore concludes that the controllability principle provides us with a paradox as to the appropriateness of using budgets as a control tool in situations where uncertainty is high, because in these situations the feasibility of formal organisational controls -such as budgets- is apparently limited while, simultaneously, they may be most needed:

“On the one hand, decentralisation is an important means of enhancing the controllability of organisational processes, and budgets are important tools to formalise decentralisation, by means of creating responsibility centres (cf. Bruns & Waterhouse, 1975; Merchant, 1981, 1984). On the other hand, a strict application of the controllability principle would hinder the subsequent use of the budget for managerial performance evaluation.”

In my opinion, this paradox may be explained by considering the different ways in which accounting information in general and budgets in particular are used within organisations. The empirical findings and conclusions based on the controllability principle support the distinction between the **technical design** of management control systems such as budgetary control and performance evaluation systems, and the **use of** (information from) these systems by managers within the organisation to evaluate performance. We may conclude therefore that when uncertainty increases, the use of formal organisational controls, and, thus, the emphasis on accounting performance measures in general may increase. However, accounting information serves several roles within an organisation, only one of which is performance evaluation. Thus, because of the limited feasibility of these performance measures for performance evaluation purposes when uncertainty is high, to evaluate the managerial performance of subordinates, superiors may not rely heavily on these performance measures, but instead use a more subjective evaluation or relative performance evaluation (cf. Merchant, 1989; Govindarajan, 1984), or be more willing to accept subordinate’s explanation of his performance (cf. Ezzamel, 1990). Despite the limited use for performance evaluation ex post, ex ante accounting performance measures may play a

central role within the organisation for planning, coordination, information, and motivation purposes, especially when uncertainty is high (cf. Chapman, 1998).

These conclusions are supported by Chenhall's (2003) review of contingency-based research on the relationship between (environmental) uncertainty and management control systems design. He summarises his findings from the review in the following three propositions (Chenhall, 2003, p.138):

1. The more uncertain the external environment, the more open and externally focused the management control system.
2. The more hostile and turbulent the external environment, the greater the reliance on formal controls and an emphasis on traditional budgets.
3. Where management control systems focused on tight financial controls are used in uncertain external environments, they will be used together with an emphasis on flexible, interpersonal interactions.

In short, several contextual factors at the organisational level may explain differences in (the distribution of) evaluative styles across different organisations. Based on the discussion above, table 4.2 summarises the expected influence of these factors on a number of underlying dimensions of perceived evaluative style as they have been identified by Otley & Fakiolas (2000). The relationships as identified in table 4.2 serve as background information against which the relevance of dimensions of perceived evaluative style that appear empirically within a specific context can be interpreted. As explained above, the relationships in table 4.2 cannot be tested within the context of a single organisation. For this reason, these relationships are not presented as propositions here.

4.2.3 Factors at the level of groups

There are not only perceived differences in evaluative styles between organizations, but there are different styles of evaluations within organizations as well. Otley (1978), for example, identified five distinct evaluative styles within a single organization. These styles were based on the perception of subordinates of how their bosses evaluated their performance. Otley's (1978) sample consisted of 41 operating unit managers selected from three geographically different groups within one single company. Each group was headed by one group manager. His results (p.140) indicated that

'despite considerable differences in perceived evaluative style perceived by managers in each group, there were significant differences between groups... These perceived differences correspond very closely to the nuances of style that senior group managers intend to transmit to their subordinates.'

Otley's findings imply three different things. First, a distinction can be made between **intended evaluative style** and **perceived evaluative style**. Intended evaluative style reflects

Table 4.2 Factors at the organizational level identified from previous literature and the expected relationship with underlying dimensions of perceived evaluative style

Factor	Expected relationship	Underlying dimensions of perceived evaluative style
A healthy financial condition	is positively related to	<ul style="list-style-type: none"> • An emphasis on other quantitative measures of performance rather than budgets; • A flexible use of quantitative measures rather than a rigid use • An emphasis on absolute measures of performance rather than comparison with pre-set targets; • An emphasis on subjective criteria of performance rather than objective criteria of performance; • An emphasis on long-term rather than short-term performance.
		<ul style="list-style-type: none"> • An emphasis on other quantitative measures of performance rather than budgets;
		<ul style="list-style-type: none"> • An emphasis on absolute measures of performance rather than comparison with pre-set targets;
		<ul style="list-style-type: none"> • An emphasis on subjective criteria of performance rather than objective criteria of performance;
		<ul style="list-style-type: none"> • An emphasis on long-term rather than short-term performance.
A high level of management hierarchy	is negatively related to	<ul style="list-style-type: none"> • An emphasis on quantitative measures of performance, whether budgets or otherwise;
		<ul style="list-style-type: none"> • An emphasis on absolute measures of performance rather than comparison with pre-set targets;
		<ul style="list-style-type: none"> • An emphasis on subjective criteria of performance rather than objective criteria of performance;
		<ul style="list-style-type: none"> • An emphasis on long-term rather than short-term performance.
		<ul style="list-style-type: none"> • An emphasis on quantitative measures of performance, whether budgets or otherwise;
Environmental Uncertainty	is positively related to	<ul style="list-style-type: none"> • An emphasis on absolute measures of performance rather than comparison with pre-set targets; • An emphasis on subjective criteria of performance rather than objective criteria of performance; • An emphasis on short-term rather than long-term performance; • A flexible rather than rigid use of quantitative measures of performance.
		<ul style="list-style-type: none"> • An emphasis on other quantitative measures of performance rather than budgets;
		<ul style="list-style-type: none"> • A flexible use of quantitative measures rather than a rigid use
		<ul style="list-style-type: none"> • An emphasis on absolute measures of performance rather than comparison with pre-set targets;
		<ul style="list-style-type: none"> • An emphasis on subjective criteria of performance rather than objective criteria of performance;
Type of responsibility centre	In comparison to cost centres, profit centres will be associated more with	<ul style="list-style-type: none"> • An emphasis on long-term rather than short-term performance.
		<ul style="list-style-type: none"> • An emphasis on long-term rather than short-term performance.
		<ul style="list-style-type: none"> • An emphasis on long-term rather than short-term performance.
		<ul style="list-style-type: none"> • An emphasis on long-term rather than short-term performance.
		<ul style="list-style-type: none"> • An emphasis on long-term rather than short-term performance.

Table 4.2 (continued)

Strategy	<div>In comparison to defenders, prospector firms will be associated more with</div> <div><ul style="list-style-type: none">• An emphasis on other quantitative measures of performance rather than budgets;• A flexible use of quantitative measures rather than a rigid use• An emphasis on absolute measures of performance rather than comparison with pre-set targets;• An emphasis on subjective criteria of performance rather than objective criteria of performance;• An emphasis on long-term rather than short-term performance.</div>
----------	---

the general manner in which a superior says he is evaluating his subordinates, while perceived evaluative style reflects the manner in which individual subordinates say their manager is evaluating their performance. The perceived evaluative style may or may not reflect the intended evaluative style. Second, between-group differences in perceived and intended evaluative style are likely to exist, a conclusion that is in line with Hopwood's (1972a) findings. Apparently, different superiors do evaluate their subordinates in different ways, which partly explains differences in perceived evaluative style between subordinates who report to different superiors. This leads to the following proposition:

- P1. *There are significant between-group differences in perceived evaluative style (where groups are subordinates reporting to the same superior).*

Third, although between-group differences exist, there are considerable within-group differences in perceived and intended evaluative style as well. Thus, within-group differences in perceived evaluative style could occur because a superior may actually evaluate different subordinates in different ways and/or because individuals may perceive the intended style differently. This leads to the following proposition:

- P2. *There are considerable differences in perceived evaluative style within groups of subordinates reporting to the same superior*
(or, stated in the null form, where I expect to reject the null form:
There is considerable agreement on perceived evaluative style within groups of subordinates reporting to the same superior.)

In this section I will identify factors that may explain between-group differences, in the next section factors that may explain within-group differences.

Otley's (1978) findings suggest that perceived differences in evaluative style across groups can be partly explained by differences across superiors in **intended evaluative style**. This in turn raises the question why superiors may "choose" or adopt a certain style (between-group differences).

An important contextual factor that may partly explain between-group differences in evaluative style within a single organisation is **task uncertainty**. **Task uncertainty** can be defined as:

"uncertainty associated with task outcomes, which is caused by the complexity and diversity of tasks performed." (Hartmann, 1997, p. 49)

In a recent review of contingency-based research on management control systems design, Chenhall (2003, p. 141) concludes that the higher task uncertainty, the more informal the

controls will be, including, for example, less reliance on standard operating procedures, less reliance on accounting performance measures, higher participation in budgeting, and more personal and clan controls.

Similarly, in a theoretical paper, Hirst (1981) argues that the appropriateness of the reliance on accounting performance measures depends on the level of task uncertainty. He states that task uncertainty could help explain the difference between Hopwood's (1972a) and Otley's (1978) findings. Hirst (1981, p. 781) claims that although Hopwood (1972a) did not measure task uncertainty, it can be inferred that some uncertainty existed:

"In the organisation which served as the source of empirical data, Hopwood (1973, pp. 41-42, 191) suggests that the departments` he studied were highly inter-dependent. Thompson (1967, p. 159) claims that an important source of uncertainty is task interdependence. Given that some degree of uncertainty was associated with the tasks undertaken in the departments, then Hopwood's sample of departments may not have included evaluative situations in which task uncertainty was low."

In contrast,

"Otley (1978, p.123) was concerned with observing the operation of a well-designed accounting system in a type of organisation that was well suited for the application of budgetary control. The units studied were described (p. 126) as being independent of each other, and operating in stable, closed environments. In terms of the previous analysis, the tasks undertaken in these units faced relatively low task uncertainty."

Clearly, Hirst (1981) uses task uncertainty as if it operates at the level of particular organisations, or, at least, as a descriptive general characteristic of a total sample. This implicitly assumes that there will be little variance in individual managers' perceptions of task uncertainty. However, the definition of task uncertainty suggests that task uncertainty is a job-related characteristic that operates at the level of tasks. This could imply that the appropriate level of analysis of task uncertainty is the unit (work group) level when individuals face similar tasks (Van de Ven & Delbecq, 1974; Whitey et al., 1983; Brownell & Dunk, 1991, p. 696), or, perhaps, the level of an individual when each individual faces a different, specific task. Thus, although similar to environmental uncertainty, uncertainty stemming from task uncertainty operates at a lower level. While environmental uncertainty is an organisational level variable, task uncertainty operates at the level of tasks (cf. Hartmann, 2000). Therefore, if differences exist in the level of task uncertainty across groups, Chenhall's (2003) proposition can be applied to groups. This suggests that, given a certain technical design of the evaluative system, superiors may use this system in different ways, depending on the level of task uncertainty in the work group.

In short, I propose the following:

- P3. *As task uncertainty increases, evaluative styles will be characterised by less emphasis on budgets and other quantitative targets, a higher emphasis on qualitative, interpersonal aspects of performance, and a higher willingness of superiors to listen to subordinates' explanations of their performance (a more flexible use of budgets and other quantitative targets).*

4.2.4 Factors at the individual level: subordinate characteristics

Even though Otley (1978) found differences in perceived evaluative style between groups reflecting differences in intended style, he also concluded (p.141):

"It is evident that unit managers were perceiving differences in perceived evaluative style that were intended to be transmitted by group managers. However, it is also clear that there was a significant degree of bias in either group managers' projection of their intended style or of unit managers' perceptions of it."

Thus, subordinates who are evaluated by the same superior may report different perceived evaluative styles that do not accurately reflect intended differences in evaluative style. Kahn (1972) raises related issues in a discussion of Hopwood's (1972a) findings. According to Kahn (1972), Hopwood's measurement of perceived evaluative style, which is based on the responses of subordinates, casts doubt on the objective accuracy of the report of subordinates about their superior's style of evaluation. He (1972, p. 185) argues that

"ideally, this question would be answered by means of expert observation or scoring of evaluative transactions between manager and superior. Less expensive and more feasible within the present (i.e. Hopwood's (1972a)) design would be analysis of agreement and disagreement between the manager and the evaluating supervisor, and among the managers evaluated by the same supervisor."

In his response to Kahn (1972), Hopwood (1972b) agrees that the use of expert observation and the scoring of evaluative transactions "are promising" means to improve the measuring and validating of perceived evaluative style. He adds that "perhaps even more objective means can be derived from an analysis of formal evaluation reports and the reported budgetary data" (p. 192). However, Hopwood (1972b, p. 193) is more pessimistic about Kahn's suggestion to validate the reports by analysis of agreement and disagreement between the manager and the evaluating supervisor, and among the managers evaluated by the same supervisor:

"... There is no reason to always expect self-reported behaviour to correspond to descriptions by others. As recent studies have suggested, the two sets of reports are different, one at times reflecting no more than intended behaviour. There are similar problems with presupposing

agreement among the managers evaluated by the same supervisor. Disagreement may, of course, reflect biased reports, although several previous studies have found that supervisors do vary their styles in relation to different subordinates. In practice, the data needs analyzing in this manner, although it is also necessary to move beyond looking for simple correspondence to consider the reasons for agreement or otherwise.”

The discussion above supports proposition 2: although perceived evaluative style may partly be a group level phenomenon, differences in perceived evaluative style between subordinates who are evaluated by the same superior are likely to occur as well. In fact, in the empirical study reported in chapter 3 I also found and reported differences in perceived evaluative style between branch managers reporting to the same area manager³⁷. The review above also suggests that at least part of perceived differences in evaluative style within groups may reflect intended differences in evaluation. Therefore, a possible explanation for these within-group differences in perceived style is that superiors adapt their style of evaluation to the specific, individual characteristics of his or her subordinates (cf. Hopwood, 1972a).

To my knowledge, the RAPM-literature has not addressed within-group differences, even though this issue was raised in the early studies of Hopwood (1972a) and Otley (1978). Therefore, the RAPM-literature offers no explaining variables for differences in perceived evaluative style within groups. Based on the Situational Leadership theory, however, I expect that three individual-level factors may explain why a superior may vary his style of evaluation in relation to different subordinates are **age**, **tenure** and **education**. According to the Situational Leadership theory, as developed by Hersey & Blanchard (1993), the ability of people is a function of knowledge, experience, and skill or (previous) performance. Consequently, age, tenure, and, perhaps to a lesser extent, education may influence the ability of managers to fulfil their tasks. Because of differences in ability across subordinates, it is likely that superiors will evaluate subordinates in different ways. Particularly, leadership literature thus suggests that young, inexperienced managers will need more structure and guidance from their superior than older, experienced managers.

The following propositions summarise the review above:

- P4. Within-group differences in age, tenure, and education of subordinates are related to within-group differences in perceived evaluative style.*
- P5. Within-group differences in perceived evaluative style partly reflect differences in evaluative style that the superior intended.*

³⁷ See § 3.4.5 and § 3.5.2.

4.3 Superior characteristics: Leadership style

Evaluating subordinates is just one of many things expected by superiors in an organisation. How a superior evaluates his subordinates, i.e. his evaluative style, is most likely part of the more general way in which a superior approaches his job as a leader. Therefore, I expect that differences in perceived evaluative style are part of more general differences in perceived leadership style (cf. Hopwood, 1973; 1974). Hopwood (1973; 1974) used leadership style to provide further evidence on the validity of the distinction among the three perceived evaluative styles that he identified³⁸. To measure leadership style, Hopwood used the two dimensions of leadership style taken from the Ohio State University leadership studies, i.e. Consideration and Initiating Structure. Judge et al. (2004, p. 36) summarise the dimensions as follows: "Consideration is the degree to which a leader shows concern and respect for followers, looks out for their welfare, and expresses appreciation and support (Bass, 1990). Initiating Structure is the degree to which a leader defines and organizes his role and the roles of followers, is oriented toward goal attainment, and establishes well-defined patterns and channels of communication (Fleishman, 1973)." Hopwood (1973; 1974) was able to show that respondents gave both budget-constrained and profit conscious supervisors high ratings on the Initiating Structure dimension, whereas non-accounting supervisors were given significantly lower ratings on this dimension. Furthermore, profit-conscious and non-accounting supervisors were given significantly higher ratings on the Consideration dimension than budget-constrained supervisors were. According to Hopwood (1974, p. 493), "the evidence allows the initial distinctions which were made among the three styles of using accounting data in performance evaluation to be placed in a wider managerial perspective". And he concludes by stating (p. 495):

"The present findings suggest, however, that the relationship between accounting data and decision behaviour is moderated by the leadership style and the social context of the decision maker. Much greater explicit consideration needs to be given to these factors in further research, although, in so doing, there may be a need to rely less on the abstractions of laboratory experimentation and more on the complexities of field research."

Surprisingly, despite these conclusions, leadership style has been ignored in the RAPM-literature ever since. Besides Hopwood (1973; 1974), no other study has investigated perceived evaluative style as part of more general patterns of superiors' managerial behaviour as reflected in, for example, leadership style. This may have contributed to the validity problems with the measurement of perceived evaluative style as described in chapter 3. There are two reasons for including leadership style in future research on perceived evaluative style. First, given the current state of the literature and the exploratory character of the research that is presently needed, it is still impossible to identify different

³⁸ See § 3.2.2.

"patterns" of perceived evaluative style, or even relevant, single dimensions of perceived evaluative style a priori. Yet, assessing and reporting the validity and reliability of the measurement of perceived evaluative style is important. To validate the interpretation of differences in perceived evaluative style, linking single dimensions or "patterns" to other variables that have been shown to validly and reliably measure related aspects of superior behaviours would be helpful. Secondly, the management accounting and control literature provides some evidence that leadership style has a relationship with subordinates' managerial behaviour and managerial performance. Otley and Pierce (1995), for example, using the same two dimensions as Hopwood (1973) in a survey under all audit seniors in three Big Six audit firms, concluded that a leadership style characterised by high structure and low consideration was associated with the highest level of dysfunctional behaviour (under-reporting of time, and audit quality reduction behaviour). In contrast, a leadership style characterised by low structure and high consideration was associated with the lowest level of dysfunctional behaviour. In this respect, the inclusion of leadership style in future research provides an opportunity to develop and validate the measures and behavioural consequences of evaluative style.

Moreover, when studying the relationship between leadership and perceived evaluative style we can make good use of the developments in the leadership field. Many previous studies in the accounting literature (DeCoster & Fertakis, 1968; Hopwood, 1973; 1974; Pratt & Jiambalvo, 1981; Pratt & Jiambalvo, 1982; Brownell, 1983; Otley & Pierce, 1995) have used the two key variables from the Ohio State approach to leadership, i.e. Initiating Structure and Consideration. According to Dansereau this is now "one of the oldest approaches to leadership in organizational behaviour" (Dansereau et al., 1995). Research on leadership has, however, changed from an emphasis on exchange, reward and control towards an emphasis on vision, value transformation, symbolic behaviour and management of meaning, reflecting the more explicit focus on leaders and on the development of followers in contemporary leadership approaches than in the earlier approaches (Dansereau et al., 1995). Recent leadership theories attempt to explain how leaders are able to change people rather than respond to them, and leaders are seen as effective when they succeed in changing and creating the environment, rather than complementing the environment. These new types of leadership are often referred to as 'transformational' and 'charismatic' leadership, as opposed to the traditional 'transactional' or 'instrumental' leadership styles. This development in the leadership literature coincides with similar developments in the management control literature. While the conventional view with its strong emphasis on budgetary control systems is prevailing in most of the RAPM-literature, in the management accounting and control literature the definition of management control has evolved "to one that embraces a much broader scope of information" (Chenhall, 2003, p. 129), including external information, non-financial information, predictive information and informal personal and social controls. As a consequence, the RAPM research has hardly investigated the nature of evaluative styles

in which budgetary data is (relatively) unimportant (i.e., the so-called non-accounting category), while contemporary management control theory implies that these styles are highly relevant. Thus, to better understand these styles, it will be necessary to give account to the developments in the leadership literature.

In accordance with the changed focus of leadership approaches, Den Hartog (1997) distinguishes three different types of leadership and a residual category: inspirational, transactional, laissez-faire, and other. Because I am interested in the relationship between leadership style and perceived evaluative style, i.e. specific behaviours when evaluating subordinates as part of more general patterns of leader behaviours, and not in leadership style per se, I will limit the inclusion of leadership style to two types, i.e. inspirational and transactional. Inspirational leadership indicates a leader who is motivating, stimulating, supporting, changing his subordinates. It emphasises interpersonal aspects of leadership. Therefore, I expect that inspirational leaders will tend to attach high importance to interpersonal, qualitative aspects of performance evaluation, such as subordinates' explanations of their performance, possibilities to improve/develop their performance, their plans for the years to come, etc. On the other hand, transactional leader behaviour entails pointing out to subordinates 'what is expected of them and provide them with feedback when their behaviour is not up to standard' (Den Hartog, 1997, p. 50). Therefore, I expect transactional leaders to attach high importance to quantitative measures of performance and deviations from targets.

This results in the following propositions:

- P6. *Inspirational leadership will be more positively associated with an emphasis on interpersonal, qualitative aspects of performance evaluation than transactional leadership.*
- P7. *Transactional leadership will be more positively associated with an emphasis on quantitative measures of performance and deviations from targets than inspirational leadership.*

4.4 Consequences of perceived evaluative style: fairness of evaluation, trust in superior, and subordinate managerial behaviour

4.4.1 Introduction

In the previous chapter I concluded that in later RAPM-research the concept of evaluative style has become narrowly focused on "the extent to which a superior uses budgetary data (or other quantitative data) to evaluate the performance of subordinate managers", thus excluding the complex behavioural and attitudinal patterns in which superiors may in fact

use a much broader set of information. By excluding the behavioural dimension, recent RAPM research has also failed to appreciate the relationship between **the use** of management control tools (such as budgeting) and the behavioural impact of the **design** of these tools. Yet, this relationship was at the heart of the early studies by Hopwood (1972a; 1973) and Otley (1978) that induced later RAPM-research. Their argument was that the utilisation of accounting information for evaluative purposes (design) in itself did not pose a threat. Rather, the manner in which superiors used accounting information, depending on their judgment of the quality and relevance of accounting performance measures, influenced the subsequent functional or dysfunctional behavioural responses of subordinates. In this perspective the effectiveness of an evaluative style is determined by the quality of a superior's response to³⁹ (compensation for) the degree of "fit" between control system design and the task it has to fulfil in its immediate organisational context.

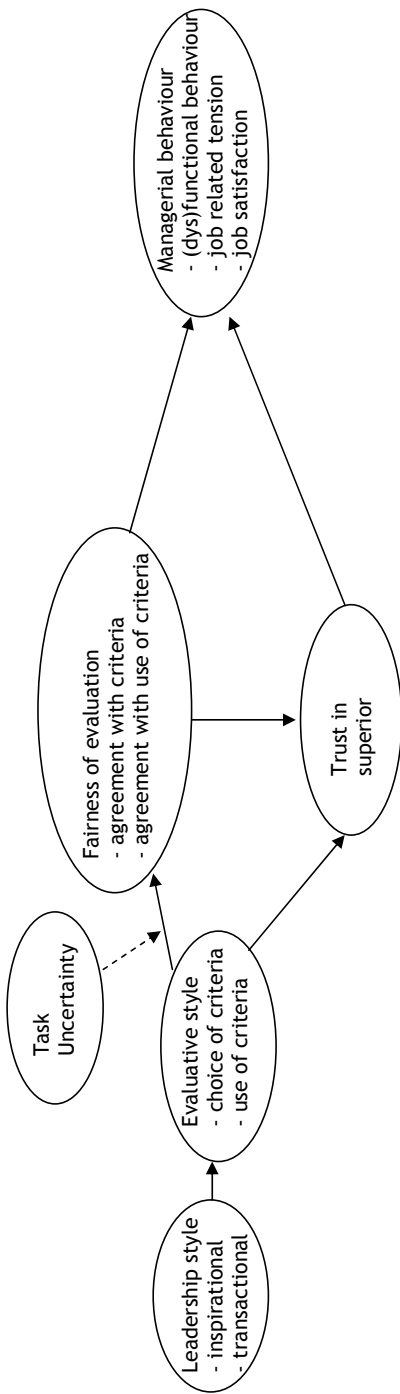
As this implies that the definition of an effective evaluative style is context-dependent, I have argued in § 3.4 that to deal with these issues effectively we need field studies studying the effectiveness of different evaluative styles in the context of a specific organisational control system, more than quantitative surveys across organisations.

Studies of single organisations allow researchers to assess the behavioural impact of differences in how superiors handle the evaluation process, i.e. evaluative style as a characteristic of superiors, while controlling for the behavioural impact of the **design**, i.e. the technical aspect or architecture of control systems (Flamholtz, 1983). Such research does not aim to identify the contexts in which the use of particular management control tools such as budgets in itself would enhance functional or dysfunctional decisions or behaviours. Rather, such research aims to identify how differences in the manner in which superiors use particular management control tools for evaluation purposes strengthen or limit the motivational and behavioural impact of these tools within a particular (organisational) context.

Therefore, in this section, I will discuss a framework that identifies factors that may explain how perceived evaluative style is associated with subordinate behaviour within the context of a specific type of management control architecture. The framework is depicted in figure 4.2. The general idea behind this framework is that I do not expect that there will be a direct effect of evaluative style on subordinate's behaviour. I expect that perceived fairness of evaluation and trust of subordinates in their superior will be important intervening variables. Particularly, I expect that evaluative styles that are perceived as fair and/or enhance the trust of subordinates in their superior will result in more positive and less negative behaviours. I will explain each of the expected relationships in more detail below.

³⁹ By using the word "response", I do not intend to suggest that evaluative style is necessarily a conscious choice of the superior, i.e. intended behaviour. This in fact is one of the issues that has remained largely unexplored despite decades of research on evaluative style, but that future research on evaluative style needs to address, as I have argued earlier in this chapter.

Figure 4.2 General framework identifying antecedents and consequences of perceived evaluative style



4.4.2 The role of fairness of evaluation

An important finding in the RAPM-literature that is worth considering in more detail, is Otley's (1978, p. 131, 133, 145) repeated observation that the prevalent norms and values in the particular organisation exerted a significant effect on the level of job-related tension. The importance of these norms and values is shown in two different ways in his results. First, managers seem to have a feeling of the appropriate level of participation. When this norm is violated in either direction, job-related tension increases. Secondly, a manager's feelings of job-related tension increase with increasing disagreement with the appropriateness of the most important criteria used in his performance evaluation. In Otley's (1978, p. 132) own words: 'There is some evidence that such tension (*i.e. job-related tension, JN*) is associated with the degree of agreement a manager has with the criteria being used to evaluate his performance and the extent to which he considers he ought to participate in setting his own budget.' (p.132)

Other authors have recognised the importance of (dis)agreement with evaluation criteria and the level of participation in setting these criteria for understanding the behavioural associations of perceived evaluative style too. Dunk (1990), for example, argues that participation and agreement between superiors and subordinates on evaluation criteria interact to affect subordinate manager's performance. Harrison (1993, p. 319), reviewing the RAPM literature, states that earlier studies implicitly assumed 'that a match between a superior's perceived evaluative style and contingency factors will mean that the style is seen as appropriate by subordinates. Subordinate agreement with the style was argued by both Otley (1978, p. 146) and Kahn (1972, p. 184) to be the crucial element in the relation between perceived evaluative style and subordinate responses.'

Although Harrison's statement seems to be in congruence with both Otley (1978) and Dunk (1990), the differences in wording are important. While Otley and Dunk both refer to agreement with evaluation **criteria** (and participation), Harrison refers to agreement with perceived evaluative style. In the previous chapter I have argued that differences in perceived evaluative style could, amongst others, refer to differences in the choice of criteria for evaluating purposes (*i.e. what* results the superior emphasises) or to differences in the use of these criteria (*i.e. how* the superior handles the evaluation **process**) or both. For an understanding of the behavioural implications of perceived evaluative style, ideally, the influence of perceived evaluative style on subordinate behaviour should be separated from the influence of the design of the evaluative system, just like in Hopwood's (1973) and in Otley's (1978) study. For it is conceivable that within a particular organisation the design of the evaluative system does not provide superiors much space in choosing the evaluation criteria and/or in how to use these criteria, while in another organisation this is left up to the superior's own choice entirely. If the choice of criteria is not a result of superior's choice, but a result of the technical design of the evaluative system that is used in a particular organisation, agreement with evaluation criteria would be a dimension of agreement with the design of the evaluative system. If the

superior is free in choosing the criteria, agreement with evaluative criteria would be a dimension of agreement with perceived evaluative style. It should be clear that that agreement with evaluation criteria is not the same as agreement with evaluative style, although they could be (strongly) associated.

The implication of all this is that agreement with evaluation criteria in itself is no guaranty that the subordinate will think the performance evaluation is fair. The perceived fairness of the performance evaluation will depend also on whether the subordinate agrees with how the superior uses these criteria when evaluating performance. The subordinate may expect that the superior will use these criteria in a certain way, for example that the superior will interpret the outcomes with care and is willing to listen to the subordinate's explanation of why actual performance is as it is. If the superior evaluates performance in a different way, the subordinate may perceive the performance evaluation as unfair. So, a subordinate may agree with the criteria used, but not with how his superior uses these criteria. The perceived fairness of performance evaluation may therefore consist of two different components: agreement with the evaluation criteria and agreement with how the supervisor uses these criteria. These are two different aspects of performance evaluation, which should be recognised as such. Most likely, it is the perceived fairness of the performance evaluation process as a whole that is the critical factor in understanding the behavioural consequences of perceived evaluative style, rather than the single components. This would be in accordance with Hartmann's (2000) suggestion to develop a framework for further research around the notion of "fairness".

We can summarise the arguments above in a general two-component model (see Figure 4.2), in which the perceived fairness of evaluation would depend on

1. the criteria that are used to evaluate subordinates' performance, and
2. the manner in which superiors use these criteria to evaluate performance.

These two components are very similar to the distinction between system procedural justice, rater procedural justice and interactional justice, as made by Erdogan (2002). Erdogan (2002, p. 559) defines system procedural justice as "the perception that the appraisal procedures adopted by the organisation are fair", rater procedural justice as "the perception that the application of procedures by raters is fair", and interactional justice as "the degree to which the interpersonal communication during the appraisal process is fair". Clearly, system procedural justice refers to the technical design of the system at the organisational level. The other two types of justice refer to fairness perceptions of the interaction between superiors and subordinates.

Based on the discussion above, I expect that the manner in which managers (are perceived to) evaluate their subordinates will affect subordinates' behaviour through perceived fairness of evaluation.

The discussion in this section results into the following propositions:

- P8. *The relationship between perceived evaluative style and managers' behaviour will be explained by an indirect effect whereby situational appropriateness of the perceived evaluative style affects fairness of evaluation, which in turn is positively related to managers' behaviour.*
- P9. *If subordinates agree with the criteria that are used to evaluate their performance, feelings of unfairness will arise if superiors ignore these criteria or do not attach as much importance to these criteria as they should according to the subordinate. On the other hand, if subordinates disagree with the criteria used to evaluate their performance, feelings of unfairness will increase as superiors attach more importance to these criteria and ignore the objections of subordinates to these criteria.*

4.4.3 Agreement with evaluative criteria: the role of task uncertainty

To explain why subordinates (dis)agree with the criteria that are used to evaluate their performance, several (motivational) theories provide a useful theoretical background. Both goal setting theory and motivation systems theory stress the importance of clear and specific goals to motivate behaviour. Therefore, I will shortly discuss the central thoughts of these theories.

Goal Setting Theory, which is among the most scientifically valid and useful theories on motivation (Locke, Latham & Erez, 1988; Ford, 1992; Pinder, 1984; Wood & Bandura, 1989), assumes that human behaviour is regulated by goals and intentions. Goals are defined as what the individual is trying to do (achieve, attain, or accomplish). According to the goal setting research, harder goals result in a higher level of performance than easy goals. Furthermore, specific (hard) goals result in a higher level of performance than do no goals or a generalized goal such as "do your best". Finally, intrinsic and extrinsic rewards and incentives, such as money, feedback, competition, and participation, only affect behaviour to the extent that they lead to the setting and/or acceptance of specific hard goals. Specific (hard) goals that are not accepted, will not lead to action. Therefore, goal acceptance, or goal commitment, is a necessary condition for goals to direct behaviour in the manner described by Goal Setting Theory. Goal commitment and goal acceptance are similar, but not identical concepts. Goal commitment implies a determination to try or keep trying for a goal, whether it is an assigned goal, a participative set goal, or a goal that one has set on

one's own. Goal acceptance implies that one has agreed to commit oneself to a goal assigned or suggested by another person (Locke *et al.*, 1981, p.143).

In **Motivation Systems Theory (MST)** Ford (1992, p.3) defines motivation as

"the organized patterning of three psychological functions that serve to direct, energize, and regulate goal-directed activity: personal goals, emotional arousal processes, and personal agency beliefs."

This definition reflects the fact that MST is designed to represent the selective **direction** of behaviour patterns (where people are heading and what they are trying to do); the selective **energisation** of behaviour patterns (how people get "turned on" or "turned off"); and the selective **regulation** of behaviour patterns (how people decide to try something, stick with it, or give up) (Ford, 1992, p.2-3). Symbolically, this can be expressed as:

Motivation = Goals x Emotions x Personal Agency Beliefs.

Because each of the motivational components is necessary to motivate people successfully, but none of them sufficient, all three components should be facilitated in order to motivate people successfully (Ford, 1992, p.80). The most important component of motivation in MST, however, seems to be people's goals. In a review of Ford's book, Pintrich (1995, p.1000) writes that although MST is an integrative motivational theory, he would classify MST as basically a goal theory approach to motivation.

MST is based on the assumption that people's behaviour is often guided by the simultaneous pursuit of multiple goals. Ford (1992, p.116) argues that to determine which goals should be prioritised in a specific situation, the criteria used generally focus on one or more of the following properties:

1. goal relevance (i.e., what goals are meaningful or appropriate in a particular context);
2. goal importance (i.e., to what extent are the relevant goals in a particular context personally significant to the individual);
3. goal attainability (i.e., personal agency beliefs); and
4. emotional salience of the actions and consequences associated with pursuing and achieving the goal.

Once a specific goal is (or multiple goals are) "chosen"⁴⁰, the importance of the different motivational components tends to shift from goals to personal agency beliefs and emotions.

⁴⁰ The goal in place does not necessarily have to be chosen consciously. A person may very well unconsciously try

From this it follows that personal agency beliefs only matter if people pursue a goal (or multiple goals) that is relevant and important to the individual (the first two criteria mentioned above). This is also implied by the definition Ford (1992, p.125) gives of personal agency beliefs:

“Personal agency beliefs are evaluative thoughts involving a comparison between a desired consequence (i.e., some goal) and an anticipated consequence (i.e., what the person expects to happen if they pursue that goal).”

Therefore Ford (1992, p.126) warns that personal agency belief information should be carefully interpreted if it is not accompanied by evidence verifying that these beliefs reflect significant personal concerns. Negative personal agency beliefs and emotions (i.e., anxiety, worry, despair etc.) cannot emerge as major motivational barriers until people realize the importance of achieving the particular goal. On the other hand it is also true that simply having a relevant and important goal is not enough to motivate people. People must also believe that they have the capabilities and the opportunities that are needed to attain this goal. So, in personal agency beliefs a distinction should be made between capability beliefs and context beliefs. Capability beliefs are defined by Ford (1992, p.124) as ‘evaluations of whether one has the personal skill needed to function effectively’, while context beliefs refer to ‘evaluations of whether one has the responsive environment needed to support effective functioning’. According to Ford (1992, p.124), personal agency beliefs are particularly crucial for motivating people in situations in which they have the relevant skills and a responsive environment, but have not used them as much as possible, i.e. in situations in which challenging but attainable goals were set.

Identical to what is written about personal agency beliefs, Ford (1992, p.140) claims that emotions only affect behaviour if a goal is in place that is relevant and important, i.e. a goal that is currently directing or influencing the individual's activity. The difference with personal agency beliefs is that personal agency beliefs are mostly concerned with long-term consequences of certain actions and with identifying alternative actions, while emotions focus mostly on short-term regulatory problems and opportunities.

Clearly, both Goal Setting Theory and Motivational Systems Theory stress the importance of goal characteristics (specific, hard, clear, etc.) both to motivate people and to understand people's behaviour. Accounting performance measures, such as budgetary targets, often provide a clear, specific, measurable goal, so motivation theory may be relevant. Motivation theories, like Goal Setting theory and Motivational Systems Theory, however also show that such clear goals will only be motivating to people when they agree with them. Motivational Systems Theory states that whether people agree with the goals and are motivated by them

to attain a goal.

depends, among others, on context and capability beliefs about the attainability of the goals. Analogous, in a performance evaluation context, **characteristics of the evaluative criteria** that a superior uses to evaluate subordinate managers' performance are most likely important determinants of subordinate's agreement with these criteria.

The importance of characteristics of evaluative criteria as an explanation of why different perceived evaluative styles relate to different managerial behaviours has been recognized within RAPM research from Hopwood's (1973) study onwards. As discussed in chapter 2, in explaining his findings, Hopwood (1973) pointed out the technical shortcomings of accounting performance measures in the specific organisational context that he studied. In contrast, Otley (1978, p.123) observed "the operation of a well-designed system in a type of organization that was well suited for the application of budgetary control", to eliminate the influence of technical shortcomings of accounting performance measures on the findings. Therefore, Otley (1978) claimed that the difference in results with Hopwood's study was a consequence of accounting performance measures being more complete measures of performance in the organisation he studied. In a theoretical paper trying to reconcile the results from Hopwood (1972a) and Otley (1978), Hirst (1981) has argued that dysfunctional behaviours may result when subordinates perceive measures used in the evaluation of their performance as being incomplete, but not when they perceive them as complete. While Hirst only considers incomplete measures in his paper, in a footnote he refers to other perceived characteristics of accounting performance measures which have also been associated with dysfunctional behaviour: the inability to interpret (Argyris, 1952; Hofstede, 1968), the uncontrollability, and the subjectivity (Ridgway, 1956). Hirst (1981; 1983) was also the first to argue that the completeness of accounting measures would depend on the level of **task uncertainty** that managers face, suggesting the inclusion of this variable in further research. Several other authors have repeated Hirst's (1981; 1983) argument (Govindarajan, 1984; Brownell, 1985; Brownell (1987)), hypothesising that the higher (task) uncertainty⁴¹, the more inaccurate accounting performance measures will be, because of the increasing uncontrollability, incompleteness and irrelevance of such measures. Therefore, these authors concluded that high reliance on accounting performance measures when (task) uncertainty was high would result in an increase in tension. Conversely, where task uncertainty was low, Hirst (1983) found that a decrease in the reliance on accounting performance measures resulted in higher reported tension too. To explain these findings, Hirst (1983) argues that the reported tension increases either because the replacement performance measures are more general and less precise, thereby increasing role ambiguity, or because the

⁴¹ Many studies refer to uncertainty in general rather than to task uncertainty as it was defined previously. Hartmann (2000) argues that it is essential to differentiate between environmental uncertainty and task uncertainty. The first refers to uncertainty arising from the external environment of a particular organisation, while the latter refers to uncertainty that is a job characteristic. In this thesis I will focus on task uncertainty. The general hypothesis as stated here is applicable to task uncertainty.

replacement performance measures are incomplete, thereby inducing conflict over the way in which performance is evaluated.

In contrast to the typical hypothesis in RAPM studies that task uncertainty is negatively related to the appropriateness of the reliance on accounting performance measures when evaluating subordinates, similar to the influence of task uncertainty on the emergence of perceived evaluative style, Hartmann (2000) has shown that the empirical evidence supports both a negative and a positive effect of task uncertainty on the appropriateness of (reliance on) accounting performance measures. The negative relationship could be explained by the limited feasibility of accounting performance measures when task uncertainty increases, because of the increasing uncontrollability, incompleteness and irrelevance of such measures. The positive relationship could be explained by the increased need of performance measures that are clear, specific, and objective. Because accounting performance measures are relatively specific and well-defined, accounting performance measures may provide a clear direction to subordinates, thereby providing structure and guidance that is especially useful when subordinates face complex and diverse tasks. Yet, as indicated earlier, providing structure and guidance may be especially important *ex ante*, or, in management control terminology, for planning and motivating purposes, but not for performance evaluation *ex post*. Therefore, I expect that as uncertainty increases, the use of quantitative measures to evaluate performance, whether budgets or otherwise, will lead to higher disagreement with the evaluative criteria.

The discussion above results in the following propositions:

- P10. The level of agreement with evaluative criteria reported by subordinates can be explained by giving explicit consideration to characteristics of the criteria that are used to evaluate performance in combination with the level of task uncertainty:*
- P10a. As task uncertainty increases, the felt appropriateness of accounting performance measures will decrease.*

Recall, however, that higher disagreement with evaluative criteria does not necessarily imply higher job related tension. My two-component model suggests that the perceived fairness of the evaluation process cannot be explained solely in terms of agreement with evaluation criteria. Agreement with how the superior uses these criteria has to be part of the explanation as well. The manner in which the criteria are used will become more important when subordinates disagree with the criteria. This leads to the following two propositions:

- P10b. Evaluative styles that are characterised by an emphasis on quantitative measures of performance and deviations from targets will result in lower levels of*

perceived fairness of evaluation in high task uncertainty situations than in low task uncertainty situations.

- P10c. *Evaluative styles that are characterised by an emphasis on qualitative, interpersonal measures of performance will result in the same level of perceived fairness of evaluation in high task uncertainty situations as in low task uncertainty situations.*

4.3.3 The role of trust

Trust is conceptualised in this study as the trust subordinates have in their superiors, that is, interpersonal trust. Therefore, trust is taken as a characteristic of the relationship between a subordinate and his superior. This conceptualisation is consistent with the conceptualisation of trust in prior RAPM studies (Hopwood, 1973; Otley, 1978) and other relevant accounting studies (Lau & Buckland, 2001; Lau & Tan, 2006), but also with the conceptualisation of trust in other studies on the perceived fairness and accuracy of performance evaluations (e.g. Fulk, Brief & Barr, 1985) and studies from the leadership/ management field (e.g. Pillai et al., 1999).

Prior RAPM studies suggest that trust is most likely positively correlated to the level of agreement with perceived evaluative style (Hopwood, 1972a; Otley, 1978). Hopwood (1972a) hypothesised that subordinates who disagree with the perceived evaluative style of their superior are more likely to report lower levels of trust in their supervisor than subordinates who do agree with the perceived evaluative style of their superior. His results confirmed this hypothesis. Similarly, Otley's (1978) results indicate that disagreement with the level of influence a subordinate perceives he has on setting his budget is associated with lower levels of trust in superiors.

In a recent study, Lau & Tan (2006) hypothesised and empirically found a positive correlation between procedural fairness and trust in superiors. Additionally, they found empirical support for their hypotheses that both procedural fairness and trust would reduce job related tension. Finally, through path analysis, they also found support for their hypothesis of an indirect effect of procedural fairness on job related tension through trust.

Fulk, Brief & Barr (1985) provide further evidence for the importance of trust in relation to the level of agreement with how a superior evaluates performance. They investigate predictors of perceived fairness and accuracy of performance evaluations in a sample of telecommunications engineers engaged in research and development activities. Their results suggest that the perceived fairness and accuracy in performance evaluation "may depend as heavily on the level of trust in the on-going superior-subordinate relationship as on characteristics of the performance appraisal process itself" (p. 301). The characteristics of

the performance appraisal process in their study refer to (1) supervisor's knowledge and (2) development of action plans related to performance weaknesses. Supervisor's knowledge refers to "the degree to which the superior is willing to obtain the appropriate information required to provide informed feedback- and to communicate that sincerity to the subordinate", while development of action plans related to performance weaknesses refers to "the degree to which performance appraisal was conducted with a view toward creating constructive, relevant outputs" (p. 309). Their results, obtained by the method of path analysis, suggest that:

"subordinate perceptions of the supervisor's knowledge, in addition to being directly related to perceived fairness and accuracy, acts indirectly through plans related to performance and trust in superiors" (p. 310).

In contrast to trust and the characteristics of the appraisal process, more formal system characteristics, such as the existence of a formal evaluation program and the frequency of evaluation, "may be relevant but nevertheless insufficient conditions for perceptions of fair and accurate performance appraisals." (p.310)

In a discussion of their results for trust, Fulk, Brief & Barr (1985, p.310) conclude:

"It is possible that the trust variable was particularly salient here due to the nature of the sample. R & D engineers and scientists tend to perform a greater proportion of less-well-structured activities, making it difficult to develop and apply a preponderance of clear-cut, objective performance standards (Pelz & Andrews, 1976). Because a greater proportion is not objectively verifiable in the short term, more reliance must be placed on the supervisor's subjective judgments. This may inflate the importance of the subordinate's confidence and trust in the supervisor."

These results not only imply that trust may be important in relation to the level of agreement with how the superior uses the performance criteria, they also support my theoretical model and previously developed propositions. First, these results confirm the importance of my two component model that distinguishes between the performance evaluation criteria and the manner in which a superior actually uses these criteria, while recognising the interaction of the two components on the perceived fairness of performance evaluation. Additionally, these results also back up my conclusion that the level of task uncertainty may be an important contextual factor that helps explain the level of agreement with the evaluative criteria, as "less-well-structured activities" implicitly refers to the level of task uncertainty as defined earlier in this chapter. Finally, the importance of including the design of the evaluative system as a contextual factor that should be clearly delineated from the manner in which superiors use this system is strengthened by these findings.

Fulk et al. (1995) treat trust as an antecedent variable of perceived fairness of performance appraisal. In contrast, Hopwood (1972a), Otley (1978) and Lau & Tan (2006) treat trust as a consequence of fairness perceptions. Other studies (Aryee et al., 2002; Pillai et al., 1999) also support the role of trust as a consequence of justice perceptions. Indeed, in these studies trust has been shown to act as an intervening variable between justice perceptions and behavioural or attitudinal outcomes. For example, in a study from the organisational behaviour field, Aryee et al. (2002) investigate the role of trust as an intervening variable between justice and work outcomes. In particular, amongst others, they hypothesise an indirect effect of interactional justice on organisational citizenship behaviour and task performance through trust in superior. Using LISREL, their results show that trust in superior fully mediated the relationship between interactional justice and work-related behaviours of organisational citizenship behaviour and task performance. Pillai et al. (1999) investigate both fairness perceptions and trust as mediators of the relationship between transformational and transactional leadership and three outcome variables, i.e. job satisfaction, organisational citizenship behaviours, and organisational commitment. Their results suggest (Pillai et al., 1999, p. 923) that "transformational leadership seems to influence procedural justice, which in turn builds trust", while "transactional leadership, on the other hand, only appears to influence distributive justice and has no impact on trust". The results also supported the indirect effect of transformational leadership on organisational citizenship behaviour through procedural justice and trust. No indirect effect was found for either job satisfaction or organisational commitment.

Based on the review above, in short I propose the following:

- P11. *The relationship between perceived evaluative style and subordinate manager's behaviour will be explained by an indirect effect whereby perceived evaluative style affects trust in superior, which in turn is positively related to subordinate manager's behaviour.*
- P12. *The relationship between perceived evaluative style and trust in superior will be explained by an indirect effect whereby perceived evaluative style affects perceived fairness of evaluation, which in turn is positively related to trust in superior.*

4.5 Conclusion

In this chapter I have reviewed the literature to identify antecedents and consequences of evaluative style. Based on a review of existing measures of evaluative style, I have concluded in chapter 3 that conducting studies in single organisations seem relevant for advancing the RAPM field. In this chapter, I have reviewed the literature to identify what this means for our understanding of antecedents and consequences of perceived evaluative style. Below I will

discuss how the findings from the review can benefit future research on evaluative style, and, thus, have been incorporated into the empirical research to be reported in the remainder of this thesis.

Relationship between contextual factors and evaluative style

With regard to antecedents, the review suggests that contextual factors at different levels may help explain why differences in evaluative style emerge. Although I developed some propositions with regard to contextual factors that could explain differences between organizations (see table 4.2), in studies that are limited to single organizations it is impossible to test these propositions. However, explicit consideration should be given to a description of the organizational context in which future studies take place, in terms of factors that may influence perceived evaluative style. This way, in discussing the empirical findings, these organizational level factors can be used to gain a deeper understanding of these findings. Thus, in future studies data should be collected on at least some of these factors, through interviews, personal observation, or documents and charts available from the organisation, or, preferably, a combination of these data sources.

Besides contextual factors that may explain differences in perceived evaluative styles across organisations, there are also contextual factors that may explain differences in perceived evaluative style between subordinates grouped by superior or organisational unit within a single organisation. Two important contextual factors at the group level that have been identified in this chapter are the intended evaluative style of the superior and task uncertainty. To test the relationships between task uncertainty and evaluative style at the group level, a measurement issue arises. To measure perceived evaluative style at the group level, future studies should aggregate the responses of subordinates reporting to the same superior in order to get a reliable "estimate" of the evaluative style of the superior, as perceived by subordinates. The aggregation of responses would only be valid if there is considerable agreement on reported perceived evaluative styles between subordinates. Thus, future studies need to assess the level of agreement on perceived evaluative style within groups. High levels of agreement within groups, where groups are defined as subordinates reporting to the same superior, may suggest that perceived evaluative style reflects a trait or inherent or consciously chosen attitude of the superior, so that there is consistency in the manner in which he evaluates subordinates' performance. On the other hand, if there is hardly any agreement within groups, this may suggest that either subordinate's responses are not very accurate, or that perceived evaluative style is more situational in nature, so that a superior will evaluate different subordinates in different ways. The review of the literature showed that although some agreement on perceived style can be expected, there is also some evidence in the literature that superiors will change their style of leadership and their style of evaluation depending on individual subordinates characteristics. Rather than relying on data obtained from subordinates only, future studies should also gather interview data from superiors to explore why superiors evaluate subordinates in different or similar ways

(i.e., data on intended evaluative style). By focusing more on the superior and his intended style in addition to evaluative style as perceived by subordinates, future studies could provide a better understanding of the "nature" of evaluative style than research up to date has presented.

Relationship between leadership style and evaluative style

The literature review also suggests a relationship between perceived leadership style and perceived evaluative style. This relationship has been discussed separately, because this relationship is theoretically different than the relationship between contextual factors and evaluative style. In fact, many of the contextual factors are likely to influence perceived leadership style in the same way as they influence perceived evaluative style, since leadership style and evaluative style are similar concepts. Furthermore, leadership style has been consistently related to managers' attitudes and behaviours. In this respect, given the inevitable exploratory character of research on evaluative style, the inclusion of leadership style in future research provides an opportunity to develop and validate the measures and behavioural consequences of evaluative style.

Relationship between perceived evaluative style and subordinate managers' behaviour

As to the consequences of perceived evaluative style, the literature has been reviewed to develop a framework that links perceived evaluative style to subordinate managers' behaviour. The framework is depicted in figure 4.2. The review suggests that examining the role of variables that intervene the link between perceived evaluative style and subordinate managers' behaviour could contribute to an improved understanding of the behavioural consequences of evaluative style. The framework indicates that fairness of evaluation and trust in superior may be important intervening variables. Future research will be needed, however, to empirically test the relationships outlined in this framework and summarized in the propositions. Since many of the variables in the framework have been measured in previous studies, future research could test (parts of) the framework through a written questionnaire. However, although the framework indicates that I expect perceived evaluative style to be related to fairness of evaluation and to trust in superior, these expectations cannot be made specific until the precise concept and measure of evaluative style in a particular organization is known. Therefore, future research will need to complement written questionnaires with interview data, to gain a deeper understanding of the exact concept of evaluative style and its relationship to subordinate managers' behaviour against the background of the particular organizational context.

5 In-depth field study: research design

5.1 Introduction

While the chapters 2, 3 and 4 have been mainly theoretical, the remainder of this study will describe the empirical part of this study: an in-depth field study conducted within a single organisation. The findings and conclusions from the literature review in the previous chapters were used to design the research in such a way that it would address some of the shortcomings and gaps of earlier research on evaluative style. In this chapter I will explicate the research design, while chapter 6, 7 and 8 will describe the findings from the study. This chapter is organised as follows. First, I will describe the purposes and topics that I will explore in this field study (§ 5.2). Next, in § 5.3, I will consider the methods by which data have been collected in more detail. After a discussion of the measurement of variables that I used in the written questionnaire (§ 5.4), I will finally spend a few words on the methods by which the data from both the interviews and from the questionnaire will be analysed. The chapter concludes with an overview of the data analyses that will be presented in subsequent chapters.

5.2 Purpose and method of the field study

In general, the field study serves a twofold purpose. The first purpose is to gain a better understanding of the construct of perceived evaluative style, taking into account shortcomings of, criticisms on and concerns raised in the existing RAPM-literature. The second aim of this field study is to improve our understanding of the relationship between perceived evaluative style and subordinates' behaviour.

As to the first purpose, gaining a better understanding of the construct of evaluative style, based on the preceding review of the literature four topics in particular have to be explored:

1. The conceptualisation and valid measurement of underlying dimensions of perceived evaluative style that are meaningful in the particular organisation (see chapter 3);
2. The level of agreement in perceived evaluative styles within and between groups of subordinates who are evaluated by the same superior (see § 4.2.3);
3. The relationship between contextual factors at different levels of analysis and perceived evaluative style (see figure 4.1);
4. The relationship between perceived leadership style and perceived evaluative style (§ 4.3).

For the second purpose, I will first explore the relationships that have been outlined in the framework displayed in figure 4.3, and then try to interpret and explain the findings by giving explicit reference to the particular organisational context, and the particular design of the evaluative system.

These two purposes and the different topics that I will address are listed in Table 5.1. For each of the topics, propositions and research questions that I have identified in earlier chapters are included.

5.3 Data collection

5.3.1 Sample

Although initially, as a follow-up to my findings reported in chapter 3, I hoped to conduct this in-depth case study within Service Group, due to a reorganisation process at Service Group they told me: "it would not be ideal if you would be walking around at this time". Therefore, I had to find another organisation. An important criterion for the organisation in which I wanted to do the field study on evaluative style was that the research site should have similar units, and preferably be a branch-type organisation like Service Group. The limitation to a single organisation and the similarity of units were deemed necessary to make sure that if differences in evaluative style were found, these could not be attributed to differences in the design of the performance evaluation system, but would be traceable to differences in the use of the system.

Based on this criterion, I was able to get access to a Dutch organisation, Van den Bergh Netherlands (VDBN). In total, twelve superiors and their subordinates were asked to participate in this research project. Seven superiors were Works Managers, i.e. heads of a production site; the other five superiors were Marketing Managers, i.e. heads of a marketing department. These twelve superiors and their units were selected by the company's HR manager, based on my request that I wanted around ten different units showing as much similarity as possible in terms of hierarchical level, but at the same time with expected differences in evaluative style. The HR manager wrote a letter to the superiors, explaining the importance of participation, asking their full co-operation with this research project and asking them to inform their subordinates of this project too.

Because of the explorative nature of this study, and because of the importance of studying perceived evaluative style in the organisational context, I collected data from different sources combining different methods: interview data from all twelve superiors and 27 subordinate managers, a written questionnaire that was sent to all subordinates reporting to one of the twelve superiors, and several internal documents from the organisation.

Table 5.1 Overview of propositions and research questions to be explored empirically, arranged by purpose and topic.

1 Purpose: gain a better understanding of the construct of evaluative style	
1.1	<p>Conceptualisation and measurement of perceived evaluative style</p> <ul style="list-style-type: none"> What are meaningful underlying dimensions of perceived evaluative style in this particular organisation? What is the role of (financial) accounting information in this particular organisation?
1.2	<p>Between- and within-group agreement in evaluative style</p> <p>P1. There are significant between-group differences in perceived evaluative style (where groups are subordinates reporting to the same superior).</p> <p>P2. There is considerable agreement on perceived evaluative style within groups of subordinates reporting to the same superior (<i>null hypothesis, expected to be rejected</i>).</p>
1.3	<p>Relationship between contextual factors and evaluative style</p> <ul style="list-style-type: none"> How does the organisational context explain the underlying dimensions of perceived evaluative style that are found to be relevant within this organisation? How does the organisational context explain the role of (financial) accounting information?
	<p>At the group level (leader level):</p> <p>P3. As task uncertainty increases, evaluative styles will be characterised by less emphasis on budgets and other quantitative targets, a higher emphasis on qualitative, interpersonal aspects of performance, and a higher willingness of superiors to listen to subordinates' explanations of their performance (a more flexible use of budgets and other quantitative targets).</p>
	<p>At the level of individual subordinates:</p> <p>P4. Within-group differences in age, tenure, and education of subordinates are related to within-group differences in perceived evaluative style.</p> <p>P5. Within-group differences in perceived evaluative style partly reflect differences in evaluative style that the superior intended.</p>
1.4	<p>Relationship between (perceived) leadership style and (perceived) evaluative style</p> <p>P6. Inspirational leadership will be more positively associated with an emphasis on interpersonal, qualitative aspects of performance evaluation than transactional leadership.</p> <p>P7. Transactional leadership will be more positively associated with an emphasis on quantitative measures of performance and deviations from targets than inspirational leadership.</p>

Table 5.1 (continued)

2 Purpose: improve understanding of the relationship between evaluative style and subordinate managers' behaviour.	
2.1 Relationship between perceived evaluative style and subordinate managers' behaviour	
P8.	The relationship between perceived evaluative style and managers' behaviour will be explained by an indirect effect whereby situational appropriateness of the perceived evaluative style affects fairness of evaluation, which in turn is positively related to managers' behaviour.
P9.	If subordinates agree with the criteria that are used to evaluate their performance, feelings of unfairness will arise if superiors ignore these criteria or do not attach as much importance to these criteria as they should according to the subordinate. On the other hand, if subordinates disagree with the criteria used to evaluate their performance, feelings of unfairness will increase as superiors attach more importance to these criteria and ignore the objections of subordinates to these criteria.
P10.	The level of agreement with evaluative criteria reported by subordinates can be explained by giving explicit consideration to characteristics of the criteria that are used to evaluate performance in combination with the level of task uncertainty.
P10a.	As task uncertainty increases, the felt appropriateness of accounting performance measures will decrease.
P10b.	Evaluative styles that are characterised by an emphasis on quantitative measures of performance and deviations from targets will result in lower levels of perceived fairness of evaluation in high task uncertainty situations than in low task uncertainty situations.
P10c.	Evaluative styles that are characterised by an emphasis on qualitative, interpersonal measures of performance will result in the same level of perceived fairness of evaluation in high task uncertainty situations as in low task uncertainty situations.
P11.	The relationship between perceived evaluative style and subordinate manager's behaviour will be explained by an indirect effect whereby perceived evaluative style affects trust in superior, which in turn is positively related to subordinate manager's behaviour.
P12.	The relationship between perceived evaluative style and trust in superior will be explained by an indirect effect whereby perceived evaluative style affects perceived fairness of evaluation, which in turn is positively related to trust in superior.
2.2 Explanation of findings by giving account to the organisational context in general and the design of the evaluative system in particular	

Table 5.2 provides an overview of the total sample. For each of the twelve leaders, table 5.2 lists the function (second column), the number of subordinates interviewed (third column), the total number of subordinates (fourth column), and the number of subordinates who filled out the complete questionnaire (fifth column). The type of data that was collected through each method will be explained below.

Table 5.2 *Total sample: number of subordinates per leader*

Leader (n = 12)	Function*	Number of subordinates interviewed (n = 27)	Questionnaire	
			Sample (n = 75)	Respons (n = 57)
1	M	2	2	2
2	M	2	4	4
3	M	3	3	3
4	M	2	3	2
5	M	2	2	2
6	P	3	6	5
7	P	3	7	4
8	P	3	8	5
9	P	2	8	6
10	P	1	16	10
11	P	2	10	10
12	P	2	6	4
* M = marketing/sales; P = production				

5.3.2 Interview data

In each unit, I attempted to interview the superior and two or three subordinates reporting to the superior, who were working in a management function. This criterion was applied to warrant the comparability of responses. This criterion was important since at the production sites not all subordinates working under the Works Manager held a management function: production heads also reported to the Works Manager, but they were not considered managers. Therefore, at the production sites only subordinates in a staff function were interviewed. In total, twelve superiors and 27 subordinates were interviewed (see Table 5.2).

The interviews with subordinates were used to collect data on the following ten topics:

- 1 The kind of performance measures available;
- 2 The use of these measures in performance evaluation;
- 3 The role of (financial) accounting information;
- 4 The role of perceived fairness of performance evaluation;
- 5 The consistency of evaluative style across subordinates, in time, and across systems
- 6 Examples of dysfunctional effects

- 7 Examples of functional effects
- 8 Subordinate's opinion on the performance evaluation system
- 9 Job description
- 10 General background of subordinate (age, education, etc.)

These interviews served to gain a more complete picture of (1) the underlying dimensions of evaluative style that are relevant and meaningful in this particular organisation, (2) factors that may be of importance in explaining the relationship between evaluative style and subordinates' behaviour, in line with or in addition to the expected relationships outlined in figure 4.4, (3) the "nature" of perceived evaluative style, and (4) the organisational context in which the evaluation of performance takes place, with a particular interest in the design of the evaluative system.

The interviews with superiors covered many of the same topics as the interviews with subordinates. The main purpose of interviewing superiors in addition to subordinates was to improve our understanding of the "nature" of perceived evaluative style, by exploring whether superiors intended to evaluate different subordinates in similar or in different ways, and for what reasons. Furthermore, by including interview data collected from superiors, I hoped to find some evidence for the possible influence of the hierarchical level in the organisation at which perceived evaluative style is measured on the relevant underlying dimensions of perceived evaluative style (see figure 4.1 and table 4.2).

The interviews were semi-structured, using an interview protocol that contained several sections intended to cover each of the ten topics. Although similar in many ways, different protocols were used for the interviews with the superiors and the interviews with subordinates. Both protocols are described in Appendix C. All interviews started with a short description of the research project and a short introduction of myself. After this introduction, I would ask the interviewee to introduce himself or herself. The first questions (1 through 3) in the interview protocol are mainly included for this purpose, although some insight into the background of the interviews could help explain some of the findings from the interviews. Subsequently, I asked the interviewee to describe his or her own job, and the job of their superior (question 4 through 6). These questions were intended to provide information on the subordinates' perception of what is important in his or her job in terms of responsibilities, targets and means. Questions 7 and 8 focus on the evaluation criteria that are used in the organisation to evaluate performance, and the role of PDP in this process, whereas questions 9 through 14 were intended to cover the role of superiors in the process of performance evaluation. Since I expected that (financial) accounting information would not play an important role in the performance evaluation cycle, questions 15 and 16 explicitly addressed the role of accounting information. It is important to note that the precise wording of these questions depended on the answers on earlier questions. After talking about the

evaluation criteria and how the superior uses these criteria, interviewees were asked whether they agreed with these criteria and how they are used (17 through 20). Questions 21 through 23 were intended to shed some light on the "nature" of evaluative style. I tried to assess whether and why evaluation criteria would change with a change in evaluation system, in time, and between different subordinates reporting to the same superior. Finally, the interview protocol contained some questions about the incidence of functional or dysfunctional behaviours, possibly in relation to PDP (24 through 28).

The interviews were held from November 1998 till March 1999. All interviews were tape-recorded, since none of the interviewees objected to that.

5.3.3 Written questionnaire

In addition to the interview data, a written questionnaire was used which enabled me to include subordinates in my research project that, for reasons of time constraints, could not be interviewed. Furthermore, by using a written questionnaire in combination with interview data, triangulation of findings is enabled, thereby increasing the validity of these findings. Finally, the written questionnaire served to connect the findings to previous literature. Therefore, the largest part of the questionnaire measures concepts that have been used previously in accounting research. Based on the constructs identified in figure 4.3, and, to a lesser extent, in figure 4.1, the questionnaire was developed to measure the following constructs: evaluative style, leadership style, task uncertainty, trust in superior, fairness of evaluation, job related tension, job satisfaction, and age, tenure, and education. The choice of variables to be included in the questionnaire to capture these constructs will be described in § 5.4.

The written questionnaire was left with the respondent directly after the interview with the request to fill it out sometime during the next week and send it back. After interviewing subordinates that reported to one superior, a questionnaire was sent to all other subordinates reporting to the same superior, being a manager or not. In total, 75 questionnaires were sent, and 61 questionnaires returned. Four questionnaires were omitted from the sample due to incomplete answers on the key variables, i.e. perceived and preferred evaluative style. The criterion was that for each of these two variables containing 18 items at least 14 items should have been filled out. If less than 14 items were filled out, the questionnaire was omitted from the sample. This applied to four questionnaires, which all contained no answers to all 18 items. Of the remaining 57 questionnaires, 6 lacked an answer on the item "performance delivered in the past few years (in de afgelopen jaren geleverde prestaties)". Further analysis of these six questionnaires revealed that those six respondents had either been working in their present function less than a year or they had been working under their current supervisor for less than a year, which made this item inapplicable. This however applied to 21 other individuals who did answer this question. But because it is doubtful whether their answers have real meaning, I decided to drop this item

in the further analysis. Of the 57 questionnaires in this study, 13 questionnaires came from marketing managers, 19 from managers at a production site, and 25 from production heads.

With respect to the design and administration of the questionnaire the Total Design Method as described by Dillman (1978; 2000) proved to be very helpful. Following the procedures suggested by Dillman I was able to get an effective response rate as high of 76%⁴². In particular, careful consideration was given to the design and lay-out of the questionnaire's frontpage, such that it would attract respondents' attention, and would invite them to fill the questionnaire out. In addition, a cover letter was attached to the questionnaire that shortly introduced the respondents to the aim and topic of the research project, and explained (and stressed) the importance of returning the questionnaire. Furthermore, my business card was attached to the questionnaire as well, so that people could contact me if for any reason they felt the need to do so. The questionnaire itself was divided in five different sections, each section starting with a short explanation about the relevance of this section for the study, and what the items in each section were about.

Attached to the back cover, on the inside of the questionnaire, a return envelope with the university's logo on which the return address at the university had already been printed and that did not need a stamp was provided. Two weeks after the questionnaire had been distributed, I would contact the respondents either by mail, e-mail, or telephone, to remind them to return the questionnaire. If they did not respond to this reminder, about one or two weeks later, I would send them a new questionnaire, together with a letter which once again stressed the importance of filling out and returning the questionnaire.

Since the questionnaire was administered in Dutch, the variables taken from English scales were translated into Dutch by a colleague and myself, after which we discussed the differences. This procedure led to a few minor adjustments in the translation of some items. Subsequently, all translated items were translated back into English by another person, who unfortunately was not a native English speaker, but could speak and write English fluently. This translation was compared with the original items, which revealed no significant deviations. As a final check, of course, my supervisors, one of which is a native English speaker who also speaks and writes Dutch fluently, checked the items both in Dutch and English as well.

Finally, I asked the Human Resource Manager to fill out the questionnaire and to inform me of possible difficulties or inappropriate items for VDBN. This too has led to some minor adjustments to the questionnaire items.

⁴² The support from the HR department probably played an important role as well, as the fact that a new performance evaluation system had been implemented quite recently.

5.3.4 Internal documents

Several documents from the organisation were consulted in order to improve our understanding of the organisational context in which performance evaluation took place, and, again, to triangulate results. These documents ranged from organisational charts and balanced scorecards to booklets and presentations on the performance evaluation system. Obviously, the most important contextual factor on which information was obtained from such documents, in addition to the interview data, was the design of the performance evaluation system that was in use throughout the organisation. However, information on other contextual factors, especially factors at the organisational level, was gathered as well. To triangulate results, I had access to some performance evaluation reports from the organisation's personnel archive.

5.4 Variable measurement in the written questionnaire

5.4.1 Introduction

Several criteria helped to choose which variables to include in the written questionnaire. The most important criteria, obviously, was that the variable should be theoretically relevant to the topic studied. Secondly, if possible, the variable should have been used in prior research and should have a reported Cronbach α higher than .70 (Nunnally, 1978). Thirdly, because the questionnaire was administered in Dutch, if the first two criteria had been met, variables that had been used in Dutch questionnaires should be chosen, in order to overcome problems inherent in using translated items. Finally, the total number of items and total number of pages for the questionnaire as a whole should be limited.

In this section, I will describe the characteristics and reliability of each of the variables that were measured through the written questionnaire. Before I turn to the description of each variable, a few general remarks will be made. Some of the instruments contained reversed items. The scores of these items were recoded. Subsequently, the internal reliability and unidimensionality (homogeneity) of all the scales were assessed. Both alpha-coefficients and factor analyses were performed for this purpose. All final scales as they are used in subsequent analyses in later chapters, including Cronbach's α , inter-item correlations and item-total correlations (average and range) are reported in Appendix D.

5.4.2 The measurement of perceived evaluative style

Perceived evaluative style. As I have argued in chapter 2 and chapter 3, existing scales to measure perceived evaluative style from the RAPM literature should not be used, because of the many problems attached to these scales. Therefore, to measure subordinates' perception of his or her leader's evaluative style, I developed my own list of 17 theoretically

relevant aspects of performance evaluation that could possibly help me to explore differences in perceived evaluative style. Respondents were asked to answer the following question on a 5-point Likert-scale with 1= no, 2 = low, 3 = some, 4= high, 5 = very high:

"When evaluating your performance as a manager (actual situation), how much emphasis does your superior lay on (Als uw baas uw prestaties als manager beoordeelt (huidige situatie), hoeveel nadruk leg hij/zij dan op):

1. Short-term goals (korte termijn doelen).
2. Long-term goals (lange termijn doelen).
3. Financial information (financiële informatie).
4. Non-financial information (niet-financiële informatie).
5. Deviations from agreed performance (afwijkingen van afgesproken prestaties).
6. Your explanation of your performance (uw uitleg van uw prestaties)
7. Objectively measurable performance (objectief meetbare prestaties)
8. His/her own intuition (zijn/haar eigen intuïtie)
9. Performance delivered in the past year (in het afgelopen jaar geleverde prestaties).
10. Your plans for the next year (uw plannen voor het komende jaar).
11. Positive aspects of your performance (positieve aspecten van uw presteren).
12. Negative aspects of your performance (negatieve aspecten van uw presteren).
13. Performances of the unit that you are responsible for ((prestaties van de eenheid waarvoor u verantwoordelijk bent.)
14. Personal, individual performance (persoonlijke, individuele prestaties).
15. Causes of failings in performance (oorzaken voor het achterblijven van prestaties).
16. Possibilities to improve/develop your performance (mogelijkheden tot verbetering/ontwikkeling van uw prestaties).
17. Information he/she gets from the conversation with you (informatie die hij/zij ontvangt uit het gesprek met u.)

These 17 aspects were chosen to measure aspects of performance evaluation that could theoretically explain differences in managerial attitudes and behaviour, and capture underlying dimensions of evaluative style. For this reason, the first 14 items were partly based on the five general dimensions identified by Otley & Fakiolas (2000). In their review of existing measures of evaluative style they distinguished the following dimensions:

- (a) Hopwood's rigid use of budgets vs a more flexible use.
- (b) Hopwood's short-run vs long-run emphasis (which may be a sub-dimension of (a) above).
- (c) As above, but referring to any quantitative targets, not just financial budgets.
- (d) An emphasis on absolute (quantitative) measures of performance rather than comparisons with pre-set targets.
- (e) The relative importance of objective and subjective criteria of evaluation.

It is important to stress here that although some of the items were based on these five dimensions, I did not expect these items to directly measure a specific underlying dimension. One of the problems with these five dimensions is that it is not clear how these dimensions relate to other dimensions. Furthermore, each single dimension as identified by Otley & Fakiolas (2000) is described in terms of opposite aspects. For example, the first dimension is

described as a relative emphasis on financial rather than non-financial information, but theoretically and statistically financial and non-financial information are not necessarily negatively (cor)related to one another.

Discussions with HR managers helped to get the terminology of the items correct, and led to the addition of the last three items. These items refer to specific characteristics of performance evaluation within this particular organisation that were not captured by the dimensions suggested by Otley & Fakiolas (2000).

Since the conceptualisation and measurement of perceived evaluative style is one of the main topics in this study, whether these 17 items proved to be helpful in identifying and measuring meaningful and relevant differences in perceived evaluative style within VDBN, and in what way, will not be addressed here. The further analysis of these 17 items, including an analysis of possible underlying dimensions, will be part of the data analysis in chapter 7 (see § 7.4), in which I will explore the concept and measurement of perceived evaluative style in this study in much more detail.

Interpersonal evaluation. In addition to the 17 items developed for this study, a second measure of perceived evaluative style was used in the written questionnaire. This was a five-item instrument which Farh et al. (1997) developed, by adapting a number of items from Folger & Konovsky (1989). The five items are:

1. My supervisor is thoroughly familiar with my job performance
2. My supervisor allows me to tell my side of the story in performance evaluation
3. My supervisor lets me know my appraisal outcomes and provides justification
4. My supervisor lets me know my pay raise and annual bonuses and provides justification
5. My supervisor reviews my performance with me and discusses plans or objectives to improve my performance

This instrument was originally intended to capture interactional justice by focusing on the “interpersonal behavior of the supervisor in performance appraisal and pay decisions” (Farh et al., 1997, p. 432). They argue that interpersonal behavior of supervisors has been shown in several studies to be an important influence on the perceived fairness of formal procedures within organisations. Although Farh et al. (1997) used these items to measure interactional justice, the items refer to performance appraisal process characteristics that have been used in different, but similar forms in the human resource and organisational behaviour literature, such as the superior’s knowledge of subordinates’ performance (Fulk et al., 1985; Landy et al., 1978; 1980), the development of action plans related to performance weaknesses (Fulk et al., 1985; Landy et al. 1978; 1980), and the extent to which the superior allows subordinates to present their side (i.e. have a “voice”) (Blader & Tyler, 2003). Since it is clear that the interpersonal behavior of superiors as measured by this instrument refers to the manner in which superiors evaluate the performance of subordinates, the level of interpersonal behavior when evaluating subordinates would be an important aspect of what I have called perceived evaluative style throughout this study. Therefore, this five-item

instrument was used to capture this aspect of perceived evaluative style. By using an existing instrument in addition to the instrument developed in this study, I hoped to be able to cross-validate the conceptualisation and measurement of perceived evaluative style, and further data analyses involving perceived evaluative style.

Cronbach α was .73 which is above the .70 criterion. Furthermore, item-total correlations were all above .30 and all five items loaded on a single factor. One of the items had an average inter-item correlation of .29, which is slightly below the .30 criterion. Taking into account that Farh et al. (1997) reported a much higher Cronbach alpha of .88 for the five-item instrument than in this study, and the fact that the correlation between the four-item instrument and the five-item instrument was very high (i.e., Kendall's tau = .933, $p = .000$), I decided not to drop this item. Therefore, the summed score on five items was used as a measure of Interpersonal Evaluation.

Preferred evaluative style. Because I also expected beforehand that agreement with evaluative style or with evaluation criteria could be an important explanatory variable with respect to gaining an understanding of the behavioural outcomes associated with different styles, I decided to measure preferred evaluative style as well. Respondents were asked the following question: "When evaluating your performance, how much emphasis should your boss in your opinion in the ideal situation lay on:" (Hoeveel nadruk zou uw baas bij het beoordelen van uw prestaties naar uw mening in de ideale situatie moeten leggen op:), followed by the same 17 items as described above. By measuring both perceived and preferred evaluative style, I hoped to be able to explore differences in agreement with evaluative style. This too will be analysed and discussed in more depth in chapter 7.

5.4.3 The measurement of leadership style

To measure leadership style, four subscales (21 items) from a questionnaire developed by Den Hartog (1997) to measure different types of leadership were used. Den Hartog's original questionnaire contained 92 items to measure thirteen dimensions of leadership. To limit the total length of my questionnaire, I used four of the thirteen dimensions of leadership that were measured by Den Hartog. These four dimensions were active management-by-exception, demonstrating trust and confidence in subordinates, individual consideration, and team-building. Of these dimensions, active management-by-exception measures an aspect of transactional leadership, while the other three subscales all measure aspects of inspirational leadership. Sample items from each scale can be found in Box 5.1. These subscales were selected because I expected them to measure aspects of leadership that were seen as important in the particular organisation (team-building), or were most likely to be related to perceived evaluative style.

Box 5.1. Sample items from the ILO-subscales used in this study

- Active management-by-exception
'points it out to me when my work is not up to par'
- Demonstrating confidence and trust in subordinates
'shows confidence in my ability to contribute to the goals of this unit'
- Individualized consideration
'is genuinely concerned about the growth and development of subordinates'
- Team-building
'develops teamspirit among employees'

To validate the measurement and internal consistency of the four subscales, first, using the same procedure as Den Hartog (1997), the missing values, if less than 25% in a scale, were replaced with the rounded mean of the respondent's score on the other items in the a-priori scale. Next, exploratory factor analysis was done to assess whether the same factor structure as described by Den Hartog (1997) would be found. Principal component factor analyses were done, and the factor structure found after varimax rotation was interpreted. Although five factors had an eigenvalue larger than one, both the three and the four factor solutions were well interpretable. In this study, I interpreted the three-factor-solution for further use in my analyses, because differences between this solution and the four a-priori subscales could theoretically be explained. The three empirical factors found are reported in Table 5.3. Two items of which the difference between factor loadings on two factors was less than .20 were discarded. Factor 1 is identical to the a-priori scale for team-building. Factor 2 is almost identical to the a-priori scale for active management-by-exception. The only difference is that the item "expects a lot from us" is dropped. This item also had the lowest item-rest correlation in Den Hartog's study and was not part of the a-priori scale for active management-by-exception in Den Hartog's study. Den Hartog placed it in this scale because it correlated higher with this scale than with the a-priori scale it was in originally. Finally, the third factor found in my data contains items from the other two a-priori scales, i.e. demonstrating trust and confidence in subordinates and individualised consideration. These two factors were highly correlated in Den Hartog's study ($r=.76$, $p<.01$) and loaded on a single factor, combined with the participation subscale that was not measured in my study, in a second-order factor analysis of the inspirational scales (see Den Hartog, p.126). Collapsing those two subscales into a single scale seems therefore justified.

Finally, Cronbach's α , average inter-item correlations and corrected item-rest correlations were calculated to test the internal consistency of the three empirical scales. The criteria were that Cronbach's α should be $>.70$, and item-rest correlations and average inter-item correlation should be $>.30$.

Table 5.3. Leadership Subscales: Results from Factor Analysis (n = 57)

Factor 1: Team-building (team)	factor loading	item-rest correlation
Encourages employees to be 'team players'	.90	.85
Develops team-spirit among employees	.87	.83
Gets the group to work together for the same goal	.86	.81
Works at creating a climate of trust among members of the management team	.85	.82
Breaks down barriers to communication between work groups	.75	.73
5 items Cronbach alpha = .93 average inter-item correlation .72 mean = 3.67 sd = .90		
Factor 2: active management-by-exception (ambe)	factor loading	item-rest correlation
Points it out to me when my work is not up to par	.83	.76
Monitors performance for errors needing correction	.82	.74
Shows his or her displeasure when my work is below acceptable standards	.76	.61
Focuses attention on irregularities, exceptions and deviations from what is expected of me	.72	.62
Would indicate disapproval if I performed at a low level	.71	.65
Focuses attention on errors I make	.69	.60
Keeps careful track of mistakes	.65	.54
7 items , 1 item dropped; Cronbach alpha = .87 average inter-item correlation .49 mean = 3.23 sd = .68		
Factor 3: supportive (support)	factor loading	item-rest correlation
Treats me as an individual rather than just a member of the group	.73	.50
Listens to my concerns	.70	.67
Allows me a strong hand in setting my own performance goals	.66	.39
Shows confidence in my ability to contribute to the goals of this unit < organisation >	.65	.57
Demonstrates total confidence in me	.64	.58
Is genuinely concerned about the growth and the development of subordinates	.54	.54
Looks out for my personal welfare	.45	.40
7 items Cronbach alpha = .79 average inter-item correlation .35 mean = 4.00 sd = .55		

5.4.4 The measurement of contextual variables

Task Uncertainty. To measure task uncertainty, three scales have been used in the accounting literature. Some studies (Hirst, 1983; Ross, 1995) use the instrument developed by Duncan (1972) to measure task uncertainty. This instrument contains 12 items, and is meant to measure environmental uncertainty. The scale distinguishes two dimensions within environmental uncertainty: environmental complexity and environmental dynamism. Other studies (Hirst, 1983; Brownell & Dunk, 1991; Lau et al., 1995) use the 14 item, 7-point Likert-type scale developed by Van de Ven & Delbecq (1974). This scale measures two dimensions, each consisting of 7 items: task difficulty and task variability. However, while the Cronbach alpha for the task variability seems acceptable (0.76 in the Brownell & Dunk (1991) study), the reliability of the task difficulty dimension is questionable. Cronbach alpha reported by Brownell & Dunk (1991) and Lau et al. (1995) respectively are 0.57 and 0.41. A third, more recent instrument, used in the literature to measure task uncertainty is a nine-item, 7-point Likert-type scale developed by Whitey et al. (1983). Cronbach alphas reported in the literature for this scale are satisfactory: Brownell & Hirst (1986) report a Cronbach alpha of 0.82, and Lau et al. of 0.85.

The most reliable instrument therefore seems to be the Whitey et al. (1983) instrument. Since this is also the most recent instrument, this scale was used in the questionnaire to measure task uncertainty. Since Hartmann (1997) used the Whitey et al. instrument as well, reporting a Cronbach α of .87, I used his translation in my questionnaire.

Cronbach's α for the total scale consisting of 9 items was .81 which indicates that this scale has a high internal reliability. However, Hartmann (1997) reported a higher Cronbach's α , i.e. 0.87. A further analysis of the scale revealed that one item had an item-total correlation well below the .30 criterion. The average inter-item correlations for this item was 0.091, which is far below the .30 criterion too. All other items in the a priori TU scale had average inter-item correlations higher than .30. Therefore, I concluded that only one item was problematic in my study. This was not the case in Hartmann's (1997) study. He reported inter-item correlations ranging from .2619 to .7929 and item-total correlations ranging from .4859 to .6860.

Subsequently, factor analysis was carried out on the 9 items. This revealed multidimensionality, because two factors were extracted explaining 58,43% of variance. A closer look at the rotated factor solution indicates however that all items have positive loadings higher than .40 on the first component, except for the one "problematic" item discussed above, which has a small, negative loading on the first component. This item loaded highly positive on the second component, as did one other item. However, this other item had a relatively high loading on the first component too. This is another indication that one item did not fit well in the TU scale. This was confirmed by doing factor analysis on the other 8 items: this revealed unidimensionality, since just a single component was extracted.

Therefore, I decided to drop one item from the TU-scale and calculate TU as the sum of the remaining 8 items. Cronbach's α for the adapted eight-item scale was .84.

Individual differences. Three variables were included in the questionnaire to measure what I thought could be theoretically important aspects related to the person(ality) of the subordinate: tenure, age and education.

Tenure. Tenure was measured by asking three questions, each capturing a different aspect of tenure. I asked respondents how long he or she had been working (1) for the organisation, (2) in his current job, and (3) under supervision of his current leader (in his current job). A fourth question did not directly measure tenure, but was closely related to the other three questions, i.e. how many times the respondent had been formally evaluated by his or her superior.

Age. Age was measured by one open-ended question, i.e. "what is your age?"

Education. Education was measured by asking the subordinate the following closed-ended question: "Wat is de hoogste opleiding die u heeft voltooid?" Respondents could choose an answer from three categories: 1. HBO, 2. University, or "3. different, i.e. ...". It should be noted that this question is culture bound, as educational institutions vary per country. Therefore, there is no direct equivalent in English. Although theoretically there would be many different categories to choose from, I expected these two categories would be appropriate for almost all respondents, since they were all reporting to a Work Level 3 manager, although not everyone held a managerial function.

5.4.5 The measurement of fairness of evaluation

Perceived fairness of evaluation. As explained in chapter 4, two dimensions should be distinguished when assessing the perceived fairness of evaluation: first, the level of agreement with the evaluation criteria, and, secondly, the level of agreement with how the superior uses these measures. To measure perceived fairness of evaluation I used an instrument developed by Hartmann (1997). Although Hartmann (1997) refers to this instrument as a measure of "agreement with evaluative criteria", a close look at the items in this instrument reveals that it would be more appropriate to refer to this instrument as a measure of the perceived fairness of evaluation, since the instrument contains both items that relate to the level of agreement with the evaluative criteria as to the level of agreement with how the superior evaluates the performance of subordinates. Thus, it seems to combine both components into a single instrument. Hartmann's (1997) 9-item instrument showed a high Cronbach alpha (i.e. .87) in his study, and the items in the scale seemed to be theoretically well grounded. Sample items from this scale are: "My performance evaluation is

based on a complete picture of my actual performance", and "In evaluating my performance my supervisor often emphasises aspects of my job which I think are not relevant".

Cronbach α for the nine-items scale in my study was .88, which is comparable to the Cronbach α reported by Hartmann (1997). All nine items had inter-item and item-total correlations higher than .30. Factor analysis revealed that, after rotation, seven items loaded on a single factor, and two items, both reverse-coded, on a second factor. Hartmann (1997) reported that in his study two factors were found as well, but in his study all three reverse-coded items loaded on the second factor. However, when extracting a single factor, all items have factor loadings $> .500$. Therefore, although there is some doubt on unidimensionality of the nine-item scale, I decided to use the sum of the scores on all nine items as the score for Perceived Fairness of Evaluation, in line with Hartmann (1997).

Felt Appropriateness of accounting performance measures. In addition to the measure of perceived fairness of evaluation, I included another variable that does not necessarily relate to the (perceived) fairness of evaluation, but that could be important for an understanding of the role of accounting information for evaluative purposes, i.e. felt appropriateness of accounting performance measures. To measure the felt appropriateness of accounting performance measures, I used a six-item instrument also developed by Hartmann (1997). Although the original instrument developed for Hartmann's study contained twelve items, factor analysis revealed that the twelve items loaded on three factors. Six items loaded on the first factor, and subsequent factor analysis of these six items showed unidimensionality, a satisfactory Cronbach α of .84, high item-total and inter-item correlations, and a high, significant correlation ($r = .91$; $p = .000$) with the twelve-item index. Therefore, Hartmann (1997) used the mean score of the six items when testing his hypotheses. Sample items to measure felt appropriateness of accounting performance measures are: "Whether my department functions well or not can be expressed accurately in financial figures." and "Budget variances provide a complete overview of the functioning and performance of my department."

In my study, Cronbach's α for the six-item scale was .72. However, two items had an item-total correlation that did not meet the .30 criterion (.27 and .28 respectively). Furthermore, the average inter-item correlations for these two items were below the .30 (.20 for both items). These analyses cast some doubt on the internal reliability of the six-item scale, although the unidimensionality of the six-item scale was supported by factor analysis. After dropping the two problematic items, subsequent analyses showed that for the four-item scale unidimensionality and internal reliability was achieved. Therefore, I decided to use the sum of four items as the total score for Felt Appropriateness of Accounting Performance Measures. This four-item scale had a Cronbach's α of .76.

5.4.6 The measurement of trust in superior

To measure trust I used a five-item instrument that has been used in previous accounting research (Hopwood, 1973; Ross, 1994). The questions in this instrument were originally intended to reflect the subordinate's trust or confidence in the superior's motives and intentions with respect to matters relevant to the subordinate's career and status in the organisation (Read, 1962). The respondents were asked to answer the questions on a 5-point Likert-type scale. Read (1962) used only four of the five items, and reported intercorrelations between these four items ranging from 0.39 to 0.68. A sample item from this instrument is "Does your superior take advantage of opportunities that come up to further your interests by his actions and decisions?".

In my study, all analyses to assess the internal reliability and unidimensionality for the five-item scale revealed no problems: Cronbach α was .78, average inter-item and item-total correlations were well above .30 for all items, and all five items loaded on a single factor. Therefore, the summed score on all five items was used as a measure of Trust.

5.4.7 Variables relating to behavioural outcomes

Typical outcomes used throughout the accounting literature are job-related tension, job satisfaction, and managerial performance. Researchers have been quite unanimous in their choice of instruments for measuring these constructs, despite the fact that several authors have pointed out that relating these measures to (organisational) performance is problematic. Especially the use of self-reported measures of managerial performance has been subject to critiques (e.g. from Briers & Hirst, 1990; Hartmann, 2000). In my questionnaire, I limited the measurement of outcomes to job-related tension and job satisfaction.

Job satisfaction. Several instruments exist that measure job satisfaction. In the accounting literature, three instruments have been used: an instrument developed by Kahn et al. (1964), the MSQ (1967) and the JDI (1969). The instrument developed by Kahn et al. (1964) is used by Hopwood (1973) and Swieringa and Moncur (1975). The MSQ is used by Brownell (Spring 1982), Brownell (1983), and Harrison (1992), while the JDI, or the short version thereof, is used by Kenis (1979) and Brownell (1987). Since all three instruments meet the first two criteria, I decided to use the instrument used by Hartmann's (1997) which is a slightly adapted version of Kahn's (1964) instrument. Cronbach's α for this instrument was .72 in Hartmann's study.

Cronbach's α for the 7-item Job Satisfaction scale was .61 which is below the .70 criterion and lower than the Cronbach's α coefficient reported by Hartmann (1997), i.e. 0.71. Of the 7 items in the a priori JSA scale, as many as four items had an item-total correlation below the .30 criterion. Furthermore, the average inter-item correlations for these four

items were well below the .30 criterion. The average inter-item correlations of the other three items were all lower than .30 too. These results indicate that the internal consistency of the JSA-scale is doubtful. Hartmann (1997) however reports inter-item correlations ranging from .1102 to .5438 and (almost) satisfactory item-total correlations ranging from .2663 to .5922. These numbers are much higher than in my study, and indicate a satisfactory internal consistency.

Subsequently, factor analysis was carried out on the 7 items. This revealed multidimensionality, because three factors were extracted explaining 66,94% of variance. The first component, which contained three items, seemed to represent "satisfaction with the organization", the second component, which also contained three items, "satisfaction with the job", and the third component is accounted for by one item. Considering the low item-total correlation of this item, it did not seem to fit well in the JSA scale. One other item was "problematic" as well, since it had high loadings on both the first and the second component, i.e. "I would rather have some other work". This however could be explained by the fact that this item is a general statement, which could refer to both the organisation and the job itself. Hartmann (1997) also reported multidimensionality: his factor analysis revealed two factors, which seemed to differentiate between satisfaction with the job and satisfaction with the organisation too.

We can conclude therefore that the internal reliability of the JSA scale is below the criteria, and problems occur with respect to both the internal consistency and the unidimensionality. The problem of multidimensionality was reported in an earlier study too, but regarding the internal reliability, no problems were reported in earlier studies using this scale. Further analysis of the first two factors that resulted from the factor analysis indicated that the three items in the first factor had high average item-total and inter-item correlations, while the three items in the second factor did not meet the criteria. Since my study is explorative, and not intended to replicate earlier findings, I decided to use the sum of three items from the first factor as the score for job satisfaction. Cronbach's α for this three-item scale was .67. However, the results concerning job satisfaction should be interpreted with some caution, because this three-item scale seemed to represent "satisfaction with the organisation" more than "satisfaction with the job" as explained above.

Job-related tension. In the RAPM literature, all authors that measure job-related tension use the instrument developed by Kahn et al. (1964). This is a 14-item, Likert-type scale, which has been shown a reliable instrument. Cronbach alphas reported in the literature vary from 0.78 (Hirst, 1983) to 0.89 (Brownell & Hirst, 1986). However, Hartmann (1997) used a different, shorter instrument, containing 8 items. Since Cronbach's α for his instrument was .81, my questionnaire was not used to replicate earlier studies, Hartmann's instrument was shorter than the instrument developed by Kahn et al. (1964) and no

translation problems would occur, again, I decided to use the eight-item instrument developed by Hartmann (1997).

Cronbach's α for the total scale consisting of 8 items was .74 which is satisfactory but below the Cronbach's α reported by Hartmann (1997).

Of the 8 items in the *a priori* JRT scale, two items had item-total correlations below the .30 criterion. Furthermore, the average inter-item correlations for these two items were .081 and .159 respectively, far below the .30 criterion. The average inter-item correlation of one other item was below .30 too, i.e. .207. These results cast some doubt on the internal consistency of the JRT-scale. Hartmann (1997) however reports inter-item correlations ranging from .1036 - .5514 and satisfactory item-total correlations ranging from .3536 to .6703. These numbers are higher than in my study, and indicate a satisfactory internal consistency.

Subsequently, factor analysis was carried out on the 8 items. This revealed multidimensionality, because two factors were extracted explaining 54,39% of variance. A third component had an eigenvalue slightly below 1.00; analysis revealed that one item loaded high on this third component. Hartmann (1997) reports that factor analysis in his study 'revealed no sign of multidimensionality, as only one factor was extracted, that explained 43.0 percent of variance'.

We can conclude therefore that in this study problems occur with respect to both the internal consistency and the unidimensionality of the JRT scale. These problems were not reported in Hartmann's earlier study using this scale. After dropping the three items that did not meet the criteria for average item-total and inter-item correlations, the remaining five items met the criteria, and loaded on a single factor. Cronbach's α for this five-item scale was .78. Furthermore, this adapted five-item scale was significantly and highly correlated with the overall JRT scale with eight items ($r=0.939$, $p=0.000$). Since my study is explorative, and not intended to replicate earlier findings, I decided to use the sum of the five items in the adapted scale as the score for JRT.

5.4.8 Descriptive statistics

Descriptive statistics of all scales that have been measured in the questionnaire, with the exception of perceived evaluative style, are summarised in Table 5.4. The conceptualisation and measurement of evaluative style will be explored in chapter 7.

Table 5.4 *Descriptive Statistics*

	N	Mean	Std. Deviation	Observed		Theoretical	
				Minimum	Maximum	Minimum	Maximum
Leadership style:							
AMBE	57	22.63	4.791	10	32	7	35
SUPPORT	57	27.98	3.875	19	35	7	35
TEAM	56	18.34	4.494	5	25	5	25
Evaluative style:							
PERSONAL	54	18.59	2.716	11	25	5	25
Context:							
TU	57	20.11	4.938	10	33	8	40
Fairness of evaluation:							
FAIREV	55	32.35	5.271	22	45	9	45
FAAPM	56	10.23	2.979	4	15	4	20
Trust:							
TRUST	57	18.60	3.035	10	25	5	25
Outcomes:							
JRT	57	13.82	3.670	6	23	5	25
JSA	57	11.86	1.894	8	15	3	15

5.5 Methods of data analysis

5.5.1 Method of analysis of interview data

Since the interviews were semi-structured, covering practically the same questions and topics in each interview, the interviews were analysed by using a method similar to the method that Miles & Huberman (1984) put forward. Based on the ten topics that had been specified before the interviews, each transcribed interview was divided into ten parts, each part covering a different topic. A separate document was prepared for each topic, bringing together parts from the different interviews that related to that specific topic. For example, one document is called "Available performance measures", which consists of all statements from different respondents covering the available performance measures within the organisation. Another document is called "Role of accounting information" which contains all sayings from the interviews with respect to the role of accounting information. All paragraphs in these documents were numbered for quick and easy reference purposes. Subsequently, as a first analysis, I read each of these ten documents, marking telling quotes, highlighting important remarks, identifying different performance measures, and writing down some notes or concepts in the left margin that could help me analyse the processes within the organisation. In the next step, for each document a matrix was prepared in a new

document, by drawing up the twelve leaders along the top of the page, and different questions from the semi-structured interviews and/or concepts that emerged in the preceding phase identified down the left margin. Finally, these tables were summarised.

5.5.2 Method of analysis of questionnaire data

To analyse the questionnaire data, first, I explored all variables that were measured in the questionnaire to assess the normality of the distribution of these variables. Although there were no large deviations from normality, given the relatively small sample size of 56, I decided to use nonparametric or distribution-free statistics rather than their parametric equivalents in most analyses⁴³. For example, correlations between two variables were tested by using Kendall's tau correlation, rather than Pearson correlations. Furthermore, when testing for significant differences in scores among groups of subordinates, a Kruskal-Wallis test was conducted rather than ANOVA. A notable exception will be the test for indirect effects in chapter 8, where I will use hierarchical regression analysis. In this case, before conducting the analysis, I will compare and report the (significance of the) zero-order Pearson correlations with the Kendall's tau correlations. The techniques to explore the data for statistically significant relationships or differences will be discussed in detail in each section of subsequent chapters, wherever appropriate.

5.6 Summary and conclusion

In this chapter I have described the research design of the field study that was conducted within Van den Bergh Nederland (VDBN). The study was designed in such a way that it would enable me to address the topics that emerged from the literature review in the preceding chapters. These topics are:

1. The conceptualisation and valid measurement of underlying dimensions of perceived evaluative style that are meaningful in the particular organisation;
2. The level of agreement in perceived evaluative styles within and between groups of subordinates who are evaluated by the same superior;
3. The relationship between contextual factors at different levels of analysis and perceived evaluative style;
4. The relationship between perceived leadership style and perceived evaluative style;

⁴³ The use of nonparametric tests has as a disadvantage that nonparametric tests are less powerful than parametric tests.

5. The relationship between perceived evaluative style and managerial behaviour;
6. The role of the organisational context, and the design of the evaluative system in particular.

Different sources of data have been collected to address these topics. In subsequent chapters I will present the data analysis and findings on each of these topics. Because of the importance of the organisational context in explaining both the conceptualisation of evaluative style and its behavioural implications in a particular organisation (topic 6), in chapter 6 I will describe this context based on interview data and data collected from organisational documents. This chapter will focus on contextual factors that operate at the level of the organisation as a whole. Next, in chapter 7, I will address the conceptualisation and measurement of perceived evaluative style (topic 1), with explicit reference to the level of agreement among subordinates reporting to the same superior (topic 2). For these analyses I will use interview and questionnaire data. Finally, in chapter 8, I will first report the findings on the relationship between contextual factors at the group and individual level and perceived evaluative style (topic 3), using data from the questionnaires and interviews. Subsequently, I will describe the relationship between perceived leadership style and perceived evaluative style (topic 4), using only questionnaire data. Lastly, the relationship between perceived evaluative style and managerial behaviour (topic 5) will be analysed and reported, combining questionnaire data with data from the interviews. Chapter 8 will be concluded by a discussion of the findings, in which I will give explicit account to the role of the organisational context (topic 6) in understanding these results.

6 Description of the organisational context

6.1 Introduction

Before exploring differences in evaluative style (chapter 7) and the relationship with managers' behaviour (chapter 8), in this chapter I will provide a more general description of the (organisational) context in which this study was conducted. First, I will give a general description of the organisation (§ 6.2). Next, I will describe the technical design of the performance evaluation system (§6.3). Against this background, I will describe some general observations from the interview data that could be important in understanding the outcomes of the analyses in subsequent chapters. Finally, I will give consideration to contextual differences at the level of function groups, leaders, and individuals.

6.2 Description of organisation (VDBN)

The study was carried out within Van den Bergh Netherlands (VDBN). With 1,800 employees and annual revenues of 1.6 billion guilders, VDBN is the largest food manufacturer in the Netherlands. VDBN is a subsidiary of Unilever. Unilever's Corporate Purpose as found in the Annual Report 1998 is shown in figure 6.1. Two elements from this corporate purpose are of special interest for a good understanding of the findings from the field study. First, Unilever's strategic objective is "sustainable, profitable growth for the business and long-term value creation for the shareholders and employees". The other elements in the corporate purpose are "the road" to this objective. Second, Unilever "requires a total commitment to exceptional standards of performance and productivity, to working together effectively and to a willingness to embrace new ideas and learn continuously". In line with the strategic objective, in 1997 a challenging, stretching target was given to all of Unilever's business units: a sustainable, profitable growth of five to ten percent per annum for the next five years. Unilever has taken several initiatives to help realise this corporate purpose. Firstly, to encourage personal growth Unilever has introduced "the Integrated Approach" which brings together target setting, personal development and career planning. The Integrated Approach will be described in detail in the next paragraph (§ 6.3), since this approach is essential for a good understanding of the way in which performance is evaluated within Unilever units. Secondly, to encourage the development of employee skills and responsibilities at the workfloor Total Productive Maintenance, a productivity programme originating in Japan, has been introduced in Unilever factories. Total Productive Maintenance (TPM) is a management strategy that aims at maximising production system efficiency by identifying and eliminating losses. According to Nakajima (1989, p.2) "the key innovation in TPM is that operators perform basic maintenance on their own equipment. They maintain their machines in good running order and develop the ability to detect potential problems before they generate breakdowns". The goal of TPM is zero breakdowns and zero defects, and achievement of this

goal will increase effectiveness in many other areas (cf. Nakajima, 1989). For example, based on the results of a number of Swedish empirical studies, Lycke (2003) argues that TPM has been associated with increased employee morale, higher productivity, better working environments, higher levels of safety and improved quality. Similarly, in a survey across Singaporean firms, Brah and Chong (2004) found evidence for significant positive correlations between TPM and financial performance (such as cost savings, higher operating profit, and increased sales), managerial performance (such as fewer customer complaints, lower absenteeism among workers, and increased worker's skills), and operational performance (such as reduced number of defective products, shorter set-up times, and decrease of minor stoppages). Despite the fact that VDBN operated in decreasing markets, the objective of sustainable, profitable growth applied to VDBN as well. Therefore, management at VDBN believed that existing paradigms and mindsets had to change. In the words of VDBN,

there was a need to reinvent and realign the six drivers of organisation performance, i.e. (1) vision, (2) strategic programs, (3) people, assets and capabilities, (4) organisational alignment, (5) managerial alignment, and (6) personal habit alignment. Behind this call for reinventing the business was the strategic thought that sustainable growth could not be achieved by operating in existing mature markets. VDBN had been primarily interested in operating efficiency and increasing volumes of existing brands in the past. Management believed operating innovatively and proactively and competing aggressively for market share should be the focus of the business in the future. To convince employees of the need for those strategic changes, and motivating and empowering them to help realise those changes,

Figure 6.1 *Unilever's Corporate Purpose*

Corporate Purpose of Unilever

Our purpose in Unilever is to meet the everyday needs of people everywhere - to anticipate the aspirations of our consumers and customers and to respond creatively and competitively with branded products and services which raise the quality of life.

Our deep roots in local cultures and markets around the world are our unparalleled inheritance and the foundation for our future growth. We will bring our wealth of knowledge and international expertise to the service of local consumers - a truly multi-local multinational.

Our long-term success requires a total commitment to exceptional standards of performance and productivity, to working together effectively and to a willingness to embrace new ideas and learn continuously.

We believe that to succeed requires the highest standards of corporate behaviour towards our employees, consumers and the societies and world in which we live.

This is Unilever's road to sustainable, profitable growth for our business and long-term value creation for our shareholders and employees.

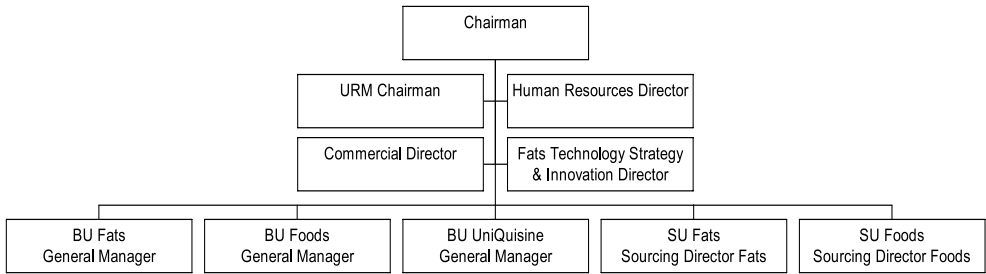
Source: Unilever Annual Report, 1998

several Human Resource activities were undertaken guided by the slogan "no company growth without personal growth". At the level of individuals, for each management level a training curriculum was written and executed to help managers understand the necessary changes, to help them identify how they could contribute to those changes and to provide them with the necessary skills and knowledge. The training therefore emphasised proactiveness, taking responsibility and leadership principles. At the level of teams, working in teams was heavily emphasised and stimulated, and several team building activities were undertaken. Furthermore, team leaders received training in leadership skills, and team members received training in how to be cooperative team members. At the organisational level, as described above, both the Integrated Approach and Total Productive Maintenance were introduced, and great effort was made to get a shared vision and shared beliefs throughout the organisation.

Organisation structure

The organisation structure of VDBN is depicted in figure 6.2.

Figure 6.2 Organisation structure of VDBN



The three marketing sales units (BU Fats, Foods, and UniQuisine) are responsible for selling and delivering products to customers and consumers in the Netherlands. The two sourcing units, Foods and Fats, produce products and sell these products to the three Dutch marketing sales units as well as to European (Unilever) marketing sales units and third parties. The sourcing unit Foods has three plants, producing many different products, ranging from peanut butter and ice-tea to dry and "wet" soups. One of these plants has three different, though related factories. The business unit Fats has two factories, one of which is a refinery producing edible oils and fats and the other mainly produces margarine.

The organisation structure of the BU Foods and the BU Fats had recently undergone some changes when I conducted the interviews in these business units. Instead of arranging the departments by functions, such as market-research, product-development, and marketing, in April 1998 the business unit Foods had been split in two departments. The business unit Fats followed in december 1998. The first department would concentrate on 'value creation'.

This department bore no profit responsibility and had no turnover goals, but was more long-term oriented, focusing on product and brand development. In this department people from marketing, market-research, and product development would work together on project basis in multidisciplinary brand teams. The other department would be more short-term oriented, focusing on 'value delivery'. In this department account managers and category managers worked together to manage the ongoing business. This structure had been chosen to enable the growth of existing core markets, increasing market shares and establishing new core markets.

6.3 Performance evaluation at VDBN: description of the Integrated Approach

6.3.1 Purpose of the Integrated Approach

One of the strategic programmes implemented at Unilever in 1998 (see Figure 6.3), and thus at VDBN, to help the organisation realise its strategic objective of sustainable, profitable growth, was the Integrated Approach, as discussed above (in § 6.2). This approach was implemented also because of shortcomings of existing systems. Firstly, job evaluation was based on many job classes, which often resulted in very short job tenures. Within a few years people would be promoted to the next job class(es). Secondly, the remuneration system ignored potential, and payment was not correlated with performance and responsibility as high as with age and/or tenure. Thirdly, existing management development activities were too general and missed clear objectives.

Figure 6.3 *The Integrated Approach as described in Unilever's Annual report 1998, p.24*

"We are Committed to the personal growth of all our employees. Individual progression both creates a more fulfilling workplace and drives the Company towards outstanding performance.

1998 saw the first full year of a major initiative designed to encourage personal growth. Called the Integrated Approach, it is a new way of thinking about individual career planning and professional development. At its heart are three aims:

- To give employees new opportunities to shape their progress in Unilever;
- To streamline work levels to create a leaner and more flexible organisation;
- To align reward and recognition processes so that we can nurture outstanding performance and maintain our competitive position in the remuneration market.

Initially conceived as a programme for managers, the Integrated Approach is a process we believe is important for all employees. It is therefore being developed for everyone."

According to an internal booklet, the purpose of the Integrated Approach is "to invest in personal and professional growth through development of individual skills and competencies, so as to continuously improve individual and organisational performance." The main elements of the Integrated Approach, aimed at continuous learning and adaptation, are:

- Organisation
- Reward & Recognition
- Performance Development Planning (PDP), which brings together:
 - ◆ Target Setting
 - ◆ Personal Development Plan
 - ◆ Performance Assessment

Each of these elements will be explained below.

6.3.2 Elements of the Integrated Approach

Organisation

To simplify the organisation, and reduce the number of management layers, the job classes are replaced with six Work Levels. The higher the Work Level, the broader are the nature of decisions to be taken and the more complex the environment in which those decisions have to be taken. Work Level 1 involves decisions that are largely operational in nature, Work Level 6 (Chairman) involves strategic decisions. This should result in a lean and flexible organisation, clear roles and accountability for every manager in the organisation, longer job tenure and more meaningful career moves than in the past. Each job is assigned to a specific Work Level. Young academics in their first job typically enter in either a Work Level 1 or Work Level 2 job.

Reward & Recognition

As to reward & recognition, VDBN argues that instead of annual increases in salary based on age or years within the organisation, guiding principles for reward policies should be to pay competitively for responsibility, performance, potential and results. Each job is linked to a particular Pay Scale; these pay scales are determined by a market survey comparing remuneration packages between the most important competitors in the Dutch market. Within each pay scale a distinction is made between the competitive zone (80 - 110% of competitive salary) in which 75% of the managers is rewarded and a premium zone (110-125% of competitive salary) for 25% of the managers who are either on the High Potential or on the Sustained High Performance List. Managers on the High Potential List are expected to fill positions at the next Work Level within five years, while managers on the Sustained High Performance List have acknowledged skills and expertise, irrespective of Potential. Furthermore, part of the remuneration of managers comes from Variable Pay, the level of which depends on Business Results and Personal Targets. Variable Pay has a maximum of 15% of salary for Work Level 3 managers, and 10% for Work Level 2 managers.

Performance Development Planning (PDP)

The Performance Development Planning (PDP) integrates different components on one form, which contains three sections: Target Setting, Personal Development, and Performance Assessment. This form provides the basis for nominations on either the High Potential List or Sustained High Performance List, supports the determination of Variable Pay, and supports the Career Planning of managers.

Target Setting. In the Target Setting section of PDP, leaders and employees are expected to identify five targets, which will be used to determine the Variable Pay. Two of these five targets should be overall targets, and in most cases these two targets are the Volume and Net Profit of Sales targets of VDBN. The other three targets are Personal Targets. At the implementation of PDP as part of the Integrated Approach, managers were instructed that the targets should be stretching and should be agreed upon at the beginning of a year. When setting the targets, the manager and his assessor have to define in quantitative terms the criteria to assess when the target has been realised, and when it has been more than realised. With respect to the key concept of this case study, the concept of evaluative style, it is important to note that these criteria are rigidly applied; in other words, whether targets have been realised or not is assessed by applying the criteria, and is not subject to discussion. This means that there is no room for differences between leaders on how they evaluate performance with respect to the financial bonus (variable pay).

Personal Development Plan. In the second section of PDP, development of competencies and skills, performance is evaluated by comparing the competencies and skills needed for a specific job at a specific level with the competencies and skills held by the employee. To make these comparisons, for each function at each Work Level, the Human Resource Department has developed Job Profiles, which define the competencies and skills needed in a particular job. If the comparison shows that there is a gap between needed and actual competencies and skills, the manager and his assessor need to write down the gaps in the PDP, accompanied by an action plan which describes what will be done in the next year to improve these competencies and skills.

Performance Assessment. The third section includes a Summary of Performance, Wishes of the manager concerning his Career, and the Company (assessor's) View on Career Planning of the manager. The Summary of Performance is the outcome of the overall Performance Assessment. Before PDP was implemented, managers were instructed that Performance Assessment should be an open and honest discussion, should include a review of the targets and other key events, should assess "what" has been achieved as well as "how" it was achieved, should support learning from previous year's performance, and should provide the input for on or off the job training, courses, reading etc.

Conclusion

The Integrated Approach as a whole, and the element Performance Development Planning (PDP) in particular, provide the context of the empirical study on evaluative style and its relationship with subordinates' behaviour. Therefore, when interpreting the findings from this study later on, we will have to be aware of possible influences of this context on the findings. Where needed in the following chapters, I will give explicit reference to this context.

6.4 General observations

Before turning to differences between (the role of) individual leaders in performance evaluation, some general observations concerning performance evaluation style will be described.

First, all marketing subordinates interviewed are confronted with volume and profit targets for the unit as a whole. These targets are on the weekly agenda of the team meeting. During those meetings, these numbers are not discussed in great detail, but everyone can see the budgeted and actual results for the past few weeks in a glance. These volume and profit targets are "translated" into personal targets or personal projects, which should help to realise the target for the unit as a whole. In contrast, many subordinates from production units that were interviewed supported operational processes in the factories. Their leaders primarily focused on operational processes and results. Thus, the interviewed subordinates were subject to less supervision and directions than their colleagues in marketing. Therefore, feedback on performance of these subordinates tends to be more directed towards how subordinates contribute to and support operational processes, through different projects, than towards the realisation of pre-defined targets. Furthermore, many of these subordinates were in functions that were, as regards content, quite different from the educational background of their leaders, which makes it hard for leaders to evaluate the performance of their subordinates through the use of targets. In the marketing units, leaders had more intrinsic knowledge of the functions and responsibilities of the subordinates.

Second, subordinates have considerable influence on the choice of targets and the level of these targets, and on their Job Profile. However, subordinates in the production units had more influence than subordinates in the marketing units, since for the marketing units volume and profit targets were determined top-down.

Third, many leaders partly depend on informal controls: what they see and hear about/from the subordinate, either informally on the workfloor or, more formally, in management team meetings. In general, the organisational culture is informal and open, and not hierarchical at all. For example, one of the subordinates said: 'I don't really see my leader as my boss; only a little.' Or in the words of another subordinate reporting to a different leader: 'We are more colleagues of one another than being in a boss-subordinate relationship.' Another illustration of the lack of hierarchical structure was provided in the first meeting I had with

the Human Resource Department, when I tried to gain access to this particular organisation: They told me right away that they never used the word "subordinate" like I did, but that they always used the term employee instead⁴⁴. This observation is in line with another observation: that leaders provide their subordinates (employees) with immediate feedback. Many subordinates say they often receive feedback on their work from their leader or ask their leader to provide feedback. Most subordinates also indicate that they feel free to consult their leader when there is a problem, and that in many cases they actually ask for his/her opinion. Furthermore, most leaders have very frequent contacts with their subordinates; some even daily. Many leaders seem to be well "informed", i.e. being aware of what their subordinates are working on, what their problems are, and how they cope with it. Culture may not only be important at the organisational level, but at the national level as well. There is some evidence in the interview data that the open, direct organisational culture is typical Dutch. Some of the interviewees worked in other countries or in international settings, and they all indicated that international colleagues find the directness and openness of Dutch people odd, and that in other countries relationships between managers are more formal and more hierarchical than in Dutch companies.

The culture aspect, both at organisational and national level, is important because it may indicate that performance evaluation and the importance attached to formal evaluations may differ from one organisation to the other, and from one country to the other. As most of the literature on evaluative style is based on the Anglo-Saxon business practice, possibly some of the earlier defined dimensions of evaluative style are not applicable in the Dutch situation studied in this study, and other dimensions that have not been previously identified may arise that are important in a Dutch setting.

Fourth, in the accounting literature on evaluative style, the concept of evaluative style is based on the role of budgetary information in the performance evaluation of subordinates. Apart from the many problems associated with the existing instruments to measure evaluative style, most of the literature on behavioural effects of evaluative style is based on the expectation that high versus low reliance on accounting performance measures for the evaluation of subordinates have distinct behavioural consequences at the level of subordinates. In other words, differences in subordinates' job- and budget-related attitudes and performance are in part contributed to differences in the extent to which superiors use budgets or accounting performance measures to evaluate subordinates. However, in the interview data collected there is no mentioning of budgets when subordinates are asked how their performance is evaluated and by which criteria. This is at least surprising, and may suggest that the role of budgets for performance evaluation is not as important in this organisation as it is according to the accounting literature.

⁴⁴ In Dutch, I used the word "ondergeschikte" which had to be "medewerker" according to the Human Resource Department.

6.5 Sample characteristics: the use of categorical principal components analysis

The section above suggests that there is a difference in context for the respondents from marketing as compared to production in my sample. Furthermore, the production sample is quite varied in itself, as some leaders have subordinates in management positions and subordinates who are head of production. Other production leaders have subordinates in managerial positions only. Thus, although all leaders were leaders operating at Work Level 3, there seemed to be important differences between these leaders. I have argued in chapter 4 that, theoretically, task uncertainty could be an important contextual factor to consider. Therefore, it would be interesting to explore whether there are significant differences in task uncertainty between managers in different functions, and between managers and non-managers in the production sample. In total, I distinguished four different groups:

1. marketing managers;
2. production managers A: these are managers reporting to a Works Manager that had no production heads reporting to him or her;
3. production managers B: these are managers reporting to a Works Manager that also had production heads reporting to him or her;
4. production heads: although they were not called managers, they reported to a Works Manager, and, thus, were "colleagues" of production managers B.

The different function groups probably do not differ in task uncertainty only. It is likely that there are differences in education, and, based on personal observation when conducting the interviews, in age and tenure. Because such differences, when they exist, could be important for the analyses in subsequent chapters, I decided to describe characteristics of the respondents in this study for the different function groups as well. I used categorical principal components analysis (CATPCA) that is available in the SPSS software to display these differences. Because categorical principal components analysis is a relatively new data analysis technique, I will give some consideration to the technique before discussing the results on the sample characteristics. First, however, I will explore differences in task uncertainty among the function groups.

6.5.1 Differences in task uncertainty

Table 6.1 shows the average (mean and median) task uncertainty scores for each function group. Higher scores on the task uncertainty measure that I used in this study indicates lower task uncertainty, and lower scores on the task uncertainty measure indicate higher task uncertainty. On average, marketing managers report the highest task uncertainty (lowest score), and production heads the lowest (highest score). This is as expected. The results of the Kruskal-Wallis test (see Table 6.1) indicate that there are indeed significant differences among the four function groups ($\chi^2 = 15.455$, $p = .001$). Post hoc analyses using Mann-Whitney tests, displayed in Table 6.2, indicate that production heads report significantly lower levels of task uncertainty than marketing managers ($p = .000$). None of the other

differences are significant at the 10% level when the Holms stepdown procedure is used. This procedure takes into account that 6 differences are tested⁴⁵.

Table 6.1 *Results of Kruskal-Wallis test for Differences in the Level of Task Uncertainty between Different Function Groups*

Function	N	Mean	Median	Kruskal-Wallis test	
				χ^2	P
Marketing Managers	13	16.154	16.000	15.455	.001
Production Managers A	14	19.571	19.500		
Production Managers B	5	18.400	17.000		
Production Heads	25	22.800	23.000		

Table 6.2 *Results of Multiple Comparisons of the Level of Task Uncertainty between Different Function Groups (Using the Holms Stepdown Procedure)*

I	Comparison	p-value	$\alpha/(k-i+1)$
1	Marketing Managers vs. Production Heads	.000	.100/(6-1+1)= .017
2	Production Managers A vs. Production Heads	.030	.100/(6-2+1)= .020
3	Marketing Managers vs. Production Manager A	.064	.100/(6-3+1)= .025
4	Production Managers B vs. Production Heads	.079	.100/(6-4+1)= .033
5	Marketing Managers vs. Production Manager B	.458	.100/(6-5+1)= .050
6	Production Managers A vs. Production Managers B	.606	.100/(6-6+1)= .100

6.5.2 Differences in age, tenure and education

Table 6.3 presents details of the sample characteristics in terms of age, tenure and educational background of the respondents. This table provides a first impression of differences between the function groups. The average age of all respondents was 34,9 years, and respondents had been in their current position for an average of 1 to 2 years. Age and tenure differed considerably between the four function groups, however. The educational background of the production heads was quite different from the managers. Only 1 marketing and 1 production manager did not have a "HBO" (21,9%) or a university degree (71,9%). Of the 25 production heads, 10 respondents (40%) had a "HBO" -background, 3 had a university degree (12%), and 12 respondents had another educational background (48%). Of the 57 respondents, 44 respondents (77,2%) were male. Because it is difficult to see differences between the four function groups from table 6.3 due to the level of detail and number of

⁴⁵ Using one-way ANOVA leads to similar results: There are significant differences in agreement with evaluation between the four function groups ($F=7.399$, $p=.000$). The use of Levene's test reveals that homogeneity of variance can be assumed (Levene statistic = .524, $p=.668$). Similar to the non-parametric analyses that have been reported, post-hoc analyses reveal that the only significant difference is that production heads report lower levels of task uncertainty than marketing managers.

Table 6.3 Sample characteristics

	Total		Marketing managers		Production managers A		Production managers B		Production heads	
Age	34,96 years		30,77 years		41,79 years		31,00 years		34,12 years	
	Count	%	Count	%	Count	%	Count	%	Count	%
Gender:										
1. Male	44	77,2	8	61,5	13	92,9	2	40,0	21	84,0
2. Female	13	22,8	5	38,5	1	7,1	3	60,0	4	16,0
Education:										
1. "HBO"	17	29,8	1	7,7	5	35,7	1	20,0	10	40,0
2. University	26	45,6	11	84,6	9	64,3	3	60,0	3	12,0
3. Other	14	24,6	1	7,7			1	20,0	12	48,0
Tenure with VDBN:										
1. < 6 months	1	1,8							1	4,0
2. 6 months - 1 year	1	1,8							1	4,0
3. 1-2 years	6	10,5	1	7,7			2	40,0	3	12,0
4. 2-5 years	18	31,6	9	69,2	4	28,6			5	20,0
5. 5-10 years	14	24,6	3	23,1	3	21,4	3	60,0	5	20,0
6. > 10 years	17	29,8			7	50,0			10	40,0
In current position:										
1. < 6 months	5	8,8							5	20,0
2. 6 months - 1 year	10	17,5	4	3,8	2	14,3	1	20,0	3	12,0
3. 1-2 years	10	17,5	2	15,4	2	14,3	3	60,0	3	12,0
4. 2-5 years	20	35,1	6	46,2	5	35,7	1	20,0	8	32,0
5. 5-10 years	11	19,3	1	7,7	5	35,7			5	20,0
6. > 10 years	1	1,8							1	4,0

Table 6.3 (continued)

Working under current leader:	1		3		3		1	
	1. < 3 months	1,8	3	23,1	3	21,4	3	4,0
	2. 3-6 months	10,5	6	46,2	9	64,3	8	12,0
	3. 6 - 12 months	35,1	3	23,1	2	14,3	10	32,0
	4. 1 - 2 years	42,1	1	7,7	2		3	40,0
	5. > 2 years	10,5						12,0
Number of evaluations by current leader:								
1. Null	21	36,8	5	38,5	3	21,4	2	44,0
2. One	27	47,4	7	53,8	8	57,1	3	36,0
3. Two	5	8,8			3	21,4		8,0
4. Three or more	4	7,0	1	7,7				12,0

categories involved, I used Categorical Principal Components Analysis (CATPCA) to display the relationships in this table graphically.

Categorical Principal Components Analysis (CATPCA)

Because CATPCA has not been used much in the organisational and accounting literature, I will briefly explain what CATPCA is and how it can be used, by summarising the more elaborate explanation that Linting et al. (2004) offer. In short, CATPCA "simultaneously quantifies categorical variables while reducing the dimensionality of the data" (Meulman et al., 2001, p. 27). Thus, the goal of CATPCA is to reduce the number of dimensions in a larger set of (categorical) variables with a minimal loss of variance accounted for. Unlike standard principal components analysis, CATPCA does not assume linear relationships between numeric variables, but models nonlinear relationships as well since it allows variables that are scaled at different levels. CATPCA transforms variables by assigning numeric values to each category of variables that are scaled at a nominal or ordinal level. This quantification of variables occurs through optimal scaling. CATPCA uses the optimally scaled data to find a pre-chosen number of components that explains as much variance of the transformed variables as possible. A major advantage of CATPCA as compared to standard Principal Component Analysis is that CATPCA does not require the data to be (multivariate) normally distributed, which enables a researcher to use it even with small samples with many variables as in this study. Another advantage of CATPCA is that it produces graphs that are relatively easy to interpret. Finally, CATPCA enables the use of supplementary variables, which are variables that do not influence the solution but can be displayed in the solution. This enables visualizing the relationship between group membership, such as different function groups in my study, and individual respondents in the plot, without affecting the solution. Because of these characteristics, CATPCA is a powerful technique to explore patterns in data (respondents, groups, and variables).

Transformation plots

I applied CATPCA to the variables listed in Table 6.3, using function groups and leaders as supplementary variables. Supplementary variables are variables that are fitted in the solution but do not define the dominant dimension. I treated the variables education and gender as multiple nominal, which means that the grouping of objects in categories (for example, university or "HBO") is preserved, but not the order of categories. The optimal scaling level for all other variables was set to ordinal, because I wished to retain the ordering in the categories for these variables without making the assumption of equal intervals between category numbers. Since there is no need for linear transformations, the use of ordinal transformations allows more freedom in the analysis. The transformation plots for all variables are displayed in Appendix E. The transformation plots show the original values of the variables on the horizontal axis and the optimally scaled quantifications on the vertical axis. For most variables, these plots support the choice for ordinal transformations, since

most transformations are not linear as they would have been if the variables had been treated as numeric as in standard principal components analysis. An interesting transformation occurs for the variable "number of evaluations by current leader" (nr_eval) (see panel B of figure E-2 in Appendix E). The transformation plot for this variable reveals that the quantifications for categories 2, 3 and 4 are equal. This means that whether subordinates have been evaluated just once or more than once by their current superior does not differentiate between subordinates. Thus, these categories could be collapsed into a single category.

Number of dimensions

The most important results of the CATPCA analysis are summarized in tables and figures below. Additional output from the CATPCA analysis is reported in Appendix E.

Table 6.4. Model Summary of CATPCA analysis

Dimension	Cronbach's Alpha	Variance Accounted For					
		Multiple Nominal Variables		Non Multiple Variables		Total (Eigenvalue)	% of Variance
		Total	% of Variance	Total	% of Variance		
1	.811	.718	35.921	2.558	51.153	3.276	46.801
2	.408	.438	21.903	1.100	22.006	1.538	21.976
Total	.891 ¹	.578 ²	28.912 ²	3.658	73.158	4.236 ³	60.516 ³

¹ Total Cronbach's Alpha is based on the total Eigenvalue.

² Mean over dimensions.

³ Because there are Multiple Nominal variables, total Eigenvalue and % of Variance Accounted For is not the sum over dimensions.

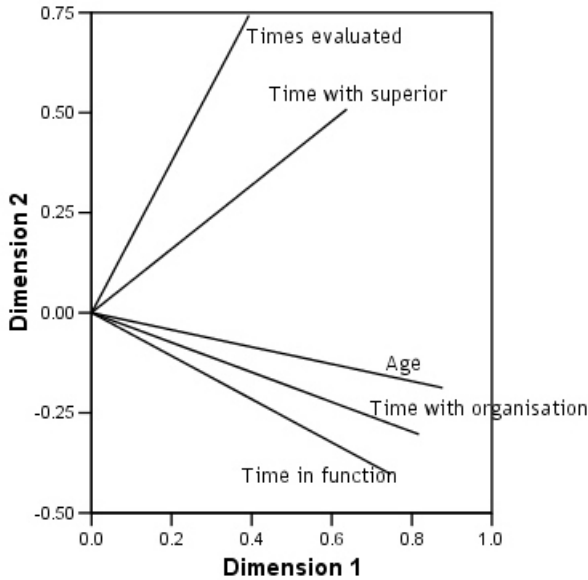
Table 6.4 displays the model summary based on two dimensions. The reduction of data to two dimensions accounts for 58.6 % of the variance in the optimally scaled variables (and not variance in the original variables). If I had used three dimensions, 62.6% of the variance would have been explained, thus increasing the explained variance with only 4.0%. In the three-dimension solution, the eigenvalue of the third dimension would be 1.13, with a value of Cronbach α of .131, which is low and indicates a bad fit. In addition, fit (the sum of eigenvalues of the dimensions) is 4,685 for the two-dimensional solution, and 5,011 for the three-dimensional solution. The increase in fit when the three-dimensional solution is chosen is small. Thus, I preferred the two-dimensional solution.

Interpretation of dimensions

For each variable, CATPCA presents the values assigned to each category through optimal scaling, i.e. the quantifications, the average scores on each dimension of all respondents (i.e., objects) in the same category, i.e. centroid coordinates, and the coordinates of the categories when they are required to be on a line, i.e. vector coordinates. This enables a

graphical display of complex relationships between variables, but also between respondents and variables. The plot of component loadings (figure 6.4) displays the vectors, representing each of the (ordinally transformed) variables.

Figure 6.4 *Plot of Component loadings*



All variables, except the number of times evaluated, have high, positive component loadings on the first dimension. The second dimension is associated with years working under present superior and number of evaluations. This means that respondents with high positive scores on dimension 2 will have a high score on years working under present superior and on number of evaluations by current leader (i.e., at least once), more or less independently of tenure (in current job and with the company) and age. The plot of component loadings also shows that the three variables that measure tenure in current job, tenure with the company, and age are closely related, because the vectors appear close to each other.

Differences between function groups

The plot of component loadings helps to interpret differences between respondents and groups of respondents. Since I am interested in differences on these variables across function groups, it would be helpful to display the average score of respondents for each function group in the same two-dimensional space. Figure 6.5 shows the category point plot for the four function groups.

Figure 6.5 Plot of Category Points: Function Groups

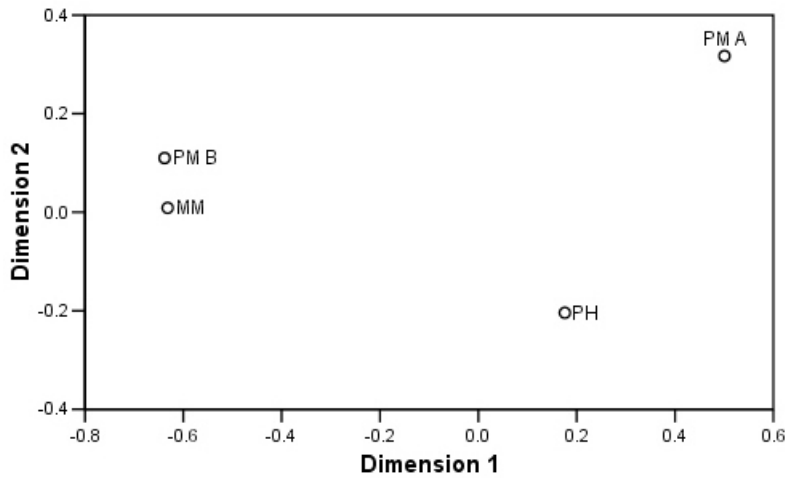
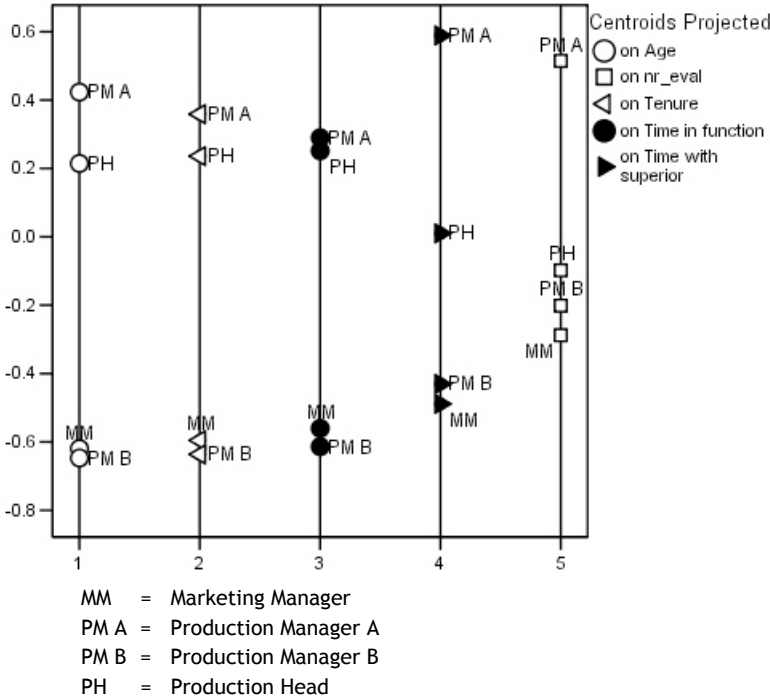


Figure 6.5 shows that, on average, marketing managers and production managers B have similar scores on both dimensions. Their scores on dimension 1 differ significantly from production managers A and production heads, who have similar scores on dimension 1 but opposite scores on dimension 2. Thus, on average, marketing managers and production managers B have shorter tenure, are younger, include more females, and are higher educated than production managers A and production heads. The major difference between production managers A and production heads is that on average production heads are less educated than production managers A, and have been working under their current leader for a shorter period of time. This analysis is confirmed in figure 6.6, that displays the centroids (average score of respondents in a category) of the four function groups projected on age, number of evaluations, time with current leader and tenure (with company and in present function). Note that the differences between the function groups refer to differences in the optimally scaled, transformed variables.

Differences between leaders

Since some leaders in production in my sample have subordinates that are managers and some that are production heads, the differences between these two groups could be important in further analyses in subsequent chapters. Therefore, in addition to the differences between function groups, figure 6.7 displays the category point plot for the twelve leaders. The right-hand side displays the plot for subordinates grouped by the twelve leaders without considering differences in within each group of subordinates reporting to the same superior. On the left-hand side, the plot displays fifteen category points. The fifteen categories were formed by splitting production heads and production managers who report to the same leader into two groups of respondents. This was the case for three of the twelve

Figure 6.6 Summary of average differences between function groups



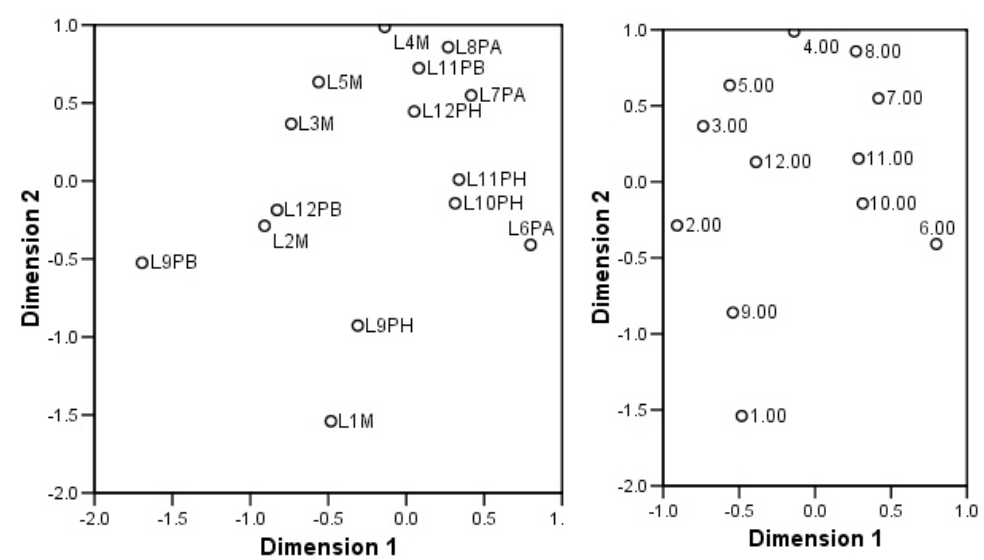
leaders (leaders 9, 11 and 12), resulting in a total of fifteen categories.

Figure 6.7 shows, first, that within function groups there are differences between subordinates grouped by leader, especially on dimension 2 (vertical). For example, the five marketing groups are all around average or slightly below average on the first dimension, but differ a lot on the second dimension. Similarly, the three production manager A groups have similar (relatively high positive scores) on the first dimension, but differ mainly on scores on the second dimension. Secondly, the plot shows that for each of the three leaders (9, 11 and 12) with two function groups the production managers B on average report lower scores on the first dimension than the production heads (see 9PB and 9PH, 11PB and 11PH, and 12 PB and 12PH in the left-hand side of figure 6.7).

6.6 Conclusion

In this chapter I have described the context of the empirical study on evaluative style and its relationship with subordinates' behaviour. I described contextual elements that are of particular interest in interpreting the findings from this study. The three most important contextual elements for understanding the findings in the following chapters are as follows. The first contextual element is the Integrated Approach, and the element Performance

Figure 6.7 Plot of Category Points: Differences between leaders



Development Planning (PDP) in particular. This element refers to the performance evaluation system as it has been formally implemented within VDBN just a few months before this study was conducted.

The second contextual element is the organisational culture (climate) at VDBN. On the hand, as part of a large multinational, formal, very detailed systems such as the Integrated Approach have been implemented within VDBN. On the other hand, managers are very direct and open in their communication. Not only in their communication with peers, but also with their boss, and also with me as a researcher. The fact that none of the interviewees objected to recording the interviews using my tape recorder is characteristic of this open culture.

The third important element described in this chapter is that there are differences in function groups within the sample. Subordinates in my sample not only report to twelve different leaders, they also come from four different function groups: marketing managers, production managers A, production managers B, and production heads. In three groups of subordinates reporting to the same leader subordinates come from two different function groups. There are important contextual differences in the level of task uncertainty, but also in age, tenure and education between these function groups. These differences suggest that even within groups of subordinates from the same function group, or even from the same leader, there appear to be contextual differences that could influence the (perceived and/or intended) evaluative style and the (perceived) effectiveness of that style within that particular context.

7 The construct of perceived evaluative style: empirical evidence

7.1 Introduction

In this chapter I will explore whether there are meaningful differences in evaluative style within Van den Bergh Nederland (VDBN). This exploration will be based on data from the interviews and on data from the written questionnaire.

The interview data will be analysed to gain a first insight into underlying dimensions of evaluative style that seem relevant within VDBN. The primary aim of identifying these general dimensions is to help us to understand the results of subsequent, more detailed analyses that are based on data obtained from the written questionnaire. In the written questionnaire, subordinates were asked to indicate the importance that their superior attaches to 17 different aspects of the performance evaluation process. To explore whether these 17 aspects captured relevant differences in evaluative style within VDBN, three topics in particular will be addressed.

First, I will analyse the relevance of the 17 aspects in this organisation. For this purpose, respondents were not only asked to indicate how much importance their superior **attaches** to each of the 17 aspects, i.e. perceived evaluative style, but also how much importance they think their superior **should attach** to those aspects, i.e. preferred evaluative style. How much emphasis subordinates think their leaders should lay on each of the 17 aspects may be used as an indication of the relevance for performance evaluation of these 17 aspects, thereby providing at least some evidence for the content validity of these items.

Second, I will explore whether there are underlying dimensions of evaluative style emerging from the written questionnaire data. To explore this, the underlying structure of the 17 aspects of perceived evaluative style will be analysed.

Third, I will explore whether evaluative style is -partly- a group level phenomenon, in addition to an individual level phenomenon. This involves two different issues. The first issue is whether the perceived emphasis on different aspects of evaluative style differs across groups, i.e. whether there are significant between-groups differences. The second issue is whether there is (considerable) agreement among the scores of subordinates working under the same superior, i.e. agreement within groups. These two issues are different, in that the occurrence of between-group differences is independent of agreement within groups, and reversed (George and James, 1993).

The chapter is organised as follows. The exploration of the interview data on evaluative style will be covered in § 7.2. Next, I will describe the exploration of the data from the written questionnaire: the relevance of the measured aspects of evaluative style will be the topic of § 7.3, underlying dimensions of evaluative style will be explored in § 7.4, and evaluative style

as a group-level phenomenon will be discussed in § 7.5. The chapter will be completed by a discussion of the findings (§ 7.6) and the articulation of some final conclusions (§ 7.7).

7.2 Differences in perceived evaluative style: a qualitative analysis

7.2.1 Introduction

The interview data has been analysed at the level of groups. Thus, the data from the interviews has been summarised by leader, and differences in perceived evaluative style will be reported in this section as differences between leaders. I followed the following procedure to conduct this analysis. First, each single interview was analysed to detect dimensions of evaluative style that seemed to be relevant in this particular organisation. Next, I tried to summarise the dimensions into a number of distinctive “common” dimensions of evaluative style that were shared by a number or even most of the interviewees. Finally, I tried to classify each leader on the common dimensions that emerged from the data.

7.2.2 Underlying dimensions of evaluative style emerging from the interview data

To explore the differences between the twelve groups (leaders), in a first analysis of the interviews I focused on the role of targets in performance evaluation, as this has been the most important difference according to the RAPM-literature. Although the role of targets seems to be a relevant dimension of evaluative style in this study too, a further analysis of the interview data revealed other differences in how leaders evaluate their subordinates as well. After I have described differences in the role of targets, I will turn to these other differences.

Role of targets

Although all leaders attach importance to the final results of actions and receive weekly or monthly reports about their unit’s performance, the role of targets in the evaluation of their subordinates differs.

Two leaders in particular, leader 1 and leader 7, evidently emphasise targets more than other leaders in their communication with subordinates. However, for leader 1 this emphasis could be due to the fact that he is leader of a project-team with a specific assignment to meet a target; all other leaders are leaders of an organisational unit, which makes comparisons with leader 1 difficult. Yet, although leader 1 emphasises the target, the emphasis is not on meeting the target. Both subordinates indicate that everyone involved in the project knows that the target will not be realised, but efforts are devoted to get a result as near to the target as possible. Leader 1 has a communicating, motivating and stimulating role to achieve this result. Therefore, although the target will not be met, the target remains important as a motivating goal. In contrast, leader 7 does emphasise meeting targets

and agreements. One of this leader's subordinates says the leader attaches more importance to the final results than to how these results were achieved, while the subordinate himself feels that when his performance is evaluated, how the results were achieved should be most important.

The leaders 6, 8, 10, 11, and 12 primarily found the evaluation of their subordinates on subjective criteria, e.g. on their observation of the subordinates, on how subordinates participate in the management team meeting, on their general impression. As one subordinate of leader 6 said:

"My performance evaluation is mainly subjective. My leader bases the evaluation on subordinates' actions. How do you behave, how do you work, what is your involvement? And whether it has a direct result in the factory comes second."

And a subordinate of leader 11 said:

"Knowing that people have confidence in you is more important than a few hard targets that do not really exist for this job."

This does not mean, however, that targets are not important at all. Targets are important, and are part of a balanced scorecard in many of the units. These targets are usually discussed weekly, or at least monthly, within the management team. Personal targets for managers have been derived from these targets, and are implicitly discussed when discussing weekly or monthly results based on a balanced scorecard. One of the subordinates of leader 8 said:

"...To be honest, we discussed those targets at the beginning of the year, we set them then. We monitored part of those targets automatically by way of the balanced scorecard, because they appeared there one way or the other. But we have never had -not the two of us, and, according to me, not or hardly as management team too- any premature consultation like: "Look, we are not going to realise that, so should we not redirect that?" That was never initiated, not by my boss and not by me. But there was a number of targets, they were so at the top of the balanced scorecard, that they were discussed indeed each week, or each month. But without converting them to someone's personal targets. But well, that was obvious to everyone. If it was about the efficiency of a particular product line that was below target, than it was clear that it was my personal target. But there has never been a premature consultation about my personal targets, or something like that. We had no need for that."

This illustrates the need to distinguish carefully between team targets and personal targets, and between the role of targets in managing the daily operations as a part of the more general planning and control cycle and the role of personal targets for evaluation purposes.

All other leaders, i.e. leaders 2, 3, 4, 5, and 9, use both hard criteria (targets) and soft criteria (how results have been achieved). In the words of a subordinate of leader 9:

“My evaluation is based on what I have done, whether I do my job well. Whether it is one percent more or less, is less important.”

Or quoting one of leader 2’s subordinates:

“Leader 2 does not focus on the numbers or the budget, but emphasises creativity, looking at alternatives.”

Coaching role of leader

In the organisation studied, in theory, leaders are stimulated by the Human Resources Department to be coaches more than bosses. Therefore, many subordinates refer to the coaching role of their leader. With regards to performance evaluation, the coaching role of the leader is most clearly shown by the extent to which subordinates receive feedback from their leaders. As discussed under the general observations in chapter 6 (see § 6.4.1), most leaders indeed provide their subordinates with immediate feedback, both positive and negative. Yet, the extent to which and the manner wherein differs between leaders.

There is clear evidence in the interview data that, based on subordinates’ perception, leaders 1, 3 and 12 take their coaching role seriously. Especially leader 12 coaches his subordinates by talking with them about how to be an effective manager. Leader 12 provides the subordinates with a lot of feedback on how they approach certain issues, on their interactions with other people, on their involvement in and contributions to meetings and discussions. As part of this feedback, leader 12 tells them how he deals with similar situations. There is no formal discussion of how things are progressing, but occasionally he asks them critical questions regarding their performance. Leader 3’s subordinates receive many remarks, suggestions or directions. He keeps his subordinates informed about what he is working on, and on what he reads and learns himself. This behaviour can partly be explained by the fact that leader 3 is the leader of a “category management” unit, which, at the time of the interviews, was a new department and function within the organisation. What leader 3 thinks is the most important for his subordinates is to promote category management as a process within the organisation: how to create added value within other units, how to organise things in such a way that others are informed and that these things are finished in time. Finally, leader 1, as discussed above, is able to motivate the subordinates to try to realise the target, despite the fact that the whole project team knows this is impossible. Leader 1 motivates and stimulates subordinates mainly through team meetings, and by providing them with immediate feedback, both positive and negative. Therefore,

even though the subordinates do not mention coaching, their remarks justify the conclusion that leader 1 coaches subordinates to help them realise the target.

As to the coaching role of leaders 2, 6, 9, and 11, the evidence is mixed and limited. Mixed, because one subordinate tells a different story than another subordinate. Limited, because at least one of the subordinates receives feedback regularly, but the nature of this feedback is unclear, and there is no explicit mentioning of coaching. Leader 2 has a monthly meeting with one of the subordinates, to discuss progress on skills and competencies. According to leader 2, this is necessary, because based on age and experience, the subordinate should have a function at the next Work Level soon. In contrast, the other subordinate says she would like to have a formal discussion of how things are progressing at the individual level more often. She describes her leader's style of leadership as "loose and giving people freedom. Leader 2 gives you space, but sometimes it feels too much like swimming." This subordinate wants more coaching, even though leader 2 "does provide confidence by saying: please come if it does not go well. If not, he will ask how it goes". There is a difference between subordinates of leaders 9 and 11 too. For both leaders, this difference could be explained by the difference in function between the subordinates: in both cases, one subordinate works in the factory and his work is linked with operational processes, while the other subordinate has a supporting function for the production site as a whole and reports to another leader as well. Both leaders communicate daily with the former subordinate, but have less direct contact with the latter subordinate. Therefore, for both leaders, only one of their subordinates indicates that he receives a lot of feedback, that his leader knows exactly what he has done and is doing, and that performance evaluation is very informal and open. Yet, these subordinates do not mention coaching explicitly, but based on the evidence the role of leader 9 and leader 11 could be described as a coaching role for at least one of their subordinates. The other subordinate of both leader 9 and 11 receives less feedback, but does have an open and informal relationship with the leader. Similar to the situation as described for the leaders 9 and 11, leader 6 communicates with one of his subordinates more often in an informal way than with two other subordinates. This however could be explained by age also: this one subordinate is much younger, and in his first job, while the other two are older and in (most likely) their last job. Again, although the youngest subordinate does not explicitly refer to coaching, as with leader 9 and leader 11, most likely the role of leader 6 towards this subordinate, involving many informal conversations, could be described as coaching.

There is limited evidence that leader 5 and leader 10 fulfil a coaching role. One of leader 5's subordinates says he receives feedback on skills and competencies, but not on a daily basis: "It's just small things: "I would have said it like this" or "You should pay attention to that". Only one subordinate of leader 10 was interviewed, since this was the only subordinate for whom a PDP-form was used. This subordinate had a supporting function in a production

environment. According to the subordinate, his relationship with the leader was very informal and open: "Criticise whenever you like. That will be short, informal, that you do not have the feeling that it was an assessment." There is no evidence in the interviews, however, that either leader 5 or leader 10 actively helps the subordinate to get better and spends time with the subordinate to discuss things.

Subordinates of leader 4, 7 and 8 receive hardly any or no coaching. According to one subordinate of leader 4:

"Leader 4 gives no coaching on how to narrow the gaps in skills and competencies. He is accessible to solve issues, but the subordinates have to take the initiative."

None of the subordinates of leader 7 mentions receiving feedback on their work, or any coaching role at all. The subordinates of leader 8 indicate that they have many informal contacts with their superior, but they do not mention explicitly that leader 8 provides direct feedback on how they perform as managers. Indeed, there seems to be little attention for career development. A possible explanation is offered by one of the subordinates of leader 8 who states:

"I think that more attention is being paid to the output of the plant, and that people than could be subordinated... You have to do it together, and the club in this plant is not that big anymore. It has been reduced considerably. And when there are a number of changes and there is already a sort of instability in the team, that then the person who really needs a next step perhaps will be held a little longer. I am not saying that it really happens, but... I have the feeling that it could happen..."

Use/implementation of PDP

As I reported in chapter 6 (§ 6.3), a new company-wide performance evaluation system (PDP) had been introduced in the same year the interviews were held which was supposed to be the basis for performance evaluation. The new system differentiates between a target setting part, the Variable Pay Schedule (VPS) and a part relating to skills and competencies. The interview data reveal that there were differences between leaders in the extent to which the new system had been actually implemented and used.

The new system has been fully implemented by leaders 9 and 12. That is, there are no signs that parts of PDP have not been implemented or filled out, or that there has been no follow-up.

Leaders 1, 2, 3, 5, 8, 10, and 11 have filled in the complete PDP-form with their subordinates, but have never used it during the year. The new system is used in the same

way as the old system: criteria are defined and written down at the beginning and performance is assessed at the end of the year. A few quotes will illustrate this approach:

- "I have not discussed my PDP with my boss for one second after filling it in. It has not received a follow-up."
- "Gaps in skills and competencies were identified, but they have not received any follow-up. I have taken some initiatives myself. The same applies to targets: they were set in the PDP-form, but during the year we have never talked about them again."
- "PDP should have more substance and should be more than an evaluation once very year. It should provide tools to improve your performance. This year that has not been fully realised."
- "We have started it last year. I really had two discussions about it, which was the kick-off. I think the real work still has to be done this year. The conversations before with the new forms, that was all very concise... And perhaps during the year too little attention has been given to the things discussed in PDP. But well, if I had felt the need for it I could have done it. I knew for myself whether things went right, whether it was a little bit aligned with what the company wants."

This last quote illustrates an important observation from the interviews in this respect: many subordinates who say that there has been hardly any follow-up to the PDPD-form during the year directly state that they could and should have taken the initiative, but did not. Thus, they do not "blame" their superior for not bringing it up, because they realize that they do not find it that important themselves. This attitude is well illustrated in the following quote from a subordinate reporting to leader 8:

"We have used PDP for one year now. I have hardly noticed it. That too is nobody's fault, because I myself could have seen to it that I had noticed it more. But I have not found the time in my daily job to really pay much attention to it. I thought it was interesting to fill it in the first time. And I though it was interesting to compare what my boss filled in for me. And it was fascinating to discuss the differences that we found. There were not that many, because my boss and I had a good understanding with one another. But there were some differences. But the step that I did not take, and that others did not take for me, is: we have these points, on which you fall short, we should do this or that about it. That would be a further step: on these points you perform on Work Level 2b, to make it Work Level 3 we should work on this and this."

Leaders 4, 6 and 7 have not implemented the PDP-system completely. Targets have been set, and skills and competencies have been identified, but a clear development plan based on these gaps has not been defined. According to one subordinate of leader 6:

"Hardly any agreements have been made, and the agreements that were made have had hardly any follow-up."

And a subordinate of leader 4 states that leader 4 did not understand what he had to fill in. They had long discussions about the meaning of “zero”, “plus” or “minus”: “There was no time left to identify gaps and to discuss how those gaps could be reduced.”

7.2.3 Conclusion

An analysis of the interview data reveals that differences in how leaders evaluate their subordinates’ performance can be described by three “dimensions”: the role of targets, the coaching role of the leader, and the use/implementation of the PDP-system. The results of this analysis are summarised by leader in Table 7.1.

Table 7.1 *Differences in performance evaluation style between leaders based on interview data*

Leader	Relative importance of targets	Evidence of coaching role		Use/implementation of PDP	
		Providing feedback	Active coaching	VPS/Full	Follow-up
1	Large	Yes	Yes	Full	No
2	Average	Yes	Mixed, limited	Full	No
3	Average	Yes	Yes	Full	A little
4	Average	Limited	Hardly	VPS	-
5	Average	Yes	Limited	Full	No
6	Small	Yes	Mixed, limited	VPS	-
7	Large	No	No	VPS	-
8	Small	No	No	Full	No
9	Average	Yes	Mixed, limited	Full	Probably
10	Small	Yes	Limited	Full	No
11	Small	Yes	Mixed, limited	Full	No
12	Small	Yes	Yes	Full	Probably

Clearly, based on these three dimensions, there seem to be differences between how leaders evaluate performance. Yet, the results as displayed in table 7.1 need to be interpreted with caution, as each leader has been classified on the three general dimensions that emerged from the data. So, the interview data has been summarised by leader, and differences in perceived evaluative style are reported in table 7.1 as differences between leaders. The analysis of the interview data however also reveals that descriptions of evaluative style by subordinates who report to the same leader do not always coincide. Possible explanations for these differences in perceived evaluative style could be differences between subordinates in function, differences in tenure, and differences in age. These issues will be explored in more detail in chapter 8 using the quantitative data from the written questionnaire.

7.3 Relevance of measured aspects of evaluative style

7.3.1 Introduction

To assess the relevance of the 17 aspects included in the written questionnaire to explore differences in evaluative style, in this section I will compare the average perceived

importance of each aspect with the average preferred importance of the same aspect at the organisational and group level. Since I have not developed specific propositions earlier in this thesis with regard to perceived and preferred evaluative style, I will first develop such propositions in this section before conducting the analyses.

I expect that on average some aspects of performance evaluation receive more emphasis than participants feel those aspects should receive, while other aspects on average receive less emphasis than participants prefer, and still other aspects receive the same emphasis as preferred. However, it is impossible to predict for which aspects the preferred emphasis will be higher than, lower than, or equal to the perceived emphasis.

Furthermore, it is assumed that the average preferred importance score on each aspect of evaluation at least partly reflects the norms and values held by participants in this organisation regarding performance evaluation⁴⁶. Thus, I expect that there is more agreement among subordinates on the preferred importance scores on aspects of evaluative style than on the perceived importance scores on those aspects.

In short, I expect:

- P13. There are significant differences between the mean perceived and preferred importance scores of some but not all aspects of performance evaluation; for some aspects these differences are positive, for others negative.*
- P14. There is more agreement among subordinates on the preferred importance scores of aspects of evaluative style than on the perceived importance scores on those aspects.*

7.3.2 Data analysis technique

To test the differences between the perceived (real) and preferred (ideal) emphasis on the 17 items of evaluative style (proposition 13), I conducted Wilcoxon signed ranks tests. Proposition 14, stating that I expected more agreement among subordinates on the preferred importance scores of aspects of evaluative style than on the perceived importance scores, was not actually tested statistically, but as an indication I looked at the standard deviations of the 17 items of perceived and preferred evaluative style.

7.3.3 Results

The results of the Wilcoxon signed rank test to test for differences in perceived and preferred evaluative style are reported in Table 7.2. Table 7.2 shows that all differences but

⁴⁶ The importance of prevalent norms and values in a particular organization for understanding the impact of evaluative style on subordinates' behaviour has been explained in chapter 4.

Table 7.2. Comparison of Average Perceived and Preferred Importance of Aspects of Evaluative Style, using Wilcoxon Signed Ranks Test

	Difference between perceived and preferred for individual respondents			Number of groups reporting lower, equal or higher perceived emphasis than preferred (n = 12)		
	(n = 57) Perceived Mean (sd)	Preferred Mean (sd)	p-value	lower	equal	higher
short-term goals	3.86 (.79)	3.61 (.62)	.014 ^a	2	2	8
long-term goals	3.49 (.83)	4.16 (.62)	.000 ^b	10	1	1
financial information	3.00 (.87)	3.42 (.71)	.001 ^b	10	1	1
non-financial information	3.45 (.74)	3.79 (.59)	.001 ^b	9	3	0
deviations from agreed performance	3.18 (.85)	3.86 (.69)	.000 ^b	9	2	1
Your explanation of your performance	3.39 (.88)	3.98 (.69)	.000 ^b	9	2	1
objectively measurable performance	3.39 (.84)	3.86 (.67)	.000 ^b	11	1	0
his/her own intuition	3.79 (.84)	3.35 (.90)	.004 ^a	2	0	10
performance in the year past	3.75 (.84)	3.88 (.68)	.184 ^b	7	1	4
Your plans for the next year	3.43 (.85)	3.91 (.91)	.001 ^b	8	2	2
positive aspects of your performance	3.56 (.73)	3.95 (.74)	.003 ^b	8	3	1
negative aspects of your performance	3.28 (.84)	3.91 (.77)	.000 ^b	11	0	1
performances of your unit	3.46 (.97)	4.02 (.58)	.000 ^b	10	1	1
personal, individual performance	3.65 (.79)	3.89 (.67)	.016 ^b	8	3	1
causes of failings in performance	3.00 (.87)	3.91 (.79)	.000 ^b	11	1	0
possibilities to improve your performance	3.16 (.94)	4.28 (.75)	.000 ^b	12	0	0
information from conversation	3.25 (.85)	3.93 (.70)	.000 ^b	10	2	0

^a Mean perceived emphasis is higher than mean preferred emphasis
^b Mean perceived emphasis is lower than mean preferred emphasis

one are significant at the 5% level. Surprisingly, of the 16 significant differences only two differences indicate that in the actual situation a higher emphasis is placed on this item than preferred. These two items are "short term goals" (3.86 vs. 3.61) and "his/her intuition" (3.79 vs. 3.35). All other items on average receive less emphasis than preferred. However, from my expectations that perceived evaluative style varies across leaders, but preferred evaluative style does not, it follows that I expect differences between importance ratings on perceived and preferred evaluative style to vary across leaders. To investigate whether the result that on average less emphasis is given to aspects of evaluation than preferred by subordinates, holds across leaders, I compared the differences between perceived and preferred also for each of the groups. The results are reported in the right-hand side of Table 7.2. The finding that preferred emphasis is higher than perceived emphasis holds in most of the groups too. Again, the two items, "short-term goals" and "his/her intuition" are exceptions, indicating that for those aspects of evaluation subordinates prefer restricted rather than expanded use.

Furthermore, the scores on the items in the preferred ('ideal') situation tend to have a lower standard deviation than the scores on the same items in the perceived ('actual') situation⁴⁷, indicating that differences in opinion about the ideal style are smaller than the perceived differences in actual evaluative style. This result supports proposition 14, and supports the use of preferred evaluative style as a measure of norms and values.

Finally, if preferred evaluative style reflects norms and values, the absolute mean scores in Table 7.2 suggest that in this particular organisation subordinates feel that the boss's intuition (3.35), financial information (3.42), and short-term goals (3.61) are the least preferred performance evaluation items, whereas possibilities to improve/develop your performance (4.28), long-term goals (4.16) and performances of your unit (4.02) are the three mostly valued items of performance evaluation.

7.4 Underlying dimensions of perceived evaluative style

7.4.1 Introduction

The results in the previous section seem to support the relevance of the seventeen items included in the questionnaire to measure evaluative style. These items were included in order to capture underlying dimensions of evaluative style that are relevant in the particular organisational context studied. As explained before, the final items were either suggested by Human resource managers or were based on underlying dimensions of evaluative style that have been identified as important and seemingly relevant dimensions in previous studies by Otley and Fakiolas (2000). Yet, a specific underlying structure cannot be identified a priori

⁴⁷ This is true for 14 of the 17 items.

since it is unclear whether the dimensions that Otley and Fakiolas (2000) have identified are related to each other, and if so, in what way. Some of these dimensions may be related in some contexts, but not in others. For example, while in some contexts there may be a conflict between long-term organisational goals and short-term (budgetary) targets, in which Hopwood's distinction between a budget-constrained style and a profit-conscious style would be relevant, in other contexts such a conflict may not exist, or to a lesser extent (cf. Otley, 1978). The implication is that despite Otley and Fakiolas' (2000) identification of, for example, a short-run versus a long-run emphasis as a separate theoretical dimension, empirically there is no reason to expect a correlation between short-run and long-term emphasis. Statistically, identification of a short-run versus a long-run emphasis as one separate dimension would require the two aspects to be correlated to each other, and, additionally, that each of these two aspects is uncorrelated to other aspects. To complicate matters even further, Otley and Fakiolas (2000) explicitly state that the short-run emphasis versus a long-run emphasis may be a sub-dimension of a rigid versus a more flexible use of budgets⁴⁸. Statistically, this statement can be transposed to stating that a short-run emphasis may be correlated to a rigid use of budgets, but not to a flexible use of budgets, while on the other hand, a long-run emphasis may be correlated with a flexible use of budgets, but not with a rigid use of budgets.

Again, such a correlation may be found in some organisational contexts, but not in others. Similar problems exist with the other dimensions as identified by Otley and Fakiolas (2000). In short, then, I conclude that the underlying structure of the seventeen items that I used to measure evaluative style cannot be specified a priori, but has to be explored empirically. Therefore, there are no specific propositions with regard to expected underlying dimensions of evaluative style as measured by the 17 items.

Yet, it is important to note that some of the propositions that I have developed in chapter 4 implicitly assume two important general dimensions of evaluative style. These two general dimensions are (1) an emphasis on quantitative measures of performance and deviations from targets, and (2) an emphasis on qualitative, interpersonal aspects of performance, and a willingness of superiors to listen to subordinates' explanations of their performance. Therefore, before empirically exploring the structure of the 17 items, I first classified the 17 items as theoretically pertaining to one of these two general dimensions. To classify the items, two raters independently indicated whether the items should be classified as belonging to one of the two categories or should be classified as unspecified. The results of both ratings were compared, and when similar the item was classified as both raters indicated. If the two raters disagreed, the item was classified as "unspecified". The resulting classification of the items is displayed in table 7.3.

⁴⁸ Or any other quantitative targets than financial budgets, as in Otley & Fakiolas' (2000) third dimension.

Table 7.3 A classification of the 17 items to measure perceived evaluative style as pertaining to quantitative or qualitative aspects of performance

Quantitative/deviations from targets	interpersonal, qualitative	Unspecified
1. short-term goals 3. financial information 5. deviations from agreed performance 7. objectively measurable performance 13. performance of the unit	6. explanation of performance 10. plans for next year 11. positive aspects of performance 15. causes of failings in performance 16. possibilities to improve performance 17. information from conversation	2. long-term goals 4. non-financial information 8. intuition 9. performance in past year ¹ 12. negative aspects of performance ¹ 14. personal, individual performance

¹ Classified as unspecified because the two raters disagreed.

This classification, that is based on the two general dimensions of evaluative style identified and discussed on the basis of the literature reviewed in chapter 4, will be used in chapter 8 to tailor the propositions that refer to the two general dimensions to the specific dimensions of evaluative style that emerge from the empirical analysis below.

7.4.2 Data analysis technique

To find the underlying structure of the seventeen items used to measure perceived evaluative style, exploratory factor analysis was done using principal component factor analysis after varimax rotation. Items of which the difference between factor loadings on two factors was less than .20 were discarded. An important decision to be made when conducting a factor analysis is how many factors should be extracted. In this study, this decision was based on a combination of different criteria. First, the root one criterion states that factors with eigenvalues greater than 1 should be retained. This is a very general and broadly applied rule that reduces the number of factors considerably. This criterion, however, has the risk of extracting factors that have incurred by chance, rather than reflect a more inherent structure among the items. Thus, even samples of items that are completely independent will produce a number of eigenvalues greater than 1 in a principal components analysis. Buja and Eyuboglu (1992) discuss parallel analysis (Horn Horn), 1965; Zwick and Velicer), 1986) as a method to overcome this chance capitalisation. In parallel analysis, the criterion of eigenvalues greater than one is replaced with the criterion that eigenvalues should exceed the mean eigenvalues generated from independent normal variates⁴⁹. As an extension and modification of parallel analysis, Buja and Eyuboglu (1992) suggest replacement of normal pseudorandom deviates with random permutations of the observed data. Random permutation implies that the scores of respondents for each item are randomly

⁴⁹ Normal variates refer to simulated (pseudorandom) samples that are based on the normal probability distribution.

reordered across the respondents (i.e., assigned to different respondents), which leaves the distribution (mean, median, standard deviation) of each item unchanged. By conducting principal component analyses on a number of permuted data sets, it is possible to generate "conditional null distributions" against which the actual observed eigenvalues can be compared. Additionally, the distributions can be used to calculate conditional significance levels and p-values. Therefore, as a second criterion, the observed eigenvalues were compared with conditional values that were obtained from a permutation test⁵⁰. As a third criterion, I used the theoretical interpretability of the factors.

After identifying the number of components to retain, for each component I created an empirical scale by summing the scores on all items that were part of that component. Cronbach's alphas, average inter-item correlations and corrected item-rest correlations were calculated to test the internal consistency of the empirical scales. The criteria were that Cronbach's alpha should be $>.70$, and item-rest correlations and average inter-item correlation should be $>.30$.

Because the results of factor analysis are sensitive to the selection of variables, inclusion or deletion of some of the 17 items may produce different components (Kim and Mueller, 1978a). Because the results of factor analysis are sensitive to the selection of variables, inclusion or deletion of some of the 17 items may produce different components (Kim & Mueller, 1978a). For this reason, finally, I performed a categorical principal components analysis (CATPCA) on the items that were part of the components that I retained. By treating all items as numerical, the results of the CATPCA are the same as from a standard principal components analysis. The main reason to use CATPCA is the advantage of the graphical display of both objects (respondents), variables and groups. This enables a graphical display of differences in evaluative style between leaders and function groups in subsequent analyses (see discussion section in § 7.6).

7.4.3 Results

The initial solution from the factor analysis revealed five factors with eigenvalues > 1.00 , which are reported in Table 7.4. Three items were dropped because the difference in factor loadings on two factors was smaller than $.20$. These items were "objectively measurable performance", "non-financial information", and "deviation from agreed performance".

The results of the permutation test are displayed in Figure 7.1. At a significance level of 5%, only one factor has an eigenvalue that is significantly higher than the eigenvalue obtained from randomly permuted data. The second factor has an eigenvalue that lies at the highest border of the 95% reliability interval of the permuted data. Therefore, this factor, using a slightly more liberal criterion than 5% significance, may be retained. This is

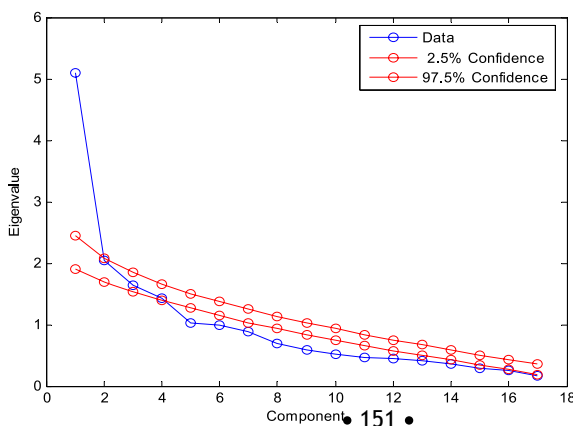
⁵⁰ I am grateful to thank Patrick Groenen for this suggestion and for running the permutation test on the 17 items.

Table 7.4. Principal Component Analysis after Varimax rotation of evaluative style items (n=57)

	Component				
	1	2	3	4	5
your explanation of your performance	.771	.303	-.145		.209
positive aspects of your performance	.690	.371	.309		.163
personal, individual performance	.688	.240	.459		-.253
performance delivered in the past year	.682	-.143	.144	.333	
performances of your unit	.602	.168			
objectively measurable performance ^a	.595	.345			.402
non-financial information ^a	.520	.136	.463	-.142	
possibilities to improve your performance		.811		.341	
information from conversation	.262	.741	-.102	-.133	
your plans for the next year	.160	.733		-.101	
Causes of failings in performance	.223	.656	.225	.209	.225
Negative aspects of your performance ^b			.769	.226	
his/her own intuition ^b	.209		.680	.207	
deviations from agreed performance ^a		.234	.520	-.286	.515
short-term goals ^b			-.130	-.826	.150
long-term goals ^b	.121	.173	.136	.690	
Financial information ^b					.918
Initial eigenvalues (before Varimax rotation)	5.107	2.001	1.584	1.418	1.031
Percentage of variance explained	30.040	11.768	9.317	8.343	6.065
Cumulative percentage of variance explained	30.040	41.807	51.125	59.468	65.532

^a Dropped from final solution because of high loadings on more than one factor
^b Dropped from final solution based on permutation test

defendable in my opinion, because permutation tests provide a rather strict test, while my analysis is mainly exploratory. For this reason I decided to see if the first two factors would be theoretically interpretable. For all other factors, the permutation test clearly illustrates that these factors cannot be separated from factors obtained by chance. Therefore, the last three factors displayed in table 7.4 were dropped from the final analysis.

Figure 7.1 Permutation test applied to principal components analysis

A close look at the five items that load high on the first factor (table 7.4) indicates that all five items refer to aspects of performance, either the performance of the respondent or the performance of his or her unit. Taken together, the items suggest that this factor emphasises the extent to which performance (in the past/ of the past year) is being assessed, based on a broad array of information (personal performance, unit performance, explanation of performance). I decided to label this factor as "Performance Assessment".

The four items that have high factor loadings on the second factor seem to emphasise the extent to which superiors use performance evaluation as a learning tool for the future. Three of the four items in this factor were actually included to capture performance development, based on interviews with the Human Resource managers and based on characteristics of the PDP system. Thus, theoretically, the second factor seems to capture a relevant dimension of performance evaluation within VDBN, and the factor can theoretically be explained. In addition, it theoretically seems to capture a different dimension of evaluative style than the first dimension. Therefore, I decided to retain the second factor, despite the results of the permutation test, and I have labelled this factor as "Performance Development".

To test the internal reliability of the first two factors, Cronbach's alpha's, inter-item correlations and item-total correlations were calculated. The results are reported in Table 7.4. For both dimensions, Cronbach's alpha, inter-item correlations and item-total correlations met the criteria. Therefore, for each dimension, the sum of the Likert scores of the items was used to calculate the total score on that dimension.

Cross-validation of interpretation of dimensions of evaluative style

To cross-validate the interpretation given to the two dimensions of evaluative style, I compared the two dimensions with the five items from the "interpersonal evaluation" measure that have been used earlier by . These five items were:

1. My supervisor is thoroughly familiar with my job performance.
2. My supervisor allows me to tell my side of the story in performance evaluation.
3. My supervisor lets me know my appraisal outcomes and provides justification.
4. My supervisor lets me know my pay raise and annual bonuses and provides justification.
5. My supervisor reviews my performance with me and discusses plans or objectives to improve my performance.

Since I labelled the first factor performance assessment, I would expect a higher correlation between the first factor and the first item of interpersonal evaluation than on the other items, because the supervisor's knowledge of the job performance would be an important aspect of performance assessment. Furthermore, I have argued in a previous chapter that within VDBN a distinction can be made between performance assessment and performance development. Determining the bonus would be part of performance assessment rather than performance development. Therefore, I would also expect a higher correlation between the first factor and the fourth item than on items 2, 3 and 5. I have defined the second factor as

Table 7.5. *Evaluative Style Subscales: Results from Factor Analysis (n = 57)*

Factor 1: Performance assessment (Assessment of past performance)	factor loading	item-rest correlation
your explanation of your performance	.77	.60
Positive aspects of your performance	.69	.67
personal, individual performance	.69	.66
performance delivered in the past year	.68	.51
performances of your unit	.60	.49
5 items, Cronbach alpha = .80 average inter-item correlation .45 mean = 3.56 sd = .63		
Factor 2: Performance development (Evaluation as a learning tool)	factor loading	item-rest correlation
possibilities to improve your performance	.81	.67
information from conversation	.74	.57
your plans for the next year	.73	.52
Causes of failings in performance	.66	.57
4 items, Cronbach alpha = .78 average inter-item correlation .47 mean = 3.22 sd = .68		

performance development, and thus, I would expect a higher correlation between the second factor and the fifth item, which clearly refers to performance development, than on any of the other items. Correlations between the two dimensions of evaluative style derived from factor analysis and the five items measuring "interpersonal evaluation" were calculated. Table 7.6 shows the results. The findings confirm the interpretation of the two dimensions.

Table 7.6 *Kendall's tau correlations between two dimensions of evaluative style and five items comprising the interpersonal evaluation measure (n=57)*

	Assess	Develop
1. My supervisor is thoroughly familiar with my job performance.	.494**	.371**
2. My supervisor allows me to tell my side of the story in performance evaluation.	.306**	.324**
3. My supervisor lets me know my appraisal outcomes and provides justification.	.324**	.287*
4. My supervisor lets me know my pay raise and annual bonuses and provides justification.	.405**	.304**
5. My supervisor reviews my performance with me and discusses plans or objectives to improve my performance.	.236*	.475**
* Correlation is significant at the .05 level		
** Correlation is significant at the .01 level		

Graphical display of dimensions

As a final check of the distinction between the two dimensions of evaluative style, Performance Assessment and Performance Development, I conducted a principal component analysis on the 9 items of the two subscales, using CATPCA, choosing to display the two-dimensional solution. Table 7.7 displays the model summary. Table 7.8 contains the component loadings and figure 7.2 shows a plot of the component loadings.

Table 7.7 Model Summary of CATPCA

Dimension	Cronbach's Alpha	Variance Accounted For
		Total (Eigenvalue)
1	.841	3.955
2	.342	1.437
Total	.916 ¹	5.392

¹ Total Cronbach's Alpha is based on the total Eigenvalue.

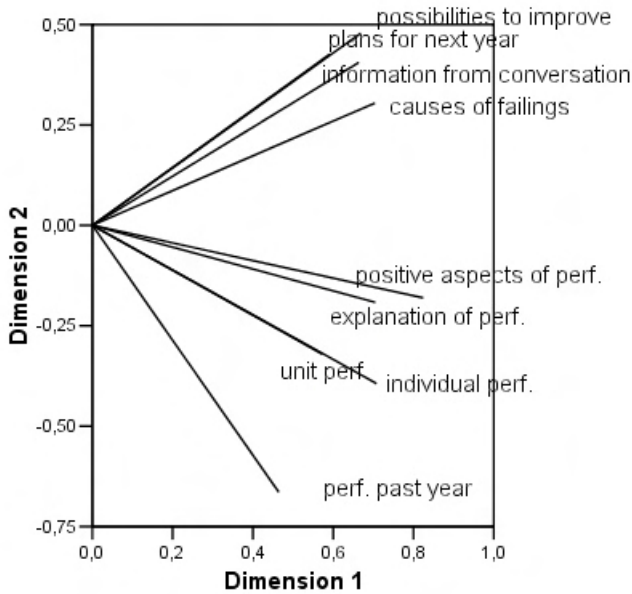
Table 7.8 Component Loadings

	Dimension	
	1	2
explanation of performance	.705	-.192
performance in past year	.464	-.665
plans for next year	.591	.426
positive aspects of your performance	.824	-.181
unit performance	.573	-.321
individual performance	.708	-.394
causes of failings	.704	.305
possibilities to improve	.669	.479
information from conversation	.664	.406

While all variables have high (positive) loadings on the first dimension, the second dimension discriminates the 4 items of the Development scale from the 5 items of the Assessment scale. Items of the Development scale load positively on the second dimension, items from the Assessment scale negatively. This indicates that objects with high positive (negative) loadings on the second dimension perceive the evaluative style of their superiors as relatively more (less) oriented towards Development than towards Assessment.

Objects with high (low) scores on dimension 1 will report relatively high (low) scores on both Development and Assessment. Together, the two dimensions discriminate between objects on the absolute importance of Development and Assessment (dimension 1) and the relative importance of Development as compared to the importance of Assessment (dimension 2). Overall, this analysis confirms the distinction between the two subscales.

Figure 7.2 Plot of component loadings



7.5 Evaluative style as a group level phenomenon

7.5.1 Introduction

In chapter 4 two propositions have been formulated with regard to evaluative style as a group level phenomenon, where groups are defined as subordinates reporting to the same superior. These two propositions are:

- P1. *There are significant between-group differences in perceived evaluative style (where groups are subordinates reporting to the same superior).*
- P2. *There is considerable agreement on perceived evaluative style within groups of subordinates reporting to the same superior.*

Because I measured both preferred and perceived evaluative style, these propositions have to be adapted. As to within-group agreement, proposition 2 states that I expect that there will be considerable within-group agreement with respect to the reported perceived evaluative style. This applies to the perceived importance scores on the 17 single items as well as to the perceived importance scores on the underlying dimensions of evaluative style. As explained in chapter 4 and supported by the interview data in § 7.2, I do not expect total

within-group agreement, since I have argued that the manner in which a leader evaluates an individual subordinate is likely to depend on individual characteristics as well, such as age, experience, and tenure of the subordinate (this expectation will be explored in chapter 8).

Proposition 1 explores whether there are differences in how leaders evaluate their subordinates. Similar to proposition 2, this proposition too will be explored for the perceived importance scores of the 17 single items, and of the underlying dimensions.

Finally, since preferred evaluative style is supposed to reflect norms and values held across the organisation, I expect less disagreement among subordinates on the importance scores of items measuring preferred evaluative style than on items measuring perceived evaluative style, both within- and between-groups.

This leads to the following adapted propositions:

P1a. *There are significant differences between groups in average perceived importance of the items and dimensions of evaluative style.*

P1b. *There are no significant differences between groups in average preferred importance of the items of evaluative style.*

P2a. *There is considerable agreement in the importance scores on the items and dimensions of perceived evaluative style within groups of subordinates reporting to the same superior.*

P2b. *Within-groups agreement on the preferred importance of aspects of evaluative style (preferred evaluative style) will be higher than within-groups agreement on the perceived importance of these aspects (perceived evaluative style).*

7.5.2 Data analysis technique

To test whether there were significant differences between leaders in mean scores on the items of perceived evaluative style (P1a) and on the items of preferred evaluative style (P1b) Kruskal-Wallis tests were used. To test the extent to which subordinates within groups agreed in their ratings of the perceived and preferred importance of aspects of evaluative style (P1a and P1b), I used the r_{wg} within-group agreement coefficient⁵¹ as developed by James et al. (1984). The logic underlying this measure is that

⁵¹ There has been a controversy in psychological and organizational literature on the labeling of within-group agreement indexes. James et al. (1984) presented their index as a measure of within-group interrater reliability, but it is more appropriate to label this index as a measure of within-group interrater agreement (Kozlowski and Hattrup, 1992; James et al., 1993). Reliability is concerned with consistency in ratings among subordinates reporting to the same leader, while agreement is concerned with consensus in ratings among subordinates. For our purpose, i.e. to assess whether evaluative style should be treated as a group-level or an individual level phenomenon and whether individual respondent scores can be aggregated to the leader level, the relevant issue is whether there is within-group agreement on the perceived evaluative style. This is measured by the within-group agreement coefficient (cf. Kozlowski and Hattrup, 1992).

"before scores that are obtained from individuals can be meaningfully aggregated and assigned to a higher level of analysis (e.g., as indicative of a work group or unit), agreement must be shown among those individuals in terms of the attitude or behavior that they have described." (Schriesheim et al., 1995, p. 114)

Details on the calculation of the $r_{wg(1)}$ within-group agreement coefficient are provided in appendix F. Using the formulas described in appendix F, r_{wg} will be calculated for each of the twelve groups of subordinates reporting to the same superior. Values of within-group agreement coefficients less than .000 or greater than 1.000 for any leader will be replaced with values of .000, as suggested by James et al. (1984).

One of the problems with r_{wg} is that it is hard to assess whether the reported values are significantly different from zero. As a rule of thumb, George (1990) suggests that values above .7 are necessary to demonstrate consistency in ratings within a group. However, application of the .70 criterion is not straightforward, as Schriesheim et al. (1995, p. 127) point out:

"... r_{wg} coefficients can be well above .7 and still not be statistically significant. Additionally, James et al. (1984) note that the number of "judges" (group members) may exert a substantial effect on obtained r_{wg} coefficients. However, <our results> show that the number of "targets" (items in a scale) also seems to substantially affect the size of the r_{wg} coefficient (and its statistical significance)."

Since single items are more influenced by measurement error than a composite measure of "essentially parallel" indicators, James et al. (1993) also state that generally $r_{wg(j)}$ will be larger than $r_{wg(1)}$, i.e. within-group agreement coefficients will be higher for scales than for single items. Therefore, within-group agreement coefficients will be calculated for the 17 single items, and for the underlying dimensions.

7.5.3 Results

Single aspects

The results of the Kruskal-Wallis tests for differences in medians among subordinates grouped by leader (P1a and P1b) are reported in table 7.9. For the 17 items measuring perceived evaluative style, the results indicate that there are significant differences in medians between subordinates grouped by leader with regard to three items of evaluative style: i.e. "short-term goals" ($\chi^2 = 24.678$, $p = .010$), "personal, individual performance" ($\chi^2 = 21.749$, $p = .026$), and "possibilities to improve your performance" ($\chi^2 = 20.561$, $p = .038$). In contrast, for preferred evaluative style, a significant difference in medians between leaders on only one item, "performance in the year past" ($\chi^2 = 21.840$, $p = .026$), was found. These results

Table 7.9. Differences in Medians between Groups (Kruskal Wallis-test) and Agreement within Groups (r_{wg}) on 17 aspects of evaluative style ($n = 12$)

	Perceived evaluative style				Preferred evaluative style			
	χ^2	p	r_{wg}		χ^2	p	r_{wg}	
			Mean	Median			Mean	Median
short-term goals	24.678	.010	.728	.850	12.716	.312	.799	.817
long-term goals	13.111	.286	.650	.750	5.501	.905	.758	.850
Financial information	13.577	.257	.642	.658	12.126	.354	.785	.842
non-financial information	9.117	.611	.768	.743	6.144	.864	.837	.850
Deviations from agreed performance	16.856	.112	.734	.761	6.383	.847	.709	.764
your explanation of your performance	11.837	.376	.640	.700	7.767	.734	.716	.750
objectively measurable performance	14.724	.195	.718	.750	11.187	.428	.802	.795
his/her own intuition	17.517	.093	.768	.753	9.069	.616	.616	.675
performance in the year past	14.465	.208	.676	.720	21.840	.026	.817	.808
your plans for the next year	13.014	.292	.729	.750	13.873	.240	.671	.811
Positive aspects of your performance	17.123	.104	.830	.854	13.711	.249	.781	.792
Negative aspects of your performance	11.366	.413	.594	.739	7.886	.723	.708	.750
Performances of your unit	14.214	.221	.514	.642	11.786	.380	.798	.863
Personal, individual performance	21.749	.026	.775	.792	13.701	.250	.769	.850
Causes of failings in performance	18.876	.063	.695	.750	18.669	.067	.813	.842
possibilities to improve your performance	20.561	.038	.615	.817	5.701	.893	.753	.792
information from conversation	16.46	.125	.710	.708	6.334	.850	.767	.750

are not totally consistent with my expectations, since I expected that scores on preferred evaluative style would not be affected by group-membership, while scores on perceived evaluative style would. In general, however, the evidence in Table 7.9 supports the expectation that group-membership is more likely to affect the reported scores on aspects of perceived evaluative style than reported scores on aspects of preferred evaluative style. With the exception of one item, "performance in the year past", p-values for the items of preferred evaluative style are much higher than for perceived evaluative style, indicating that it is more likely that the assumption of equal medians across different leaders holds for preferred evaluative style items than for perceived evaluative style items. Thus, although I found limited support for my expectations in propositions 1a and 1b, the results do support my expectation that scores on perceived evaluative style items would show more variation across leaders than scores on preferred evaluative style items.

Within-group agreement on single aspects of evaluative style

As an indication of the extent to which subordinates within groups agree in their ratings of perceived and preferred evaluative style (P2a and P2b), the mean and median value of r_{wg} for each item are reported in Table 7.9. The data in Table 7.9 reveal that all but two median within-group agreement coefficients for the 17 aspects of perceived evaluative style are above the .7 criterion. However, only 9 of the 17 mean within-group agreement coefficients are greater than .7. More specifically, of the 204⁵² within-group estimates, (a) 121 are equal to or greater than .7, (b) 85 are equal to or greater than .8, and (c) 30 are equal to or greater than .9. These results raise concern about the appropriateness of aggregating individual responses on the 17 single aspects of individual perceived evaluative style to measure perceived evaluative style at the group-level.

For the 17 aspects of preferred evaluative style, as displayed in the right-hand side of Table 7.9, only three of all median and mean within-group agreement coefficients are lower than .7. Furthermore, of all 204 within-group agreement coefficients, (a) 157 are equal to or greater than .7, (b) 107 are equal to or greater than .8, and (c) 48 are equal to or greater than .9. A comparison of the within-group agreement coefficients for perceived and preferred evaluative style reveals that 12 mean and 13 median coefficients are higher for preferred than for perceived evaluative style, and that 147 of the 204 within-group agreement coefficients for preferred evaluative style are equal to or higher than perceived evaluative style. Wilcoxon Signed Ranks Tests reveal that the median and mean within-group

⁵² Each mean or median within-group agreement coefficient reported in Table 7.9 requires the calculation of 12 within-group agreement coefficients, one for each group of subordinates reporting to the same leader. Thus, in total, $12 * 17 = 204$ within-group agreement coefficients were calculated.

agreement coefficients are indeed significantly higher for preferred evaluative style than for perceived evaluative style⁵³. These results support proposition 2b.

Within-group agreement on dimensions of evaluative style

For perceived evaluative style, similar analyses as described above for the 17 single aspects of evaluative style were conducted for the two underlying dimensions of evaluative style found through factor analysis, i.e. performance assessment and performance development, and for interpersonal evaluation. The results of the Kruskal-Wallis tests for differences in medians between groups of subordinates are shown in Table 7.10.

Table 7.10 Differences in Medians between Groups (Kruskal Wallis-test) and Agreement within Groups (r_{wg}) on 3 dimensions of perceived evaluative style (n = 12)

	χ^2	P	r_{wg} Mean	Median
Performance Assessment	15.771	.150	.874	.900
Performance Development	18.183	.077	.872	.914
Interpersonal	11.159	.430	.840	.903

The results indicate that there are no significant differences in medians between subordinates grouped by leader at the 5% level. At the 10% level, there are significant differences in medians between leaders for Performance Development ($\chi^2 = 18.183$, $p = .077$). As to the within-group agreement coefficients, the data in Table 7.10 reveal that all median and mean within-group agreement coefficients are above the .7 criterion, and even well above .8. For all three dimensions the within-group agreement coefficients seem to support the appropriateness of aggregating the data to a group-level measure of perceived evaluative style. This is confirmed by an analysis of the 36 within-group agreement coefficients: (a) 31 of these estimates are equal to or greater than .7, (b) 29 are equal to or greater than .8, and (c) 18 are equal to or greater than .9. A further investigation of the within-group agreement coefficients indicates that for two of the three dimensions of evaluative style the within-group agreement coefficients of leader 1 are below the .7 criterion. Therefore, leader 1 is excluded from further analyses at the group-level (see chapter 8).

Graphical display of between-group differences

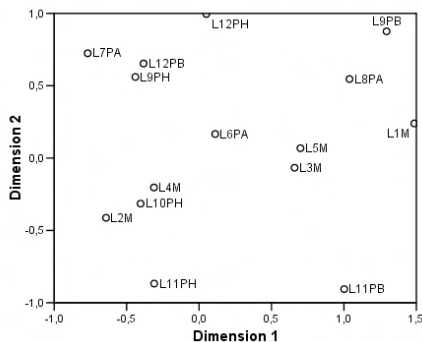
To display differences in evaluative style at the leader-level graphically, I used the results from the CATPCA on the items of the two empirical dimensions. First, however, I created an additional variable that combined function groups and leaders (*leagroup*). This was done for two reasons. First, some leaders in production (9, 10, 11, and 12), as discussed earlier, had

⁵³ For the difference in medians: $z = -2.296$, $p = .022$ and for the difference in means: $z = -2.343$, $p = .019$ (n=17)

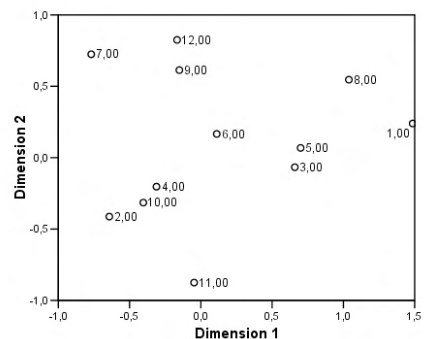
respondents from different function groups, i.e. managers and production heads. Second, the analyses in chapter 6 showed that characteristics of respondents such as age, tenure and education differed significantly between function groups. Thus, the difference in function groups in addition to differences between leaders seemed to be relevant. Since I am interested in the role of leaders, it is important to take into account possible influences from differences in function groups. The new variable differed from the variable *leader* because for each of the leaders 9, 11 and 12 two categories were used, one for respondents who were managers, and the other for respondents who were heads of production. There was no need to split respondents of leader 10, since none of the managers reporting to leader 10 returned the questionnaire. Using this new variable, the average scores from respondents of a single leader, split by function group, were plotted in the two-dimensional space as defined by the results of the CATPCA that was reported in § 7.4.3. This was achieved by adding *leader* and *leagroup* as supplementary variables into the previous CATPCA. The resulting plot is shown in figure 7.3, panel A. This figure shows that there seem to be differences between leaders and between function groups. Before analysing these differences, it is important to note that the plot for leaders, without splitting production heads and production managers, is almost similar (see figure 7.3, panel B). This is caused by the fact that for all three leaders the number of respondents in managerial positions is relatively small in comparison to the number of respondents who were production heads.

Figure 7.3 Category plots of leader-function group variable (panel A) and leaders (panel B)

Panel A. Categories of leader-function group



Panel B. Categories of leader



Marketing leaders

Marketing managers (L1M, L2M, L3M, L4M, and L5M) seem to differ mostly on the first dimension, and only little on the second. This suggests that there seems to be a difference in the absolute emphasis that marketing leaders place on Performance Assessment and Performance Development. Leader 1 emphasises both aspects the most, leader 2 and leader 4 the least, and leader 3 and 5 are in the middle.

Production leaders A

In comparison to marketing managers, except for leader 11 (L11PB), managers who report to a production leader (L6PA, 7PA, 8PA, 9PB, 12PB) on average have higher scores on the second dimension. This suggests that the relative emphasis on Performance Development as compared to Performance Assessment is higher for production leaders than for marketing managers. As in marketing, within the group of leaders who only evaluate managers and no production heads, i.e. leaders 6, 7 and 8 (L6PA-L8PA) there are differences between leaders. Leaders 6, 7 and 8 differ considerably in the extent to which they emphasise Performance Assessment and Performance Development (dimension 1). Leader 8 emphasises both aspects the most, leader 7 the least, and leader 6 is in between. While leader 6 puts almost equal emphasis on Performance Development and Performance Assessment, leaders 7 and 8 put relatively more emphasis on Performance Development than on Performance Assessment (dimension 2).

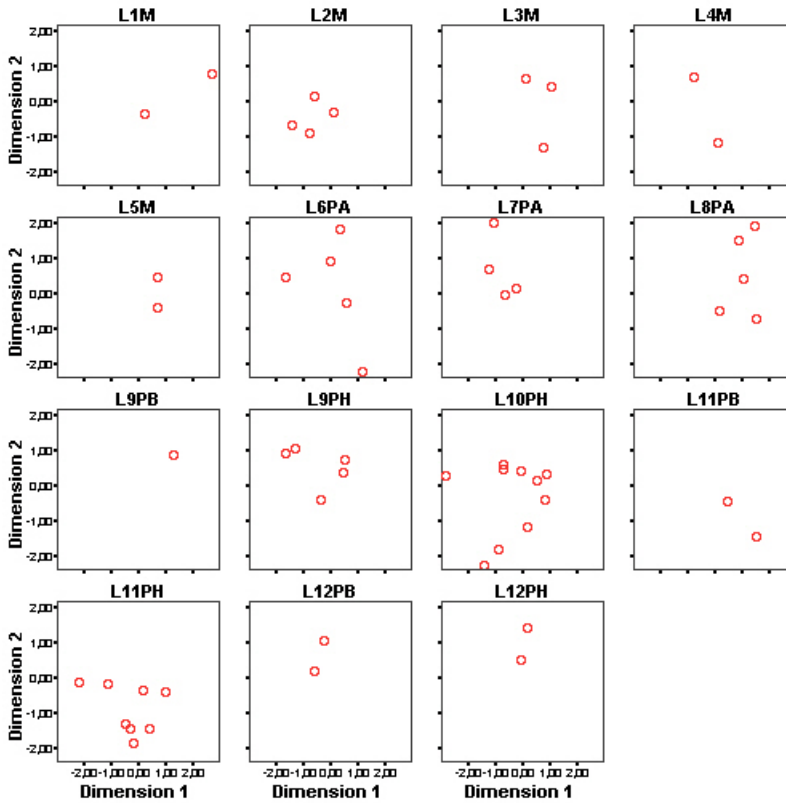
Production leaders B

The other leaders, i.e. leader 9, 10, 11 and 12 evaluate both managers and production heads, although as explained for leader 10 none of the respondents were managers. As to the evaluation of managers, leader 9 (L9PB) differs from leader 11 (L11PB) on dimension 2 only, and from leader 12 (L12PB) on dimension 1. Leaders 11 and 12 differ from one another on both dimensions. Production heads who report to leader 9 and leader 11 mainly report different scores from managers on the first dimension. For both leaders, production heads on average report a lower emphasis on Performance Assessment and Performance Development. In contrast, for leader 12, production heads report slightly higher scores than managers on dimension 1 and dimension 2. Within the group of production heads, on average, all production heads, irrespective of the leader, report similar scores on dimension 1. Production heads reporting to leaders 9 and 12 report higher scores on dimension 2 than leaders 10 and 11.

Graphical display of within-group differences

The previous CATPCA not only enables plotting leaders and function groups in the two-dimensional space as defined by the analysis of evaluative style items, but it also enables plotting individual respondents into this same space. In CATPCA, for each respondent, called an object in CATPCA, object scores on the dimensions are calculated. These object scores are the coordinates that enable the graphical display of the objects in the two-dimensional space. This is especially useful in analysing within-group differences, because the object scores for each respondent indicate how similar objects are to each other. Thus, plotting the object scores for respondents in each group gives a graphical display of within-group differences. The graphical display is shown in figure 7.4.

Figure 7.4 Object scores on two dimensions by leader-function group



The figure merely illustrates the results from previous analyses in this chapter: there appear to be some differences between leaders, between function groups, and within-groups on evaluative style, although not all differences reach statistical significance.

7.6 Discussion and conclusion

In this chapter I explored two topics: the conceptualisation and measurement of perceived evaluative style within the context of VDBN, and between- and within-group agreement in evaluative style. Table 7.11 provides an overview of the main findings in this chapter. The table displays the research questions and propositions on these two topics that were tested in this chapter and presents the outcomes. The findings will be discussed briefly below.

Table 7. 11 *Research questions and propositions addressed in this chapter and the outcomes*

<i>Research questions and propositions</i>		<i>Findings</i>
1	Purpose: gain a better understanding of the construct of evaluative style	
1.1 Conceptualisation and measurement of perceived evaluative style		
<ul style="list-style-type: none">What are meaningful underlying dimensions of perceived evaluative style in this particular organisation?		<p>Three general dimensions from interviews:</p> <ul style="list-style-type: none">- The importance that superiors attach to targets;- The coaching role of superiors;- The use/ implementation of PDP.
<ul style="list-style-type: none">What is the role of (financial) accounting information in this particular organisation?		<p>Three dimensions (subscales) from written questionnaire:</p> <ul style="list-style-type: none">- Performance assessment: assessment of past performance- Performance development: evaluation as a learning tool- Interpersonal evaluation <p>Hardly any emphasis on budgets for performance evaluation (see also § 6.4)</p>
1.2 Between- and within-group agreement in evaluative style		
<p>P1. There are significant between-group differences in evaluative style (where groups are subordinates reporting to the same superior).</p>		<p>This proposition was replaced by P1a and P1b.</p>
<p>P1a. There are significant differences between groups in average perceived importance of the items and dimensions of evaluative style.</p>		<p>Not supported: significant differences found for only three of 17 items and none of the three subscales</p>
<p>P1b. There are no significant differences between groups in average preferred importance of the items of evaluative style.</p>		<p>Supported, although significant differences found for one of 17 items</p>
<p>P2. There is considerable agreement on perceived evaluative style within groups of subordinates reporting to the same superior.</p>		<p>This proposition was replaced by P1a and P1b.</p>

- P2a. There is considerable agreement in the importance scores on the items and dimensions of perceived evaluative style within groups of subordinates reporting to the same superior.
- P2b. Within-groups agreement on the preferred importance of aspects of evaluative style (preferred evaluative style) will be higher than within-groups agreement on the perceived importance of these aspects (perceived evaluative style).
- Partly supported:** median $rwg > .7$ for 15 of 17 items and all three subscales; mean $rwg > .7$ for 9 of 17 items and all three subscales
- Supported:** 12 of 17 median and 13 of 17 mean rwg coefficients are larger for preferred than perceived evaluative style

Meaningful underlying dimensions of evaluative style within VDBN (1.1a)

The interview data suggest that there appear to be differences in evaluative style within VDBN on three dimensions: the importance that superiors attach to targets, the coaching role of leaders, and the use/ implementation of PDP. These three dimensions are somewhat different from those theoretically expected. For example, hardly any evidence is found in the interview data that supports the importance of the dimensions identified in RAPM-literature⁵⁴. Some differences exist in the relative importance that leaders attach to targets (results) as opposed to processes (how results are realised). This difference relates to one of the dimensions of performance evaluation style as identified in the literature, i.e. a rigid use of quantitative targets versus a more flexible use (Otley and Fakiolas, 2000). It is different from this dimension however because there are no signs of rigid use, or any pressure created by performance evaluation.

The written questionnaire also reveals some underlying dimensions of evaluative style. In the written questionnaire, I have measured 17 aspects of perceived evaluative style that were thought to be important in exploring and understanding differences in how superiors evaluate the performance of their subordinates. The results (not reported in table 7.11) suggest that the 17 aspects I measured were indeed important aspects of performance evaluation, since I also asked subordinates to indicate how much emphasis they think each of these aspects should be given by their superior when evaluating their performance, i.e. preferred evaluative style. On average, subordinates indicate that they feel that almost every aspect that is measured deserved more attention than it presently received. Exceptions are short-term goals and boss's intuition. The most preferred instruments were paying attention to possibilities to improve/develop your performance (score 4.28 on a 5 point scale), long-term goals (4.16) and performance of your unit (4.02), the least preferred instruments were the boss's intuition (3.35), financial information (3.42), and short-term goals (3.61). With respect to the role of financial information, this suggests that the motivational and appraisal role of budgetary targets and budgetary control in this organisation may be limited, which is in line with my conclusion from the analysis of the interview data (see also the section on the role of accounting information below).

A further analysis of the 17 items, revealed a number of underlying dimensions, of which two dimensions seemed to be theoretically relevant and interpretable. These dimensions have been described as:

1. assessing past performance (pastperf)
2. using performance evaluation as a learning tool (learn)

⁵⁴ These dimensions have been described and discussed in more detail in Chapter 3.

Because I distributed the written questionnaire at the same time that I conducted the interviews, it is hardly possible to compare the dimensions from the interviews with the dimensions of the written questionnaire. A suggestion for future research would be to develop a written questionnaire after analysing the interview data, to try to capture some of the dimensions that emerge from the interviews more directly in the written questionnaire. Although no formal attempt has been made to compare the findings from both sources at the level of leaders, the dimensions that emerge from the interview data and the dimensions from the questionnaire data both seem to be relevant and meaningful to be used in further analyses in chapter 8 as to the behavioural implications of these differences.

Role of accounting information (1.1b)

Although the role of budgets has been very important to distinguish performance evaluation styles in the existing (RAPM-)literature, none of the leaders seem to attach much importance to budgets for performance evaluation purposes. This could be due to characteristics that are typical for the organisation studied, such as the introduction of a new performance development system, or the organisational culture. Another explanation, supported by recent findings in the UK reported by Marginson (1999), is that the role of budgets in contemporary organisations is changing, and that the role of non-financial measures is becoming more and more important. If this is the case, the findings in this study would not be limited to this single organisation, but could be part of a broader trend. A final explanation for the limited role of budgets for performance evaluation purposes could be the sampling of respondents. As I described in Chapter 6, the subordinate respondents were managers at the lowest management-level within VDBN, and some of the respondents were not even managers at all. Many of these respondents had only limited responsibility and limited managerial tasks. In terms of responsibility accounting, these respondents were not heads of financial responsibility centres, and, therefore, bore no financial responsibility.

Between- and within-group agreement in evaluative style (1.2)

The interview data suggest that there appear to be differences in perceived evaluative style between leaders, but also between individuals that report to the same leader. These results are partly confirmed by the findings from the written questionnaire which do reveal differences in perceived evaluative style across groups of subordinates, although none of these differences is statistically significant. Possibly, in general, organisational level contextual factors could strongly influence perceived evaluative style within a single organisation, thereby reducing the variance. This will be especially true in organisations with a strong organisational culture, or, to be more precise, with strong cultural norms and values. This could be the case within VDBN, as there seems to be considerable consensus on preferred evaluative style across subordinates. The fact that this study was carried out at the time a new evaluative system had been launched recently, probably has contributed to the shared cultural norms and values. Finally, the role of the CEO in any organisation could be

important to consider as well. Merchant & Van der Stede (2003) for example argue that “tone at the top” is important for management control systems to be effective. Simons (1995), similarly, argues that one of the four necessary levers of control are belief systems: top management has to emphasise their vision in different ways within the organisation. Although the new evaluative system in VDBN was created by Unilever, the CEO of VDBN and the HR department actively stimulated and motivated managers and employees to use this system. Some interviewees explicitly indicated that the CEO was their “role model”. This “tone at the top” also could help explain the large amount of consensus within this organisation. Additionally, the findings also indicate that despite considerable within-group agreement, there are also considerable within-group differences between subordinates. As expected, in general, there is less agreement on perceived evaluative style than on preferred evaluative style. Together, these results indicate that differences in perceived evaluative style that exist at the level of individuals can be partly explained by differences in leaders to whom subordinates report. However, another part of the differences do occur at the level of individuals. Thus, in trying to understand the effectiveness of perceived evaluative style individual subordinates more than groups could be the most relevant level of analysis. In chapter 8 I will try to explain these differences between and within groups in more detail by studying the role of contextual factors at the level of groups and individuals.

8 Antecedents and consequences of perceived evaluative style: empirical evidence

8.1 Introduction

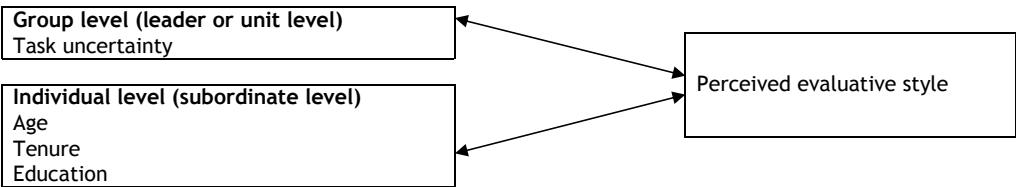
In this chapter the relationship between perceived evaluative style and other variables will be explored. I will address three topics. First, in § 8.2, I will explore whether between- and within-group differences in perceived evaluative style are associated with contextual variables as described in chapter 4. The analysis in this section is based on data from the written questionnaire and on interview data from superiors. Next, I will explore the relationship between leadership style and perceived evaluative style, by analysing data from the written questionnaire. This will be the subject of § 8.3. Finally, I will discuss the relationship of perceived evaluative style with subordinate managers' behaviour, with a focus on the intervening role of perceived fairness of evaluation and trust in superior. These relationships will be explored by analysing the questionnaire data (§ 8.4), and interview data, particularly the data obtained from the subordinates (§ 8.5). The chapter ends with a summary and conclusion.

8.2 Perceived evaluative style and context

8.2.1 Introduction

In chapter 4, I have developed a framework linking contextual factors and perceived evaluative style at different levels of analysis. Based on this framework, in this section I will analyse the relationships depicted in figure 8.1:

Figure 8.1 *The relationship between contextual variables and perceived evaluative style*



With respect to the relationship between task uncertainty and perceived evaluative style, the following proposition will be explored:

- P3. As task uncertainty increases, evaluative styles will be characterised by less emphasis on budgets and other quantitative targets, a higher emphasis on qualitative, interpersonal aspects of performance, and a higher willingness of superiors to listen to*

subordinates explanations of their performance (a more flexible use of budgets and other quantitative targets).

At the level of individual subordinates, the following propositions will be explored:

- P4. Within-group differences in age, tenure, and education of subordinates are related to within-group differences in perceived evaluative style.*
- P5. Within-group differences in perceived evaluative style partly reflect differences in evaluative style that the superior intended.*

8.2.2 Data analysis technique

Proposition 3

To test proposition 3, first I tested the association between task uncertainty and each of the 17 items of evaluative style, using the the classification of the 17 items as pertaining to quantitative or qualitative aspects of performance evaluation as developed in § 7.4.1.

Next I tested the association of task uncertainty with the three subscales of evaluative style: Performance Assessment, Performance Development and Interpersonal Evaluation. To help interpret the results, proposition 3 has to be tailored to each of the three subscales. The four items of the Performance Development scale are all classified as referring to interpersonal, qualitative aspects of performance. Thus, according to proposition 3, as task uncertainty increases we would expect a higher emphasis on Performance Development. The five items of Performance Assessment, in contrast, are spread over the different columns: one item is classified as referring to quantitative measures of performance and deviations from targets (i.e., performance of the unit), others as interpersonal, qualitative aspects of performance (i.e., explanation of performance, and positive aspects of performance), and the remaining two items as unspecified. As task uncertainty increases I expect a lower emphasis on quantitative aspects but a higher emphasis on qualitative, interpersonal aspects. Therefore, it is impossible to specify whether to expect more or less emphasis on Performance Assessment as task uncertainty increases. Interpersonal evaluation was measured by a separate instrument. Since it measures interpersonal aspects of evaluation, I would expect that as task uncertainty increases the emphasis on Interpersonal Evaluation increases as well.

Because task uncertainty operates at the level of tasks, and each group of subordinates in this study perform similar tasks, task uncertainty is treated as a group-level variable. Therefore, for each group of subordinates, aggregated scores on task uncertainty and evaluative style were calculated. As explained in chapter 7, leader 1 was excluded from the analyses due to the relatively low within-group agreement coefficient (see § 7.5.3). Responses have been aggregated by leaders and by the combination of leader-function groups. The latter aggregation contains three extra groups, as it splits respondents of leader

9, 11, and 12 into managers and production heads (see § 6.5). Since managers and production heads do not perform similar tasks, the aggregation by combining leaders and function groups into 14 groups seems to be the most pure level of aggregation. It is important to realise, however, that aggregation of responses for individual items is problematic in this study as I have shown in the previous chapter (see § 7.5.3). This aggregation is less problematic for the three subscales of evaluative style, Performance Assessment, Performance Development and Interpersonal Evaluation.

To test the association at the group level of task uncertainty with the 17 items and three subscales of perceived evaluative style I used Kendall's tau correlation, a non-parametric measure of correlation between two variables.

Proposition 4

To test proposition 4, first within-group differences in perceived evaluative style were calculated for the variables age, time with organisation, time with leader, time in function, number of evaluations, education, and for the three evaluative style subscales (performance assessment, performance development, and interpersonal evaluation). To enable a comparison of differences across leaders that would be independent of the average score of respondents, the within-group differences for each leader were calculated in such a way that the lowest individual score within each group of respondents would be coded as 1 for each variable⁵⁵. This was achieved by adding 1 to all individual scores and subtracting the lowest score within each group from this score. An example of this coding of within-group differences for the respondents of leader 2 is provided in Appendix H. To test whether within-group differences in the three subscales of evaluative style are associated with differences in age, tenure, and education, I used Kendall's tau correlations.

Proposition 5

Proposition 5 was not tested statistically, but the data obtained from interviewing the superiors is used to shed light on this issue. I originally intended to use the interview data obtained from subordinates as well. However, the majority of subordinates simply stated that their superior could possibly evaluate subordinates in different ways, but that they did not really know. Yet, most of these subordinates indicated that they trusted that their superior was evaluating all subordinates equally fair. Unfortunately, the topic has not been discussed (in depth) in all interviews with the superiors so the interview data do not allow a full exploration of this issue.

⁵⁵ This way of coding makes the within-group differences independent of the average within-group score, and is just a way of recoding the values of the original data, without changing the measurement properties within groups. The measurement properties would change if I had standardized the data within groups, making the data hard to compare across groups.

8.2.3 Results

Proposition 3

Table 8.1 displays Kendall's tau correlation⁵⁶ at the aggregated group-level between task uncertainty and (1) the 17 items of perceived evaluative style, classified into three categories, and (2) the three subscales of perceived evaluative style.

When respondents are grouped by leader, task uncertainty is only significantly correlated with an emphasis on non-financial information ($\tau = -.434$, $p < .05$), leader's intuition ($\tau = -.534$, $p < .05$), and positive aspects of the respondent's performance ($\tau = -.457$, $p < .05$). As task uncertainty increases, the emphasis on these three aspects increases as well⁵⁷. Two of these aspects have been classified as unspecified, and one item as a qualitative, interpersonal aspect. No significant correlations were found between task uncertainty and items referring to quantitative aspects and deviations from targets or between task uncertainty and other qualitative, interpersonal aspects of performance.

Table 8.1 *Kendall's Tau correlations between perceived evaluative style and task uncertainty, aggregated by leader (n=11) and leader-function group (n=14)*⁵⁸

	Task Uncertainty	
	Leader	Leader-Function Group
items referring to quantitative aspects and deviations from targets		
short-term goals	0.150	-0.113
financial information	-0.191	-0.164
deviations from agreed performance	-0.117	0.166
objectively measurable performance	-0.094	-0.012
performances of your unit	0.224	-0.113
items referring to qualitative aspects		
your explanation of your performance	0.000	-0.046
your plans for the next year	-0.135	-0.057
positive aspects of your performance	-0.457*	-0.398*
causes of failings in performance	-0.321	-0.034
possibilities to improve your performance	-0.150	0.012
information from conversation	-0.191	-0.046

⁵⁶ As explained in § 5.5.2 Kendall's tau is a non-parametric statistic. In case of small samples as we have here, it delivers more conservative estimates than Pearson correlations. The use of Pearson correlations rather than Kendall's tau correlations would lead to the same conclusions as those displayed in table 8.1. However, using Pearson correlations, the three significant correlations in table 8.1 would be significant at the 1% level (1-tailed).

⁵⁷ Note that a higher score on the task uncertainty scale indicates lower task uncertainty and vice versa. Thus, the negative signs in table 8.1 do indicate positive correlations between task uncertainty and items of perceived evaluative style.

⁵⁸ As indicated in chapter 7, leader 1 was left out of analyses involving aggregate measures of perceived evaluative style. Thus, the analyses are based on 11 leaders and 14 leader-function group combinations.

unspecified items		
long-term goals	-0.272	-0.104
non-financial information	-0.434*	-0.445*
his/her own intuition	-0.534*	-0.211
performance in the year past	-0.167	-0.093
negative aspects of your performance	-0.056	0.127
personal, individual performance	-0.220	-0.294
Subscales of evaluative style		
Performance Assessment	-0.294	-0.155
Performance Development	-0.167	-0.101
Interpersonal Evaluation	-0.019	0.167

* Correlation is significant at the 0.05 level (1-tailed)

When respondents are aggregated by leader-function groups, only two of the three correlations that were significant with subordinates grouped by leader remain significant. In general, most of the correlations are lower in comparison to the results with aggregation at the leader level.

The lower part of table 8.1 displays Kendall's tau correlations between task uncertainty and the three subscales of perceived evaluative style. All three correlations are not significant, which indicates that, contrary to expectation, aggregate measures of perceived evaluative style are not related to (group-level) task uncertainty.

These results do not support proposition 3. There is no correlation at the group-level between task uncertainty and quantitative aspects of performance evaluation and deviations from targets, between task uncertainty and qualitative, interpersonal aspects of performance evaluation, and between task uncertainty and perceived evaluative style subscales. Thus, leaders do not seem to adapt their evaluative styles to the (average) level of task uncertainty faced by subordinates.

Proposition 4

Table 8.2 displays Kendall's tau correlations between within-group differences of the three subscales of perceived evaluative style and age, tenure, and education. The results indicate that within-group differences in the different dimensions of perceived evaluative style are not associated with within-group differences in age, education, or tenure. The only significant association is between within-group differences of performance assessment and the period that a respondent has been working under his current superior ($\tau = .336$, $p = .002$)⁵⁹. Overall, there is hardly any support for the expectation that leaders may adjust their

⁵⁹ The use of Pearson correlations instead of Kendall's tau correlations would have the same results: the correlation between within-group differences in performance assessment and time with superior would be significant at the 1% level ($r = .427$, $p = .001$).

style of evaluation to subordinates' characteristics, like age, tenure, and educational level (proposition 4).

Note, however, that table 8.2 shows that there are high positive correlations between the three within-group differences in the evaluative style subscales. This indicates that, within groups reporting to the same superior, subordinates who report relatively high scores on one of the three subscales are likely to report relatively high scores on the other two subscales as well.

Table 8.2 *Kendall's Tau Correlations Between Within-Group Differences of Perceived evaluative style and Age, Tenure, and Education (n=57)*

	Within-group differences in perceived evaluative style		
	Assess	Develop	Personal
Within-group differences in:			
Performance Assessment (Assess)	1.000		
Performance Development (Develop)	0.420**	1.000	
Interpersonal Evaluation (Personal)	0.533**	0.555**	1.000
Age	-0.004	0.020	0.103
Education	0.126	0.100	0.150
Tenure:			
- Time with organization	0.154	0.123	0.184
- Time in function	0.151	0.023	0.077
- Time with leader	0.336**	0.183	0.141
- No. of evaluations by current leader	0.100	-0.001	0.176

** Correlation is significant at the .01 level (2-tailed).
* Correlation is significant at the .05 level (2-tailed).

Proposition 5

During the interviews, I asked some superiors whether they evaluated subordinates in the same way, and if not, why not. Although, as mentioned before, the interview data do not contain much information on this topic, there is some evidence that superiors will adapt their style of leadership/ evaluation to individual subordinates. Leader 2 for example indicates that one of his subordinates is not doing a good job considering his work level, age, and experience, so he definitely has to develop his skills and competencies. Leader 2 seems to have more formal discussions of how things are progressing during the year with this subordinate than with the other subordinates. Leader 12 is most explicit in his answer:

"Yes, of course I am more direct with some subordinates than with others. In the end, the result depends on the level of acceptance. And if you start to fully emphasise the two things that could improve a little with people that are very sensitive or something, they will clam up completely, start crying and walk away... That will get you nowhere, so of course you will adapt the style to how your message gets across with someone. And besides that of course to

how far someone is in his development. For example, I have one shift leader who has done this job for many years. He is of course further in his learning curve than someone who attracted attention among the operators last year. You will evaluate them differently. I can imagine that I will give the experienced guy a G (good) or an average and say: "Well, you did pretty good..." -and give the other guy a very good, although they perform at the same level, or perhaps the guy who gets a good or average performs even better - "but I just think that considering what you are capable of and how long you have been here, that we can expect more. That simply means that your development is jeopardised, because we do not see any improvement".

These interviews suggest that at least some superiors intend to adapt their style of evaluation according to individual subordinates' actual performance relative to working experience and age.

8.3 Perceived evaluative style and leadership style

8.3.1 Introduction

In chapter 4 I have developed the following propositions as to the relationship between inspirational and transactional leadership and perceived evaluative style:

- P6. Inspirational leadership will be more positively associated with an emphasis on interpersonal, qualitative aspects of performance evaluation than transactional leadership.*
- P7. Transactional leadership will be more positively associated with an emphasis on quantitative measures of performance and deviations from targets than inspirational leadership.*

Like proposition 3, these propositions will be tested first using the classification of the 17 perceived evaluative style items as described in § 7.4.1 into 1) quantitative, target-oriented aspects, (2) qualitative, interpersonal aspects, and (3) an unspecified category. Next, the propositions will be tested using the three evaluative style subscales (Performance Assessment, Performance Development, and Interpersonal Evaluation). For this purpose, the propositions have to be tailored to these subscales. All items of the Performance Development subscale were classified as pertaining to qualitative, interpersonal aspects of performance evaluation. The five items of Performance Assessment are spread over the three different theoretical categories. Thus, it is not possible to classify Performance Assessment in terms of an emphasis on quantitative measures of performance and deviations from targets or an emphasis on qualitative, interpersonal aspects of performance. Finally, Interpersonal Evaluation clearly refers to qualitative, interpersonal aspects of performance evaluation. Therefore, based on propositions 6 and 7 I expect the following relationships between the three evaluative style subscales and leadership style:

P15. Inspirational leadership will be more positively associated with Performance Development and Interpersonal Evaluation than transactional leadership.

P16. There will be no difference in association between inspirational leadership and transactional leadership with Performance Assessment.

8.3.2 Data analysis technique

To test propositions 6 and 7, associations between the 17 perceived evaluative style items as classified as described in § 7.4.1 into 1) quantitative, target-oriented aspects, (2) qualitative, interpersonal aspects, and (3) an unspecified category and three leadership subscales were calculated using Kendall's tau coefficient. The three leadership subscales were active management by exception, supportive leadership, and teambuilding. Active management by exception is an aspect of transactional leadership, while supportive leadership and teambuilding are aspects of inspirational leadership. The results are described in § 8.3.3. I will also use Kendall's tau correlations to test the association between the three evaluative style subscales and the three subscales of leadership style (propositions 15 and 16). These results will be reported in § 8.3.4.

8.3.3 Results for leadership style and 17 perceived evaluative style items

Table 8.3 presents the Kendall's tau coefficients for associations between the three leadership subscales and 17 perceived evaluative style items.

Table 8.3 *Association between leadership subscales and items measuring perceived and preferred evaluative style, using Kendall's tau coefficients (n= 57)*

	Leadership		
	AMBE	SUPPORT	TEAM
Active management-by-exception	1.000	.093	.076
Supportive	.093	1.000	.312**
Team-building	.076	.312**	1.000
	Perceived evaluative style		
	AMBE	SUPPORT	TEAM
items referring to quantitative aspects and deviations from targets			
short-term goals	.151	-.107	.023
Financial information	.276**	-.002	-.083
deviations from agreed performance	.398**	-.038	-.122
objectively measurable performance	.261*	.347**	.022
Performances of your unit	.266*	.287**	.054

Table 8.3 (continued)

items referring to qualitative aspects			
your explanation of your performance	.101	.459**	.178
your plans for the next year	.088	.313**	.107
positive aspects of your performance	.152	.463**	.107
causes for failings in performance	.375**	.425**	.238*
possibilities to improve your performance	.306**	.343**	.354**
information from conversation	.167	.513**	.265*
unspecified items			
long-term goals	.043	.089	.022
Non-financial information	.024	.160	-.065
His/her own intuition	.035	-.002	-.161
Performance delivered in the past year	.191	.226*	.008
negative aspects of your performance	.336**	-.088	-.177
personal, individual performance	.100	.412**	-.063

All five correlations of active management by exception with items classified as referring to quantitative aspects and an emphasis on deviations from targets are higher than the correlations of teambuilding with these aspects, and three of the five correlations are higher than for support⁶⁰. These results support proposition 7. In contrast, all correlations of support with qualitative, interpersonal aspects of performance evaluation are higher than the correlations of active management by exception with these aspects. For teambuilding, four of the six correlations are higher than for active management by exception. These results support proposition 6.

A closer look at table 8.3 provides interesting insights beyond the relationships hypothesised in propositions 6 and 7. These insights come from looking at the relationship of each of the three leadership subscales to the different aspects of performance evaluation in itself, rather than in comparison to each other.

First, table 8.3 shows that positive associations are found between active management-by-exception and seven perceived evaluative style items. Of these seven items, four items have been classified as referring to quantitative information or deviations from targets. In addition to these items, active management-by-exception is also positively correlated with causes for failings in performance, possibilities to improve your performance, and negative aspects of your performance. These items seem to refer to shortcomings in performance in general, without implying that these shortcomings in performance are monitored or measured by use of quantitative measures or targets. This could be true also for deviations from agreed

⁶⁰ I did not test the significance of the differences in correlations. Meng, Rosenthal & Rubin (1992) developed a test to assess the significance of the difference between two dependent correlations. The Meng, Rosenthal &

performance, an item which was classified as referring to quantitative information or deviations from targets, but, of course, could also refer to deviations from more general (oral or written) agreements, such as on the timely delivery of a project plan or project proposal, or on the progress of projects. Thus, although the results confirm a relationship between transactional leadership and the use of quantitative information and targets when evaluating performance, the results indicate that the strongest associations between active management-by-exception and perceived evaluative style items are with items that refer to shortcomings in performance in general, without necessarily implying the use of quantitative information or targets.

Next, table 8.3 shows that support is positively associated with all perceived evaluative style items that seem to refer to interpersonal, qualitative aspects of evaluation, although support is also positively associated with objectively measurable information, performances of your unit, performance delivered in the past year, and personal, individual performance. These results seem to suggest that supportive leadership is associated with a leader who is well informed on the performance of his subordinates and their units, and is willing to evaluate and develop the performance of subordinates through personal interaction.

Finally, the teambuilding subscale of leadership does not seem to be of high significance in explaining differences in perceived evaluative style. It is positively associated with only three items, two of which are positively associated with all three leadership subscales. Because teambuilding and support both measure aspects of inspirational leadership and are correlated to each other, these two associations could be explained through the influence of supportive leadership on these two items. To test this, partial correlations were calculated between all 17 items and teambuilding, controlling for supportive leadership. When supportive leadership is partialled out, teambuilding is only significantly and positively correlated with "possibilities to improve/develop your performance" (.248, $p=.046$) and significantly negatively with "personal, individual performance" (-.336, $p=.018$). In contrast, when teambuilding is partialled out, supportive leadership is still significantly correlated with the same items as in table 8.3, although the correlation with "performance delivered in the past year" is only significant at the 10% significance level. Additionally, when teambuilding is partialled out, support is also significantly positively associated with non-financial information (.292, $p=.042$).

8.3.4 Results: leadership style and three perceived evaluative style subscales

Kendall's tau correlations between the three subscales of perceived evaluative style and the three subscales of leadership style are displayed in table 8.4.

Rubin test tests the difference between Pearson correlations, while I used Kendall's tau correlations.

Table 8.4 Kendall's Tau-b Correlations between Leadership Subscales and Perceived evaluative style Subscales (n= 57)

	Leadership		
	AMBE	SUPPORT	TEAM
Performance Assessment	.182	.463**	.068
Performance Development	.277**	.480**	.272**
Interpersonal evaluation	.298**	.425**	.201*

Table 8.4 shows that performance development and interpersonal evaluation have higher significant positive correlations with supportive leadership (inspirational) than with active management-by-exception (transactional). This is as expected. However, the correlations of performance development and interpersonal evaluation with teambuilding, which is also a component of inspirational leadership, are not higher than with active management-by-exception. Thus, proposition 15 is partly supported.

Table 8.4 also shows that performance assessment is significantly positively correlated with supportive leadership, but is not significantly related to active management-by-exception and team-building. Thus, proposition 16, which states that I expected no correlation between leadership subscales and performance assessment, is supported for active management-by-exception and team-building, but not for supportive leadership.

Again, in addition to the test of the propositions 15 and 16, it is also interesting to look at the relationship between each of the leadership subscales to the performance evaluative style subscales.

Table 8.4 shows that active management-by-exception is significantly positively correlated with performance development ($\tau = .277$, $p = .005$) and with interpersonal evaluation ($\tau = .298$, $p = .003$). The correlation with performance assessment ($\tau = .182$, $p = .066$) is not significant at the 5% level of significance, but it is at the 10% level of significance. Supportive leadership is significantly positively related at the 5% level with all three subscales of perceived evaluative style.

Teambuilding is significantly positively related to performance development ($\tau = .272$, $p = .007$) and interpersonal evaluation ($\tau = .201$, $p = .049$), but not with performance assessment ($\tau = .068$, $p = .497$).

These findings, except for the relationship with interpersonal evaluation, do not need much explanation, since the same items as in the previous section on the relationship between leadership and individual items are used. The only difference here is, of course, that the individual items have been classified differently in the empirical scales than the theoretical distinction between quantitative, target-oriented versus qualitative, interpersonal aspects. The results for the individual items in the performance assessment and performance development subscales from table 8.3 in the previous section have been repeated in table 8.5, now classifying these items according to the empirical scales.

Table 8.5 Association between leadership subscales and items measuring perceived and preferred evaluative style, using Kendall's tau coefficients (n= 57)

	Perceived evaluative style		
	AMBE	SUPPORT	TEAM
Items of Performance Assessment			
your explanation of your performance	.101	.459**	.178
positive aspects of your performance	.152	.463**	.107
personal, individual performance	.100	.412**	-.063
performance delivered in the past year	.191	.226*	.008
performances of your unit	.266*	.287**	.054
Items of Performance Development			
possibilities to improve your performance	.306**	.343**	.354**
information from conversation	.167	.513**	.265*
your plans for the next year	.088	.313**	.107
causes of failings in performance	.375**	.425**	.238*

The table shows that only one of the five items of performance assessment (performances of your unit) is significantly positively correlated with active management-by-exception. This item is the only item that has been classified as pertaining to quantitative, target-oriented aspects of performance.

All three leadership dimensions are significantly positively correlated with interpersonal evaluation, although after controlling for the influence of supportive leadership, this correlation is no longer significant for teambuilding. The positive correlation between supportive leadership and interpersonal evaluation is as expected, since the findings using individual items earlier confirmed the expectation that supportive leadership would be positively associated with interpersonal aspects of evaluation. However, I did not expect a relationship between active management-by-exception and interpersonal evaluation. To help interpret these findings, I decided to calculate Kendall's tau correlations between active management-by-exception and support with the five single items of the interpersonal evaluation measure. Table 8.6 displays the outcomes.

Table 8.6 Kendall's tau correlations between two dimensions of leadership style and five items comprising the interpersonal evaluation measure (n=57)

	AMBE	Support
My supervisor is thoroughly familiar with my job performance.	.372**	.411**
My supervisor allows me to tell my side of the story in performance evaluation.	.094	.310**
My supervisor lets me know my appraisal outcomes and provides justification.	.260*	.252*
My supervisor lets me know my pay raise and annual bonuses and provides justification.	.235*	.296**
My supervisor reviews my performance with me and discusses plans or objectives to improve my performance.	.181	.308**

These results do provide some useful insights.

First, supportive leadership is significantly positively correlated with all items, including the two items that refer to the influence of subordinates on their evaluation, i.e. whether they are allowed to tell their side of the story, and whether they are allowed to participate in developing plans to improve performance. In contrast, active management-by-exception is not significantly correlated with these two items. The three items with which active management-by-exception is correlated all refer to the supervisor's role in performance evaluation but not to the subordinate's role: whether the supervisor is thoroughly familiar with the job performance of the subordinate, and whether the supervisor clearly communicates and justifies appraisal outcomes, pay raise and annual bonuses. These results may suggest that as leaders (in the perception of subordinates) display more active management-by-exception, they may gain a better knowledge of the job performance of subordinates, possibly because they will monitor or measure the performance of subordinates more carefully, and may use this knowledge to make better-informed and justified evaluations.

Second, earlier in this chapter I have suggested that "supportive leadership is associated with a leader who is well informed on the performance of his subordinates and their units, and is willing to evaluate and develop the performance of subordinates through personal interaction." The results in table 8.6 provide further evidence for this conclusion, as support is most positively associated with the supervisor's knowledge of the job performance of his subordinates, and with the two items that refer directly to personal interaction.

8.4 Perceived evaluative style and subordinate managers' behaviour: a quantitative analysis

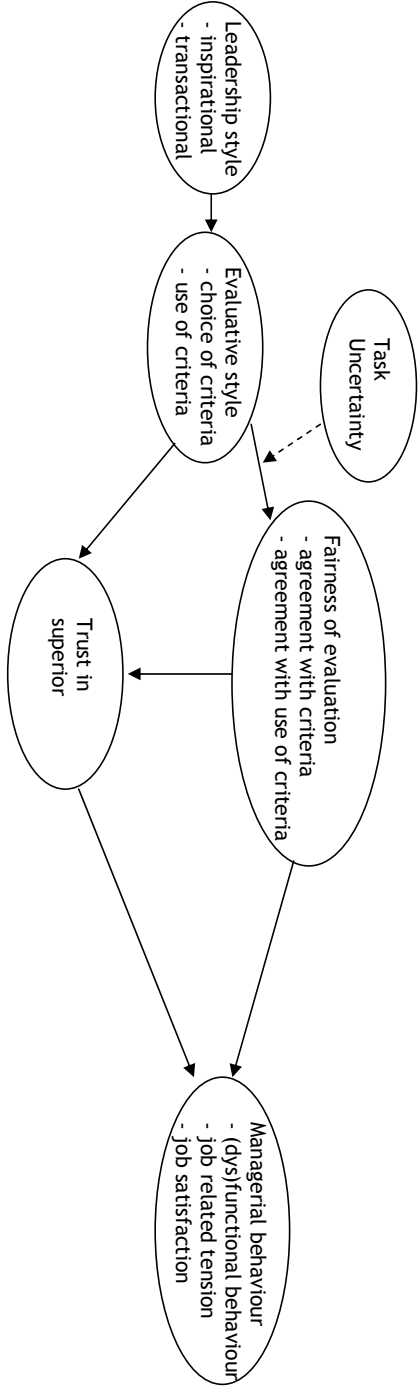
8.4.1 Introduction

The third topic that I will explore in this chapter is the relationship of perceived evaluative style with subordinate managers' behaviour, with a focus on the intervening role of perceived fairness of evaluation and trust in superior. Based on a review of the literature, in chapter 4 I developed the framework depicted in figure 8.2.

The relationships as depicted in this figure were summarised by the following propositions:

- P8. The relationship between perceived evaluative style and subordinate manager's behaviour will be explained by an indirect effect whereby perceived evaluative style affects fairness of evaluation, which in turn is positively related to subordinate manager's behaviour (i.e. enhances job satisfaction and reduces job-related tension).*
- P9. If subordinates agree with the criteria that are used to evaluate their performance, feelings of unfairness will arise if superiors ignore these criteria or do not attach as much importance to these criteria as they should according to the subordinate. On the other hand, if subordinates disagree with the criteria used to evaluate their performance, feelings of unfairness will increase as superiors attach more importance to these criteria and ignore the objections of subordinates to these criteria.*

Figure 8.2 General framework identifying antecedents and consequences of perceived evaluative style (figure 4.2)



- P10. The level of agreement with evaluative criteria reported by subordinates can be explained by giving explicit consideration to characteristics of the criteria that are used to evaluate performance in combination with the level of task uncertainty.*
- P10a. As task uncertainty increases, the felt appropriateness of accounting performance measures will decrease.*
- P10b. Evaluative styles that are characterised by an emphasis on quantitative measures of performance and deviations from targets will result in lower levels of perceived fairness of evaluation in high task uncertainty situations than in low task uncertainty situations.*
- P10c. Evaluative styles that are characterised by an emphasis on qualitative, interpersonal measures of performance will result in the same level of perceived fairness of evaluation in high task uncertainty situations as in low task uncertainty situations.*
- P11. The relationship between perceived evaluative style and subordinate manager's behaviour will be explained by an indirect effect whereby perceived evaluative style affects trust in superior, which in turn is positively related to subordinate manager's behaviour (i.e. enhances job satisfaction and reduces job-related tension).*
- P12. The relationship between perceived evaluative style and trust in superior will be explained by an indirect effect whereby perceived evaluative style affects fairness of evaluation, which in turn is positively related to trust in superior.*

In this section, I will use the data from the written questionnaire to test propositions 8, 11 and 12. Propositions 9, 10, 10a, 10b, and 10c will not be tested in this section. In § 8.5, the qualitative data from the interviews will be used to explore all propositions, and to explore the relevance and validity of the framework in more detail.

8.4.2 Data analysis technique

The propositions 8, 11 and 12 hypothesise indirect effects, and, more specific, as can be seen in figure 8.2, full mediation. Full mediation means that I expect that there is no direct effect of perceived evaluative style (subscales) on either job related tension or job satisfaction, but only indirect through an intervening variable: fairness of evaluation and/or trust in superior. According to Baron and Kenny (1986), mediating occurs if four conditions are met:

1. The independent variable(s) must be significantly and uniquely related to the dependent variables.
2. The independent variable(s) must be significantly related to the proposed mediator variable(s).
3. The proposed mediator(s) must be significantly and uniquely related to the dependent variables, while controlling for the effect of the independent variable(s).
4. The significant relation between the independent variable(s) and the dependent variables is no longer significant (full mediation) or is significantly weaker (partial mediation) when the mediator is added to the regression equation.

To test if these four conditions were met, I used hierarchical linear regression analyses. The hierarchical regression analyses consisted of three steps. Condition one and two require that the evaluative style subscale should have a significant unique effect on the dependent and on the mediating variables. Therefore, in the first step, the three evaluative style subscales

were entered into the regression to test their effect on the mediating (fairness of evaluation, and trust in superior) and the dependent variables (job satisfaction, and job related tension). The propositions 8 and 12 suggest that fairness of evaluation mediates the relationship between the three evaluative style subscales and trust in superior (proposition 12), job satisfaction (proposition 8), and job related tension (proposition 8). Condition three of Baron & Kenny requires that fairness of evaluation in this step has a unique and significant effect on trust in superior, job satisfaction, and job related tension, while controlling for the effect of the three evaluative style subscales. At the same time, the fourth condition requires that the effect of the evaluative style subscales on these variables is significantly decreased in comparison to the effect in the first step. Therefore, in the second step, fairness of evaluation was added to the regression to test the mediating effect of fairness of evaluation on the relationship between evaluative style and trust in superior, job satisfaction, and job related tension. To test whether the effect of the perceived evaluative style subscales has significantly decreased, I used bootstrapping. In small samples, bootstrap confidence intervals for indirect effects have been shown to have higher power and more accurate type I error rates (Preacher & Hayes, "Testing Mediation Models") than other tests, including the Sobel test (Sobel, 1982) that Baron & Kenny (1986) recommend. An important advantage of bootstrapping over the other tests is that it does not make assumptions about the sampling distribution of the indirect effect (See Preacher & Hayes (2006) for a complete overview). Bootstrapping was conducted using the SPSS macro that Preacher & Hayes (2006) developed, which can be found at <http://www.quantpsy.org>. Finally, in the third step, trust in superior was added to the regression to test the mediating role of this variable (proposition 11). Similar to fairness of evaluation, in this step trust in superior must have a unique and significant effect on job satisfaction and job related tension, and the effect of the evaluative style subscales and fairness of evaluation should be significantly decreased.

Although the causal steps method of Baron & Kenny has been often used (cf. Shrout & Bolger, 2002; Preacher & Hayes, 2004), there is some debate in the literature about the necessity of meeting the first condition. Shrout & Bolger (2002) argue that the first condition of Baron & Kenny should be relaxed, especially when the sample size is small (20-80 cases) and when there is no a priori expectation that the total effect of the dependent variable on the independent variable is medium ($>.30$) or large ($>.5$). Preacher & Hayes (2004) distinguish between mediated effects and indirect effects. Mediation effects imply that there is a significant total effect of the independent on the dependent variable which disappears or is weakened when entering the proposed mediator. Indirect effects imply that the statistical significance of the effect of the independent variable on the dependent variable through the proposed mediator is assessed. An indirect effect can be significant even though the total effect is not. A mediated effect thus represents a special case of an indirect effect. Therefore, if Baron & Kenny's second and third conditions are met, I will use bootstrapping to assess whether there are significant indirect effects. If significant indirect effects are

found, I will use Baron and Kenny's first criteria (i.e., assessing an overall significant effect) to assess whether these indirect effects represent mediated effects.

Since hierarchical regression analysis involves the comparison of different models, it is essential that in subsequent analyses the same cases are used. Therefore, the data set of 57 cases was further reduced by omitting cases which had missing values on any of the variables. However, if less than 25% of all items in the final scales had missing values, these missing values were replaced with the rounded mean of the other items in that scale. Cases in which more than 25% of the items on a single scale were missing were deleted from the sample. This meant that 3 cases were deleted from the sample due to missing values. Hierarchical linear regression analysis assumes linear relationships between variables and normally distributed variables. To test for the normality of the variables I assessed skewness, kurtosis and the Shapiro-Wilk statistic for each variable. The results (*not reported*) indicated that all variables did not deviate significantly from a normal distribution, except job satisfaction. Next, each of the variables was tested for univariate outliers, using box plots. This revealed no outliers for performance assessment, fairness of evaluation and job satisfaction, 3 outliers for performance development (cases 2, 19 and 50), 2 outliers for interpersonal evaluation (cases 2 and 47), 1 outlier for trust (case 60), and 1 outlier for job related tension (case 47). However, although identified as outliers, none of these cases were close to extreme values (more than three standard deviations from the mean). Therefore, I decided not to drop these cases from the sample, but to test for multivariate outliers first. Next, for each step in each regression analysis I compared the standardized residuals with the standardized predicted values. I identified one possible outlier for fairness of evaluation (case 63), none for trust and job related tension, and one for job satisfaction (case 20). Although, again, none of the standardized residuals was outside three standard deviations, I decided to drop these two cases. Because multivariate outliers (more than univariate outliers) could affect (the interpretation of) the results from the hierarchical regression analyses. Therefore, the results of all the regression analyses reported below are based on a final sample of 52 cases.

The zero-order correlations of the evaluative style subscales, fairness of evaluation, trust in the superior, and the two outcome variables are shown in table 8.7. The table shows both Pearson and Kendall's tau correlations. The table shows that there are no big differences in findings between the two techniques. Since hierarchical regression assumes parametric correlations, I will focus on the Pearson correlations. Table 8.7 shows that all variables are positively correlated with each other, except to job related tension. Higher emphasis on each of the perceived evaluative style subscales is associated with higher levels of fairness of evaluation, trust in superior, and job satisfaction. Furthermore, higher levels of fairness of evaluation and trust in superior are also associated with higher levels of job satisfaction. Job related tension is only significantly related to performance assessment, interpersonal

evaluation, and to job satisfaction, but not with any of the proposed mediators (trust and fairness of evaluation). All significant correlations with job related tension are negative, indicating that higher scores on performance assessment and on interpersonal evaluation are associated with lower levels of job related tension.

Table 8.7 *Correlations Between Variables - Raw Scores (n = 52) Pearson Correlations are shown above the diagonal, Kendall's Tau Correlations below the diagonal*

Variables	Perceived evaluative style			Intervening variables		Outcomes	
	1	2	3	4	5	6	7
1. Performance Assessment (ASSESS)	1.000	.451**	.640**	.653**	.488**	.420**	-.300**
2. Performance Development (DEVELOP)	.316**	1.000	.624**	.461**	.583**	.449**	-.116
3. Interpersonal Evaluation (PERSONAL)	.474**	.452**	1.000	.707**	.609**	.444**	-.322*
4. Fairness of evaluation (FEV)	.508**	.334**	.526**	1.000	.702**	.463**	-.170
5. Trust (TRUST)	.367**	.507**	.493**	.570**	1.000	.478**	-.092
6. Job satisfaction (JSA)	.323**	.358**	.339**	.342**	.391**	1.000	-.306*
7. Job-related tension (JRT)	-.180!	.003	-.138	-.108	-.075	-.250*	1.000

** Correlation is significant at the .01 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

8.4.3 Results

Tabel 8.9 displays the results of the hierarchical regression analysis based on the causal steps method of Baron & Kenny (1986)⁶¹. First, I will consider the hypothesised mediating effects of fairness of evaluation (proposition 8) and trust in superior (proposition 11) on the relationship between evaluative style and job satisfaction and job related tension. Next, I will discuss the results with respect to proposition 12.

Propositions 8 and 11 propose an indirect effect of perceived evaluative style on job satisfaction and job related tension through fairness of evaluation (proposition 8) and/ or trust in superior (proposition 11). Table 8.8 shows that, in the first step, none of the evaluative style subscales is significantly related to the outcome variables (job satisfaction and job-related tension). Thus, the first condition of Baron & Kenny has not been met for any of the evaluative style subscales. The results of step 2 and step 3 in table 8.8 show that neither fairness of evaluation nor trust in superior is significantly related to job satisfaction and job-related tension. Thus, the third condition of Baron & Kenny has not been met too.

⁶¹ Using AMOS 5.0.1., bootstrapped estimates and 95% confidence intervals for all parameters were obtained based on 10,000 bootstrapped samples. The results did not differ from the significance testing using a normal distribution as reported in table 8.8.

Table 8.8 Hierarchical regression of outcome variables on evaluative style

Step and Variables	Fairness of evaluation			Trust			Job satisfaction			Job related tension		
	1			1	2		1	2	3	1	2	3
1. Performance Assessment	.338**			.138	-.048		.206	.126	.135	-.172	-.241	-.237
Performance development	.004			.322*	.319*		.264	.263	.204	.154	.153	.122
Interpersonal evaluation	.487**			.320!	.052		.147	.032	.022	-.309	-.409!	-.414!
2. Fairness of evaluation					.550**			.236	.135		.206	.152
3. Trust									.185			.097
R ²	.567			.450	.580		.271	.295	.309	.133	.151	.155
Adjusted R ²	.540			.415	.545		.225	.235	.234	.079	.079	.063
F	20.993**			13.077**	16.257**		5.934**	4.910**	4.116**	2.455!	2.095!	1.691
Change in R ²					.131			.024	.014		.018	.004
F					14.645**			1.612	.959		1.012	.215
* Significant at 1% level (2-tailed)												
** Significant at 5% level (2-tailed)												
*** Significant at 10% level (2-tailed)												

This suggests that propositions 8 and 11 which propose an indirect effect of evaluative style on job satisfaction and job-related tension through fairness of evaluation and/ or trust in superior should be rejected.

Surprisingly, in contrast to my expectations, table 8.8 shows that the beta coefficients of fairness evaluation and trust in superior for job-related tension are positive, while I expected negative coefficients. As a result, the negative relationships between performance assessment and job-related tension and between interpersonal evaluation and job related tension become stronger in the second step. In the second and third step, at the a more liberal 10% significance level, interpersonal evaluation appears to be significantly negatively related to job related tension.

Proposition 12 predicts that the relationship between evaluative style and trust in superior is indirect through fairness of evaluation. The results of the first step displayed in table 8.8 show that performance assessment is positively related to fairness of evaluation ($p < .01$), but not significantly related to trust in superior. Performance development is not significantly related to fairness of evaluation, but is significantly positively related to trust in superior ($p < .01$). Finally, interpersonal evaluation is significantly positively related to fairness of evaluation ($p < .01$), and, at a 10% significance level, to trust in superior ($p < .10$). In the second step, when fairness of evaluation is added to the regression, the results show a significant positive effect of fairness of evaluation on trust in superior ($p < .01$). At the same time, the effect of performance assessment on trust in superior is reduced from .138 in the first step to -.048 and the effect of interpersonal evaluation on trust in superior is reduced from .320 to .052. The effect of performance development on trust in superior, remains significantly positive ($p < .01$), with a very small change in the beta coefficient from .322 to .319.

Based on the causal steps method, we can conclude that the first condition of Baron & Kenny, which requires a significant effect of evaluative style on trust in superior, has been met for performance development and interpersonal evaluation (although at the 10% level, not the 5% level), but not for performance assessment. The second condition, a significant effect of evaluative style on fairness of evaluation, has been met for performance assessment and interpersonal evaluation, but not for performance development. The third condition, which requires a significant effect of fairness of evaluation on trust in superior controlling for the effects of evaluative style, has been met. Thus, fairness of evaluation does not mediate the relationship between performance development and trust in superior, because condition two has not been met, and between performance assessment and trust in superior, because condition one has not been met. However, it is possible that there is a significant indirect effect on trust in superior through fairness of evaluation for performance assessment, although this is not a mediated effect. To assess the significance of the indirect effect for performance assessment and interpersonal evaluation, I used bootstrapping. The results are displayed in table 8.9. Table 8.9 shows the bootstrapped estimates and

Table 8.9 Bootstrapped estimates and confidence intervals for the indirect effects of evaluative style subscales on trust in superior through fairness of evaluation, using the SPSS macro found at <http://www.quantpsy.org>

	Point estimate	Indirect effect	Bootstrapping					
			Percentile 95% CI		BC 95% CI		BCa 95% CI	
			Lower	Upper	Lower	Upper	Lower	Upper
Assess → fairness → trust	.186	.192*	.052	.398	.060	.422	.053	.400
Personal → fairness → trust	.268	.263*	.066	.520	.088	.565	.086	.561

Note. BC = Bias Corrected; BCa = Bias corrected and Accelerated; 10,000 bootstrap samples

* Significant at 5% level (2-tailed)

confidence intervals for the indirect effects. These results are based on 10,000 bootstrap samples, using the SPSS macro developed by Preacher & Hayes (2006). The results in table 8.9 show that there is a significant indirect effect ($p < .05$) between performance assessment and trust in superior through fairness of evaluation. There is also a significant indirect effect between interpersonal evaluation and trust in superior through fairness of evaluation. Overall, I conclude that proposition 11 is partly confirmed for performance assessment, because there is no overall effect of performance assessment on trust in superior. Yet, there is a significant indirect effect on trust in superior through fairness of evaluation. The data also show that proposition 11 is fully supported for interpersonal evaluation, where fairness of evaluation mediates the positive relationship between interpersonal evaluation and trust in superior. For performance development proposition 11 is rejected. Performance development has a positive direct effect on trust in superior, but no indirect effect through fairness of evaluation.

8.5 Perceived evaluative style and subordinate manager's behaviour: a qualitative analysis

8.5.1 Introduction

In this section, the framework developed in chapter 4 will be addressed by analysing the data from the interviews. As explained in chapter 4, the framework was mainly developed to identify variables that should be measured in the written questionnaire and to identify the topics to address in the interviews. The framework is exploratory, and does not reflect strong theoretical relationships. One of the aims of interviewing subordinate managers in addition to using a written questionnaire was to improve the understanding of the relationship between evaluative style and subordinates' behaviour in this particular organisation. For this purpose, the interviews contained questions regarding the perceived fairness of performance evaluation, including the subordinate's opinion on or agreement with the performance evaluation system, and examples of dysfunctional and functional behaviours that were related to performance evaluation. While the written questionnaire was necessarily limited to variables identified in the framework, the interview data allowed the respondents to express their opinions in a more flexible way. The interview data thus enable a more flexible, context-based interpretation and "test" of the framework.

8.5.2 Subordinates' perceptions of the PDP-system

A distinction has to be made between performance evaluation for determining a financial bonus and performance evaluation for determining (future) promotion, while at the same time both approaches were integrated into a single system. The interview data reveal that subordinates indeed distinguished between targets, which were used to determine their financial bonuses, which they referred to as VPS, and (gaps in) competencies and skills - which they commonly referred to as PDP. While targets were set to focus attention and to

make clear what the priorities are, competencies and skills helped managers to develop themselves to become a better manager, and be of value for the organisation in the future. Therefore, receiving a bonus did not imply that you were a good manager, and not getting a bonus did not imply you were a bad manager, although at the same time subordinates admitted that if you were not a good manager, you probably would not receive a bonus. Although the target setting was part of the PDP-system, most subordinates used the term "performance evaluation" for the evaluation of their competencies and skills. From a career perspective, this was much more important for subordinates than whether or not they had met their targets and did or did not receive an additional financial bonus. In one of the interviews the (subordinate) manager explained this very well. Since his remarks are illustrative of the feelings that many other subordinates expressed during the interviews, I will quote from this interview at length:

"... You have to separate evaluation from VPS. Evaluation is something different than assessing what amount your bonus will be. The evaluation of performance in the past year, the way it went last year - and I expect that it will be the same this year- I agree with that one hundred percent. I don't have any problems with that. And also the criteria that go with that, yes. I suspect that I am free to change whatever I like during the year, if I think that is necessary to perform well... In the final evaluation, I will not be rigidly judged on targets, normally. The evaluation should be based on your performance during the whole year, and so all possible influences from outside may play a role. Assessing your results on the basis of your VPS, well, that are hard targets, that is not open to interpretation. I believe you can miss all targets, and be a very good manager. Because apparently you did to separate evaluation from VPS. Evaluation is something different than assessing what amount your bonus will be. The evaluation of performance in the past year, the way it went last year - and I expect that it will be the same this year- I agree with that one hundred percent. I don't have any problems with that. And also the criteria that go with that, yes. I suspect that I am free to change whatever I like during the year, if I think that is necessary to perform well... In the final evaluation, I will not be rigidly judged on targets, normally. The evaluation should be based on your performance during the whole year, and so all possible influences from outside may play a role. Assessing your results on the basis of your VPS, well, that are hard targets, that is not open to interpretation. I believe you can miss all targets, and be a very good manager. Because apparently you did something else that was important. And if this distinction is made, I think the two systems go together perfectly. But you have to be mature enough with one another to be able to discuss why. Not in order to adjust the bonus, not that, that is totally fixed. But in your evaluation."

Yet, as discussed in chapter 7, not in all units the competencies and skill section of PDP was fully covered, because this was new to the organisation and it required a lot of time to fill out this section. Besides, some subordinates indicate that there was some disagreement with their leader on either the appropriate job profile or on the actual competencies and skills of the subordinate when the "gap analysis" was performed. Some of these disagreements were

not fully resolved in the first year. The follow-up during the year on gaps that were identified is also a matter of concern to some subordinates. However, I will discuss these issues in more detail in the section "agreement with how the superior uses the system".

A variable pay system did exist before PDP was introduced. Therefore, setting targets during a PDP session did not cause much problems. Yet, not all subordinates agreed with the rigidity of determining the actual bonus, because there was no possibility to take into account changes in priorities during the year beyond the subordinates' circle of influence. If such a "priority change" occurred, sometimes a project that was part of the target setting was not important anymore, and the subordinate would miss (part of) the financial bonus. Subordinates felt this was unfair.

The difference between the determination of financial bonus and actual "performance evaluation" as subordinates described it in the interviews is summarised in Table 8.10.

Table 8.10 *Differences between two sections of PDP-system*

Target-setting	Competencies and skills
Performance assessment/determination	Performance evaluation
Objective criteria: targets	Objective criteria: job profile
Determine whether criteria have been met: mechanical	Determine whether criteria have been met: human interaction
One-way	Two-way
Short term performance	Long term performance
Results/output oriented	Development/process oriented
Financial bonus	Promotion/career
Operational	Strategic

Another issue that emerged from the interviews regarding the target-setting part (Variable Pay Schedule) was whether it truly motivates humans. In this organisation, it seems that the role of targets to motivate subordinates was severely limited. First of all, there is the problem of non-adjustable targets: once the target was set, neither the target nor the level at which the target was set could be changed. This was seen as unfair, and a number of subordinates said that they had (one or more) targets that were not relevant anymore because of changes in priorities. This decreased the relevance of the targets, and some subordinates indicated that they did not take the targets very seriously. Second, the maximum financial bonus the subordinate could earn was 15%. However, most subordinates expected they would only achieve a bonus of 6 to 10%. This would be nice, but was not enough to put a lot of extra effort in to achieve the targets. Finally, the higher the Work Level, the higher the maximum percentage for the bonus was. There is some evidence that subordinates felt this was unfair too: first, why would there have to be a difference in the percentage, and second, the higher the Work Level, the higher the base salary was anyway. So even the same percentage would lead to higher bonuses at higher Work Levels.

Table 8.11 summarises the positive and negative aspects of the system as they emerged from the interview data. In general, all interviewees were very positive about the new performance evaluation system, although subordinates indicated that the system had not reached its full potential yet, because of the problems outlined above and listed in table 8.11. Most subordinates considered these problems as inherent in implementing a new system, and in majority the problems were considered to be only minor problems that would be solved over time. Several subordinates recommended updating the targets on a regular basis, for example each month or each quarter. This would make target-setting more useful, and might ensure that targets remain relevant. In the words of one subordinate:

"If you would really want to use the system well you should... -Well, I have the opportunity myself. If I wanted it, I could have done it myself. I do not blame anyone.- but you should in fact revise the targets every three months. Check how we are doing, see if we should adjust them, and when you have to adjust them than it influences your VPS. My leader could have done that. I could have done it myself, I could have taken the initiative. But of course, because I am not really guided by these things I did not do that. But it would be an improvement of the system I think."

Others also indicated that they thought it would be useful if all agreements that were made in the PDP-form, not just targets, were reviewed more often during the year. To realise the full potential of the PDP-system, they felt that there had to be more discipline to use the system. Overall, the interviews clearly indicated that in general, subordinates accepted the system as fair and good. It is important to note, however, that these feelings were primarily based on expectations rather than actual experience with the system. This is due to the fact that the system had been implemented just a couple of months at the time of the interviews.

Table 8.11 *Positive and negative aspects of the evaluative system*

Positive aspects

- overall: systematic, complete, good system to evaluate performance
- identifies clear points of action
- objective (less subjective than old system)
- used within the whole multinational
- identification of skills and competencies for both functions and levels
- personal
- future-oriented
- no element of threat as in the old system
- better linked to business results
- much more effective because you are the owner of your own development plans instead of a one-way assessment as in the old system
- Provides tools to solve possible conflicts on evaluation between leader and subordinate
- Better insight in potential, both at the individual level as for the organisation as a whole

Table 8.11 (continued)

Negative aspects:

- Very time consuming
 - Comparability between organisations within multinational doubtful
 - Not used everywhere to the same extent
 - Risk of inflexibility when things change during the year
 - Variable Pay is less than in the old system
-

8.5.3 Overall fairness of evaluation

In chapter 4, I have argued that to understand subordinates' perceived fairness of evaluation, it is important to distinguish agreement with the formal design of the system from agreement of how leaders actually use the system. The description of differences in evaluative style as they emerged from the interview data in chapter 7 indicates the importance of this distinction. Although subordinates generally agreed with the PDP-system, most subordinates indicated that the manner in which their leader actually used it could be improved. The remarks on how leaders could improve performance evaluation were very closely related to and have already been incorporated in the more general dimensions of evaluative style as identified in § 7.2. Therefore, there is no need to discuss these again in detail, and I will confine the analysis to a few comments.

Given the differences between leaders, as summarised in table 7.1, some leaders seemed to use the PDP system in more satisfactory ways than others. In general, subordinates indicated that the follow-up on gaps in competencies and skills that had been identified in the PDP-form at the beginning of the year needed improvement. This required more active coaching from the leader, part of which would be providing feedback. However, it is important to note that despite ways for improvement, none of the interviews provided any evidence that subordinates disagreed with overall performance evaluation.

One possible explanation would be that the manner in which leaders actually use the system becomes more important if subordinates disagree with the formal design of the system. When subordinates agree with the formal design of the system and think it is a good system, as in this organisation, the fact that leaders do not use the system to its full potential may not lead to negative feelings. Instead, they may mitigate more positive feelings. Another explanation is that the organisational culture at VDBN is very informal and open, as mentioned in § 6.4. Subordinates do receive a lot of informal feedback, and thus, despite the lack of formal attention for the PDP-system, they may expect or trust that in the end their evaluation will be positive.

Finally, an important explanation that is offered in the interviews, which is closely related to the former two explanations, is that subordinates were aware of their own responsibility for their development. Using the words of one subordinate:

"And I also think that you are responsible for it yourself. I do not think that it is up to me to say: "The management does not pay any attention to it". No, you have to do it yourself. It is your own development plan."

This explanation is aligned with the strong (strategical) emphasis within Unilever and VDBN on personal growth of employees and on initiatives as Total Productive Maintenance (see § 6.2). It is part of (a change in) the organisational culture at VDBN. The importance and relevance of this explanation is also stressed by the analysis of the behavioural consequences of performance evaluation, to which I will turn now.

8.5.4 Dysfunctional behaviours

Most subordinates indicated that dysfunctional behaviours did occur, but also explicitly mentioned that this had nothing to do with the PDP-system.

As an example of dysfunctional behaviour, one subordinate indicated that she did not know how to game performance measures in her own job, but continued as follows:

"But I do know what you mean. I sometimes see it in purchasing for example. That is not just a result of PDP, but more of the way of target setting. The most simple example is that people are extremely driven by cost reduction and that may influence things like delivery time, delivery quantities. So you tell a supplier: "You will supply me glass 10% cheaper." The supplier will say: "Fine, it will take two more weeks before I can deliver it". So you have to be clear about what you want two weeks earlier. So yes, what you have to do is to make that trade-off: what do we actually want and what is in the end the most expensive? Perhaps it is much more expensive to have longer delivery times. And last year I have seen that people were focused too one-sidedly. "I simply have to get my bonus, so I am going to cut my glass purchase prices with 2%." I think it wasn't glass, but paper for the labels. It meant that we had tremendous problems in the plant with producing the right labels."

And another subordinate from marketing/ sales said:

"Of course we notice that now too: there is a very clear focus on volume. It means that you will pull out all the stops to realise that volume and sometimes you forget to ask whether it is really that profitable, the extra volume that you get in. Now it is important to realise a certain volume with all clear advantages of it. And the negative effects, well, yes, you will come across them again next year."

Most examples of dysfunctional behaviours that subordinates described had to do with a short-term focus without considering the long-term consequences or without considering the actual strategy of the organisation. Subordinates indicated that dysfunctional behaviour seemed to be related with the target-based bonus system, but was a bigger problem in the past. In the past, it was not uncommon for Unilever managers to get promoted to another job within two to three years. This encouraged a short-term focus, especially in combination with a variable pay system. Subordinates thought the PDP-system reduced dysfunctional

behaviours, as people would stay in their jobs longer than in the past and the variable pay schedule was complemented with an emphasis on competencies and skills. So, short-term pressures were released somewhat. At least some of the leaders recognised their own role in this process too. For example, when asked if dysfunctional behaviours occurred within his unit, one leader answered:

"Yes. Of course, the answer is yes. But it just depends on how tight you will hold on to your target. If you hold on to your target extremely tight, people will eventually use dysfunctional ways to realise it. They will make sure they will realise it. And then they will realise it. I think we have the flexibility here to say: "I rather adjust my target in order to align the behaviour of people with my Strategic Intent than to encourage this type of behaviour"."

8.5.5 Functional behaviours

As to the functional effects of performance evaluation, again, the fact that the PDP-system had not been fully used (see before) means that most subordinates were unable to relate improvements and learning to the PDP-system. In fact, many subordinates had learned by doing their job rather than by the PDP-system. In addition, the restructuring of the Foods business unit into value creation and value delivery (see § 6.2) had contributed a lot to improved ways of working, according to leaders and subordinates.

In contrast to most of the subordinates, many leaders did see functional consequences of the PDP-system, or at least of the changes within VDBN of which the PDP-system was part. Many leaders emphasised the rewarding, motivating role of the PDP-system to subordinates, and the learning and innovative behaviours that occurred as a consequence. Two quotes from two marketing/ sales leaders will illustrate this:

"... the PDP system is simple, clear. It is rewarding too. If it all works, if you really... That is the beauty of it. If you write a plan and you see that plan come true, then it is splendid. That gives great satisfaction. I am a marketer, so I feel the same way about my marketing plans. When you start planning people -and of course, you have to be a little careful-, but when you start planning the development of people together with them - and that's something too: it is a dialogue, not a monologue, it is not unilateral. I think the reciprocity of the PDP is a very good benefit - When you start planning people, and you finally see the result of it, and you can simply put it next to your planning, you can say: you see, I have taken these and these steps. So it is a self-learning system that way. That gives great satisfaction. So I think the benefits are: reciprocity, self-learning and -because of that- fulfilling."

"In the past we had one man working on one product on one project and very defensive behaviour was normal, while now you see more willingness to involve a larger group of people and try to find a solution. Now sometimes people give up part of their budget in favor of other projects. But this is more the result of the organisational change than of PDP."

Production leaders made similar remarks, although their remarks were broader than just their direct subordinates. They saw a change in attitude of all workers in the plant, not just the managers or shift leaders, as illustrated in the following quotes from three different leaders:

"I think that in comparison to how it used to be people now suggest more own ideas. And when you stimulate them to suggest their own ideas, and to feel responsible for things, they will do that..."

"What I notice is that if you look at how people take their responsibility and simply feel owners of what they do, start taking care of things, I really find that surprising. Then you just really see that there are tremendously good people among them. You simply notice: I do not have to spend much time in operations. People can handle that. You see more and more that if shift leaders are absent, and you do not replace them, that the team simply continues. That are very good signals. That they do not spend their whole shift in the canteen, even though it is night and no one sees what they are doing... It shows how mature people are. "

"I think technicians are not the most free-thinking people. And in this plant for the largest part you deal with this type of people. That means that you will not meet the most innovative, creative people by nature. On the other hand is the way in which some things are carried out here is innovative indeed... The fact that employees organise a number of things themselves, while before management had to organise it, that is innovative to me."

8.6 Summary and conclusions

In this chapter I explored three topics. First, I looked at the relationship between contextual factors and within- and between-group differences in perceived evaluative style. Second, I dealt with the relationship of leadership style with perceived evaluative style. Third, I investigated the relationship of perceived evaluative style with subordinate manager's behaviour. Table 8.12 summarises the main findings. These findings will be discussed below.

8.6.1 Contextual factors and perceived evaluative style (1.3)

Proposition 3 stated that task uncertainty could be an important contextual factor to help explain between-group differences in evaluative style. In particular, I expected that as task uncertainty increased, evaluative styles would be characterised by less emphasis on budgets and other quantitative targets, a higher emphasis on qualitative, interpersonal aspects of performance, and a higher willingness of superiors to listen to subordinates explanations of their performance (a more flexible use of budgets and other quantitative targets). The findings did not support this proposition. One explanation could be that leaders do not adapt their style of evaluation to the level of task uncertainty faced by their subordinates. This explanation is partly supported by the fact that three superiors had subordinates from two different function groups: production managers and production heads. These two groups differ in the level of task uncertainty, although these differences were not statistically

Table 8.12 Research questions and propositions addressed in this chapter and the outcomes

<i>Research questions and propositions</i>		<i>Findings</i>
1	Purpose: gain a better understanding of the construct of evaluative style	
1.3 Relationship between contextual factors and evaluative style		
<ul style="list-style-type: none">• How does the organisational context explain the underlying dimensions of perceived evaluative style that are found to be relevant within this organisation?• How does the organisational context explain the role of (financial) accounting information?		<ul style="list-style-type: none">• New evaluative system (PDP)• Informal culture <p>Results in : relatively high relevance of coaching, interpersonal/ qualitative aspects, performance development, and performance assessment.</p> <p>Role of (financial) accounting in evaluation is limited, probably because of reasons above. Additionally, subordinates are relatively low in management hierarchy.</p> <p>See also chapter 6.</p>
At the group level (leader level):		
P3.	As task uncertainty increases, evaluative styles will be characterised by less emphasis on budgets and other quantitative targets, a higher emphasis on qualitative, interpersonal aspects of performance, and a higher willingness of superiors to listen to subordinates explanations of their performance (a more flexible use of budgets and other quantitative targets).	<p>Not supported: no significant relationship between TU and quantitative aspects of performance evaluation and deviations from targets, between TU and qualitative, interpersonal aspects of performance evaluation, and between TU and perceived evaluative style subscales.</p>
At the level of individual subordinates:		
P4.	Within-group differences in age, tenure, and education of subordinates are related to within-group differences in perceived evaluative style.	<p>Hardly supported: only significant association between within-group differences of performance assessment and the period that a respondent has been working under his current superior ($r = .336$, $p = .002$).</p>
P5.	Within-group differences in perceived evaluative style partly reflect differences in evaluative style that the superior intended.	<p>Partly supported (but limited evidence)</p>

Table 8.12 (continued)

1.4 Relationship between (perceived) leadership style and (perceived) evaluative style		
P6.	Inspirational leadership will be more positively associated with an emphasis on interpersonal, qualitative aspects of performance evaluation than transactional leadership.	Supported
P7.	Transactional leadership will be more positively associated with an emphasis on quantitative measures of performance and deviations from targets than inspirational leadership.	Supported
2 Purpose: improve understanding of the relationship between evaluative style and subordinate managers' behaviour.		
2.1 Relationship between perceived evaluative style and subordinate managers' behaviour		
P8.	The relationship between perceived evaluative style and managers' behaviour will be explained by an indirect effect whereby perceived evaluative style affects fairness of evaluation, which in turn is positively related to managers' behaviour.	Not supported
P9.	If subordinates agree with the criteria that are used to evaluate their performance, feelings of unfairness will arise if superiors ignore these criteria or do not attach as much importance to these criteria as they should according to the subordinate. On the other hand, if subordinates disagree with the criteria used to evaluate their performance, feelings of unfairness will increase as superiors attach more importance to these criteria and ignore the objections of subordinates to these criteria.	Not tested. This proposition helps to interpret the results. In this particular organisation, subordinates agree with the (new) system. Therefore, they expect their superior to use it intensively.
P10.	The level of agreement with evaluative criteria reported by subordinates can be explained by giving explicit consideration to characteristics of the criteria that are used to evaluate performance in combination with the level of task uncertainty.	Not tested. Task uncertainty was treated as a group level variable. Sample is too small to test this proposition.
P10a.	As task uncertainty increases, the felt appropriateness of accounting performance measures will decrease.	Not tested. See under proposition 10.
P10b.	Evaluative styles that are characterised by an emphasis on quantitative measures of performance and deviations from targets will result in lower levels of perceived fairness of evaluation in high task uncertainty situations than in low task uncertainty situations.	Not tested. See under proposition 10.

Table 8. 12 (continued)

P10c. Evaluative styles that are characterised by an emphasis on qualitative, interpersonal measures of performance will result in the same level of perceived fairness of evaluation in high task uncertainty situations as in low task uncertainty situations.	Not tested. See under proposition 10.
P11. The relationship between perceived evaluative style and perceived fairness of evaluation will be explained by an indirect effect whereby perceived evaluative style affects trust, which in turn is positively related to perceived fairness of evaluation.	Partly supported: supported for performance assessment (indirect effect) and interpersonal evaluation (full mediation), not supported for performance development
P12. Subordinates who report similar perceived evaluative styles but different levels of trust will also report different levels of agreement with perceived evaluative style: the higher the level of trust, the higher agreement with perceived evaluative style.	Not supported
2.2 Explanation of findings by giving account to the organisational context in general and the design of the evaluative system in particular	See discussion in § 8.6.

significant, as reported in chapter 6 (see § 6.5.1 and 6.5.2). If I take into account these two different function groups by creating 14 groups rather than 11 (the number of superiors), in general the findings are very similar. I would expect that if proposition 3 were true the associations would be stronger, because using the 14 groups task uncertainty would be studied at the appropriate level.

Another explanation could be that only effective superiors adapt their style of evaluation to the level of task uncertainty, while ineffective leaders do not. Thus, the findings could indicate that there are effective and ineffective superiors in this organisation. Yet, this explanation requires an independent assessment of the effectiveness of superiors, which I did not obtain in this study. Finally, unfortunately, the lack of significant findings could be attributed to the small sample size. We should realise that at the group level the sample has only 11 or 14 observations. Therefore, the test of Kendall's tau associations lacks power. This is a severe limitation. Unless there are very strong correlations, it is hard to find significant associations.

Earlier, in chapter 7, I found that there are considerable within-group differences in perceived evaluative style. Proposition 4 stated that within-group differences in age, tenure, and education could help explain within-group differences in perceived evaluative style. The findings did not support this proposition, as the only significant association found was between within-group differences of performance assessment and the period that a respondent has been working under his current superior ($\tau = .336$, $p = .002$)⁶². These findings suggest that superiors do not adjust their style of evaluation to subordinate characteristics such as age, tenure, and education. The interview data obtained from superiors provide some evidence that at least some superiors intend to adapt their style of evaluation according to individual subordinates' actual performance relative to working experience and age. This may explain why I could not find support for proposition 4: perhaps age, tenure, and education are not important in itself, but only in relation to actual performance. I.e., superiors may tend to evaluate performance not in absolute terms, but relative to age, tenure, and education of subordinates. Additionally, the interview data offer personality of the subordinate as another possible explanation of why superiors may adapt their style of evaluation to individuals. Since I did not measure performance and personality of the subordinate in this study, it is impossible to test these alternative explanations.

8.6.2 Perceived leadership style and perceived evaluative style (1.4)

Next, the relationship between leadership and evaluative style was explored. The results indicate that leadership style and evaluative style are related. Different aspects of leadership relate to different aspects of evaluative style. Active management-by-exception is positively correlated with an emphasis on quantitative information or shortcomings in

⁶² The use of Pearson correlations instead of Kendall's tau correlations would have the same results: the correlation between within-group differences in performance assessment and time with superior would be

performance when performance is evaluated, while supportive leadership is positively correlated with more qualitative, interpersonal aspects of evaluation. Although teambuilding is a subscale of inspirational leadership, as is supportive leadership, teambuilding is not as important in explaining differences in evaluative style. However, supportive leadership and teambuilding differ from each other in that supportive leadership concerns the leader's approach to individuals, as against his/her approach towards a group of individuals in the teambuilding subscale. Since performance evaluation concerns the evaluation of individuals' performance, this could explain why supportive leadership is positively correlated with aspects of evaluative style, while teambuilding is not.

Although leadership and perceived evaluative style are related, none of the leadership subscales is correlated with either short-term goals or long-term goals. However, it is likely that aspects of leadership that measure a leader's concern with the future, such as vision, are related to this dimension of evaluative style. Since the leadership subscales that I used in this study focused more on the interpersonal aspects of leadership than on vision, this may explain why I do not find a relationship between leadership and the emphasis on short/ long term goals when evaluating performance.

The results also indicated that leadership style was related to the three evaluative style subscales. Active management-by-exception was significantly positively correlated with performance development ($\tau = .277$, $p = .005$) and with interpersonal evaluation ($\tau = .298$, $p = .003$). The correlation with performance assessment ($\tau = .182$, $p = .066$) is not significant at the 5% level of significance, but it is at the 10% level of significance. Supportive leadership is significantly positively related at the 5% level with all three subscales of perceived evaluative style.

Teambuilding is significantly positively related to performance development ($\tau = .272$, $p = .007$) and interpersonal evaluation ($\tau = .201$, $p = .049$), but not with performance assessment ($\tau = .068$, $p = .497$).

Combined with the findings about the relationship of leadership style with quantitative, target-oriented aspects and/or qualitative, interpersonal aspects of performance these findings provide further validation of the interpretation of the three subscales of evaluative style. In chapter 7, I argued that Performance Assessment refers to "the extent to which performance (in the past/ of the past year) is being assessed, based on a broad array of information (personal performance, unit performance, and explanation of performance)" (see § 7. 4.3). The findings in this chapter support this, as performance assessment is not associated strongly with active management-by-exception, but more with supportive leadership. This supports the broad array of information on which performance assessment is

significant at the 1% level ($r = .427$, $p = .001$).

based. Performance assessment seems to refer to active involvement and continuous monitoring of the performance of subordinates, rather than management-by-exception, or, in other words, to a coaching superior more than a superior who manages by numbers. As to the second subscale, performance development, I argued in chapter 7 that it referred to “the extent to which superiors use performance evaluation as a learning tool for the future.” Performance development was found to be positively correlated with active management-by-exception. This can be explained by the fact that performance development focuses on weaknesses or shortcomings in performance. Yet, performance development was also found to be positively correlated with supportive leadership. This suggests that the focus on shortcomings does not indicate a judgement of the appropriateness of the level of performance, but rather that shortcomings are seen as opportunities to learn and develop individual performance through interactions between superior and subordinate.⁶³

8.6.3 Perceived evaluative style and subordinate managers' behaviour

The third topic explored in this chapter was the relationship between perceived evaluative style and subordinate managers' behaviour. In particular, I explored whether fairness of evaluation and trust in superior are intervening variables that help to better understand the relationship between perceived evaluative style and subordinate managers' behaviour.

Summary of quantitative analysis

The results from the written questionnaire indicate that none of the perceived evaluative style subscales is significantly related to job satisfaction or job related tension, and, thus, fairness of evaluation and trust in superior cannot intervene in this relationship. Performance development had a significantly positive direct effect on trust in superior, but no effect on fairness of evaluation. Both performance assessment and interpersonal evaluation had a significant positive indirect effect on trust in superior through fairness of evaluation. The total effect of performance assessment on trust in superior was not significant, but the effect of interpersonal evaluation on trust in superior was. Thus, fairness of evaluation and trust in superior seemed to be important and relevant consequences of perceived evaluative style.

Summary of qualitative analysis

The interview data revealed that subordinates distinguished between targets, which were used to determine their financial bonuses, which they referred to as VPS, and (gaps in) competencies and skills - which they commonly referred to as PDP. The bonus was

⁶³ In this sense, both performance assessment and performance development seem to be aspects of a coaching leader. These two aspects indicate how a coaching leader evaluates the performance of subordinates.

determined mechanically, and as such did not provide much opportunity for discretionary use by leaders. The bonus system could cause short-term pressure, and, accordingly, a short-term focus and dysfunctional behaviours, and in the past it did. These short-term pressures were released by the explicit focus of the (newly developed and implemented) PDP-system on competencies and skills, which were seen as much more important from a career perspective than the possibility of obtaining a bonus. Therefore, although subordinates could provide examples of dysfunctional behaviours, most subordinates explicitly indicated that these behaviours were a bigger problem in the past. Subordinates expected that the PDP-system would reduce dysfunctional behaviours for two reasons. First, people would stay in their jobs longer than in the past, when it was not uncommon for managers to change jobs every two or three years. Second, the short-term pressure induced by the variable pay schedule in the past was reduced because it was now complemented with an emphasis on competencies and skills.

The interviews also indicated that in general the subordinates accepted the evaluative system that integrated VPS and PDP as a fair and good system. Yet, many subordinates stated that the manner in which their leader used the system could be improved. An important improvement would be that superiors give more attention to the follow-up on gaps in competencies and skills that have been identified in the PDP-form at the beginning of the year. This requires more active coaching from the leader, which includes providing subordinates with feedback on a more continuous basis. Indeed, this seemed to be the main reason why subordinates did not link functional behaviours -such as learning and innovative behaviour- to performance evaluation. Many subordinates felt that they had learned by doing rather than by the PDP-system. To achieve this learning by doing, again the coaching role of leaders seemed to be important, and thus it is this aspect that in the opinion of subordinates should be improved. Still, despite ways for improvement, the interviews did not provide evidence that subordinates disagreed with overall performance evaluation.

Conclusions

Combining the findings from the interviews with the findings from the written questionnaire and with earlier analyses, we can draw a number of conclusions. These conclusions will illustrate the importance of the organisational context in understanding perceived evaluative style and the behavioural consequences.

As the interviews show, dysfunctional behaviours as a consequence of performance evaluation in this particular organisation were not a major concern. Furthermore, the interviews indicated that subordinates had high expectations of the new evaluative system, and agreed with the system. Even though the manner in which their superiors used the system could be improved, subordinates did not disagree with overall performance evaluation. Possibly, the manner in which superiors use the evaluative system becomes more important if subordinates disagree with the formal design of the system. When subordinates agree with the formal design of the system and think it is a good system, as in this

organisation, the fact that leaders do not use the system to its full potential may not lead to negative feelings. Instead, they may mitigate more positive feelings.

Although the previous explanation could explain why job related tension was not affected by perceived evaluative style, I would still expect a relationship with job satisfaction. It should be noted, however, that I had problems with the reliability and validity of the scale to measure job satisfaction (see § 5.4.7). The items used in this study to measure job satisfaction seemed to refer to satisfaction with the organisation more than satisfaction with the job. Since I expect that evaluative style is more likely to affect satisfaction with the job than satisfaction with the organisation, this could explain the lack of significant findings with regard to job satisfaction.

Another explanation for my findings is that the organisational culture at VDBN is very informal and open, as mentioned in § 6.4. Subordinates do receive a lot of informal feedback, and thus, despite the lack of formal attention for the PDP-system, they may expect or trust that in the end their evaluation will be positive. Again, this could explain why I did not find a relationship between perceived evaluative style and job related tension. It also helps explain why I did find a relationship with trust in superior and fairness of evaluation.

In addition, the interviews indicated that subordinates were aware of their own responsibility for their development. Indeed, the system does not allow subordinates to be passive and wait for a one-way assessment of their performance by their superior, but rather stimulates them to recognise and take their own responsibility. This explanation is aligned with the strong (strategical) emphasis within Unilever and VDBN on personal growth of employees and on initiatives as Total Productive Maintenance (see § 6.2). It is part of (a change in) the organisational culture at VDBN. Against this background, it is not surprising that many subordinates felt that their superior should have a coaching role. The two empirically found subscales of perceived evaluative style, performance development and performance assessment, seem to capture aspects of this coaching behaviour. This is also supported by the earlier finding that both subscales are positively correlated with supportive leadership. Thus, it is not surprising that a higher score on performance development and/or performance assessment was associated with higher scores on fairness of evaluation and trust in superior.

9 Conclusion

9.1 Introduction

In chapter 1, I outlined the two purposes of this study. The first purpose of the research reported in this thesis was to contribute to a better understanding of the concept of evaluative style, taking into account the shortcomings of, criticisms on and concerns raised in the existing RAPM-literature. The second purpose of this study was to contribute to a better (theoretical) understanding of how and why evaluative style is related to subordinate managers' behaviour.

In this chapter, I will discuss how this thesis has contributed to these two purposes. I will do this by first providing a brief summary of the main findings from this study in § 9.2. Next, in § 9.3, I will discuss the main findings by focusing on the main contributions of this thesis to the scientific literature and its implications for practice. The aim of this section is not to discuss the results from each individual proposition that I developed and tested throughout this thesis in detail. This I did at the end of each chapter. Rather, the focus will be on the overall conclusions with regard to the two purposes that were set in chapter 1. In § 9.4, I will discuss the limitations of this study. Finally, I will propose a number of promising directions for future research (§ 9.5).

9.2 Summary of main findings

To achieve the two purposes that I identified at the start of this thesis, this study consists of a theoretical (chapters 2-4) and an empirical part (chapters 5-8).

In the theoretical part, starting from the existing RAPM literature, the literature on evaluative style has been reviewed and analysed in the light of contemporary developments within the management control field. One general conclusion from this review was that there is a need for qualitative in-depth field studies within single organisations rather than quantitative survey research across organisations. Conducting a field study within a single organisation enables the conceptualisation and measurement of evaluative style in ways that are relevant, meaningful and valid in the particular context studied, and 'accommodate changes in contemporary control systems' (Chenhall, 2003, p. 131). Instead of using the same measure of evaluative style for different organisations, tailored measures of evaluative style organisation may be needed in different organisations, to account and control for differences in the technical design of the evaluative system. In addition, I concluded that there is a need for future studies on evaluative style that study the behavioural impact of evaluative style as a complement to the impact of the 'fit' between design of components of performance evaluation and the wider organisational context. I called this the 'contextual approach'.

Furthermore, the following six topics emerged from the literature review as important topics that need to be addressed in future research:

1. The conceptualisation and valid measurement of underlying dimensions of perceived evaluative style that are meaningful in the particular organisation (see chapter 3);
2. The level of agreement in perceived evaluative styles within and between groups (see § 4.2.3);
3. The relationship between contextual factors at different levels of analysis and perceived evaluative style (see figure 4.1);
4. The relationship between perceived leadership style and perceived evaluative style (§ 4.3).
5. The relationship between perceived evaluative style and managerial behaviour (§ 4.4);
6. The role of the organisational context, and the design of the evaluative system in particular.

For each of these topics, I developed propositions and research questions which serve to guide their empirical exploration.

The empirical part of this research consists of an exploratory field study that I conducted within twelve units of Van den Bergh Nederland (VDBN), a contemporary Dutch food-processing company. This field study illustrates how the topics and propositions that emerged from the literature review can be applied to a particular context. Data in the field study were obtained from documents, personal observations, interviews with all twelve superiors and two or three of their subordinate managers (27 subordinates in total), and a written questionnaire that was sent to all subordinates ($n=57$) who reported to one of the twelve superiors.

Below the main results from the field study within VDBN on each of the six topics are described.

1. Conceptualisation of evaluative style

Differences in perceived evaluative style do exist, but these differences are not based on the use of budgets or the use of targets, as in most of the existing RAPM-literature. Instead, I identified three dimensions of evaluative style that seemed particularly relevant within VDBN. The first dimension was performance assessment, which refers to “the extent to which a superior actually assesses performance of the subordinate (in the past/ of the past year), based on a broad array of information (personal performance, unit performance, and explanation of performance). The second dimension was performance development, which refers to the extent to which a superior uses performance evaluation as a learning tool to

develop the subordinate's future performance. The third dimension was interpersonal evaluation. This dimension measured interpersonal behaviour of the superior in performance appraisal and pay decisions.

2. Agreement on perceived evaluative style between- and within-groups

The interview data suggested that different superiors have different styles of evaluation, and the results from the written questionnaire revealed that there was more variance across groups in perceived evaluative style than in preferred evaluative style. However, statistically significant differences in perceived evaluative style across groups were only found for performance development ($p < .10$), not for performance assessment and interpersonal evaluation. As to within-group differences, the results indicated that for all three perceived evaluative style dimensions mean and median within-group agreement coefficients were higher than .70, which indicates that there was considerable agreement on perceived evaluative style within groups. As expected, there was less within-group agreement on perceived evaluative style than on preferred evaluative style, which indicates that despite high within-group agreement coefficients there were considerable within-group differences between subordinates. This finding was supported by a graphical representation of within-group differences using Categorical Principal Component Analysis (CATPCA). Taken as a whole, these findings indicate that differences in perceived evaluative style can be partly explained by differences in leaders to whom subordinates report (between-group differences), but a significant part of the differences occurs at the level of individual subordinates (within-group differences).

3. Contextual factors and perceived evaluative style

At the group-level, I expected that differences in task uncertainty could partly explain differences in perceived evaluative style. However, no significant correlations were found between task uncertainty and the three evaluative style subscales (performance assessment, performance development, and interpersonal evaluation). Thus, no evidence was found that leaders adjust their style of evaluation to the (average) level of task uncertainty faced by subordinates. Yet, the lack of significant findings could be attributed to the small sample size, as there were only 11 to 14 groups in the sample⁶⁴.

At the individual subordinate level, the results indicated that within-group differences in performance assessment were significantly related to tenure as measured by the period that a respondent has been working under his current superior ($\tau = .336$, $p = .002$). However, no other significant relationships were found between within-group differences in the three subscales of perceived evaluative style and age, tenure, and education of subordinates. Thus,

⁶⁴ The number of groups depends on whether groups are defined as subordinates reporting to the same superior or as subordinates *from the same function group* (i.e. marketing managers, production managers A, production managers B, production heads) reporting to the same superior. In three groups reporting to the

no evidence was found that superiors may adjust their style of evaluation to subordinate characteristics such as age, tenure, and education. The interview data provided some evidence that age, tenure, and education are not important in itself, but only in relation to actual performance, as at least some superiors intend to adapt their style of evaluation to individual subordinates' actual performance relative to working experience and age.

4. Perceived leadership style and perceive evaluative style

Perceived leadership style and perceived evaluative style are significantly related. The results indicated that performance assessment was significantly positively related to supportive leadership ($p < .01$) and active management-by-exception ($p < .10$), but not with teambuilding. Performance development was significantly positively related to supportive leadership ($p < .01$), active management-by-exception ($p < .01$), and teambuilding ($p < .01$). Intepersonal evaluation was significantly positively related to supportive leadership ($p < .01$), active management-by-exception ($p < .01$), and teambuilding ($p < .05$).

5. Perceived evaluative style and managerial behaviour

Perceived evaluative style is significantly related to fairness of evaluation and trust in superior, but there is no relationship with job satisfaction, and with job related tension. More specifically, the results indicated that performance assessment was significantly positively related to fairness of evaluation ($p < .05$). Although the total effect of performance assessment on trust in superior was not significant, performance assessment had a significant positive indirect effect on trust in superior through fairness of evaluation. Performance development was not significantly related to fairness of evaluation, but was significantly positively related to trust in superior ($p < .01$). Interpersonal evaluation was significantly positively related to fairness of evaluation ($p < .01$), and to trust in superior ($p < .10$). Furthermore, fairness of evaluation was found to fully mediate the relationship between interpersonal evaluation and trust in superior.

The interview data provided further support for the findings from the written questionnaire that there was no relationship between perceived evaluative style and job satisfaction and job related tension, as there was hardly any evidence that perceived evaluative style was linked to functional (learning) or dysfunctional behaviour (myopia).

6. The role of the organisational context, and the design of the evaluative system in particular

Using a contextual approach, I have argued that the findings on the previous five topics can only be fully explained by giving account to the organisational context in which this study was carried out. There are several important features of the particular organisational context of VDBN that help to understand the results described above.

same superior subordinates came from two different function groups (see § 6.5 for the sample characteristics).

First, the theoretical framework and propositions were applied and tested empirically to an organization in which a new evaluative system had been implemented just a few months before the data were collected. There appeared to be differences between units in the degree to which superiors had implemented the new system. Additionally, the data indicated that most subordinates had high expectations of the new system and thought it was a better system than in the past. This may explain why the underlying dimensions of evaluative style were so closely related to the performance evaluation system.

Second, an important general aspect in studying evaluative style is the degree of freedom that the formal design of the system allows superiors and subordinates in using the system. Within the organisation I studied, the new evaluative system was very detailed, and consisted of two parts: a Variable Pay Schedule (VPS) and a Performance Development Plan (PDP). The first part, VPS, determined the level of the bonus. This part of the system did not allow any interpretation from either the superior or the subordinate. It was simple: targets were targets, and they were met or not. The second part, PDP, described in detail the format against which the development of people should be assessed. Profiles were made for each function and work level. Yet, this part allowed the superiors some discretion in assessing and evaluating a subordinate's performance. This too may explain why the differences in evaluative style as identified in this study seemed to refer primarily to the degree in which the new system was implemented, rather than to, for example, the choice of performance measures on which performance was assessed. Both superiors and subordinates generally accepted the format of evaluations implicated by the formal design of the evaluative system, which also prescribed the criteria of evaluation.

Third, the data revealed that dysfunctional behaviours as a consequence of performance evaluation in this particular organisation were not a major concern. An important explanation for this could be the finding that subordinates distinguished between targets, which were used to determine their financial bonuses, which they referred to as VPS, and (gaps in) competencies and skills - which they commonly referred to as PDP. While the VPS was designed and used to aid the organisation to reach its objectives, the PDP system was designed and used to help individual managers in reaching their personal (career) goals. The bonus system could cause short-term pressure, and, accordingly, a short-term focus and dysfunctional behaviours, and in the past it did. These short-term pressures were released by the explicit focus of the (newly developed and implemented) PDP-system on competencies and skills, which were seen as much more important from a career perspective than the possibility of obtaining a bonus. Therefore, although subordinates could provide examples of dysfunctional behaviours, most subordinates explicitly indicated that these behaviours were a bigger problem in the past. Subordinates expected that the PDP-system would reduce dysfunctional behaviours for two reasons. First, people would stay in their jobs longer than in

the past, when it was not uncommon for managers to change jobs every two or three years. Second, the short-term pressure induced by the variable pay schedule in the past was reduced because it was now complemented with an emphasis on competencies and skills. Furthermore, the financial bonus that could be earned was not seen as very important to the managers in this study. The maximum gross bonus a manager could earn was 15% of base salary, while it was more likely that they would get only 5%. Since the Netherlands has a rather high marginal income tax rate (42%-52% for most of these managers), the bonus was not very attractive to the managers in my sample.

Fourth, our data also indicated that in general the subordinates accepted the evaluative system that integrated VPS and PDP as a fair and good system. Yet, many subordinates stated that the manner in which their leader used the system could be improved. An important improvement would be that superiors give more attention to the follow-up on gaps in competencies and skills that have been identified in the PDP-form at the beginning of the year. This requires more active coaching from the leader, which includes providing subordinates with feedback on a more continuous basis. Indeed, this seemed to be the main reason why subordinates did not link functional behaviours -such as learning and innovative behaviour- to performance evaluation. Many subordinates felt that they had learned by doing rather than by the PDP-system. To achieve this learning by doing, again the coaching role of leaders seemed to be important, and thus it is this aspect that in the opinion of subordinates should be improved. Still, despite ways for improvement, the data did not provide evidence that subordinates disagreed with overall performance evaluation.

Fifth, the organisational culture at VDBN is very informal and open, as mentioned in § 6.4. Subordinates do receive a lot of informal feedback, and thus, despite the lack of formal attention for the PDP-system, they may expect or trust that in the end their evaluation will be positive. Since VDBN (Unilever) is a lifetime career organisation, and managers changed jobs within the organisation quite often, i.e. every two to three years, managers in our study had worked for several different leaders, and thus had many chances of receiving informal feedback on their performance. It is likely that this informal feedback has reduced the importance the managers attached to the formal annual performance evaluation meeting. Additionally, given the culture and lifetime career type of organisation, performance over just one year is probably not as important as cumulative performance records stretching over many years.

Sixth, our data indicated that subordinates were aware of their own responsibility for their development. Indeed, the system does not allow subordinates to be passive and wait for a one-way assessment of their performance by their superior, but rather stimulates them to recognise and take their own responsibility. This explanation is aligned with the strong (strategical) emphasis within Unilever and VDBN on personal growth of employees and on

initiatives as Total Productive Maintenance (see § 6.2). It is part of (a change in) the organisational culture at VDBN. Against this background, it is not surprising that many subordinates felt that their superior should have a coaching role. The two empirically found subscales of perceived evaluative style, performance development and performance assessment, seem to capture aspects of this coaching behaviour. This is also supported by the finding that both subscales are positively correlated with supportive leadership. Thus, it is not surprising that a higher score on performance development and/or performance assessment was associated with higher scores on fairness of evaluation and trust in superior.

Seventh, the organisation studied was probably atypical because as indicated above the organisation relied heavily on procedures and systems, and formal evaluation forms. Yet, the organisational climate was very informal and open, with direct communication. Given the sensitivity of the topic studied, and the fact that the organisation was in the middle of implementing a new evaluative system, the high level of cooperation that I received in carrying out this study is surprising. The Human Resource Department helped me in contacting possible respondents, and provided documents. None of the interviewees objected to the interviews being tape recorded, and there were no signs that any of the interviewees felt distracted by the presence of the tape recorder. All respondents asked to participate in the interviews, did participate. The response rate of the written questionnaire was quite high as well.

Eighth, the subordinates in this study were managers, but most of them bore no unit responsibility and did not have subordinates themselves. As such the subordinates were at the lowest managerial level within this organisation. This may have affected the relevant dimensions of evaluative style that were found to be relevant in this study.

9.3 Discussion

Although this thesis is grounded within the RAPM-literature in terms of the issues that are being explored, the theoretical and the empirical part of this thesis illustrate that research on evaluative style should move beyond the concepts and assumptions used in existing RAPM-literature. The perspective on evaluative style in this thesis is much broader than in most RAPM-studies, and as such this study deviates from earlier research and extends earlier research in important and meaningful ways.

First, this research project contributes to a better understanding of evaluative style by addressing the criticisms that have been vented on earlier conceptualisations and measurements of evaluative style in the RAPM-literature. The conclusions from the literature review of existing measures of evaluative style, the findings from the pilot study reported in chapter 3, and the findings from the in-depth field study reported in the chapters 6 through 9

consistently show the need to revise the conceptualisation and measurement of evaluative style in future research. I see two general dimensions of evaluative style that need specific attention in future research. The first dimension addresses the evaluative focus of the superior (e.g. budgets, other quantitative targets, short or long-term targets, etc.). The second dimension addresses the superior's way of handling the evaluation process (e.g. rigid or flexible, fixing blame, using it as a learning opportunity, etc.). Building on these two dimensions, there is a need for studies that assess how superiors use specific performance measures in different ways within a particular organisational context. This is important, as designing and implementing performance measurement systems is just the first step. The next step is how managers actually use these systems to influence other members of the organisation. These human interactions between employees at the same or different hierarchical levels of an organisation are essential to understand in order to be able to judge the effectiveness of management control. Building on the RAPM literature, the findings of this study indicate that the design of performance evaluation systems matters, but that the effectiveness of these systems cannot be assessed without considering how the system is used in practice. In this perspective the effectiveness of an evaluative style is determined by the quality of a superior's response to⁶⁵ (compensation for) the degree of "fit" between control system design and the task it has to fulfil in its immediate organisational context. Unfortunately and surprisingly, while this notion laid the foundation of the RAPM literature decades ago, over time this behavioural emphasis has disappeared to the background.

Second, this research project contributes to a better understanding of evaluative style by addressing the level-of-analysis issue raised by Otley (1978) that has been neglected in later RAPM studies too. This study encompasses both leaders (superiors) and subordinate managers in contrast to many RAPM studies that only include subordinate managers. Sampling individuals from several (comparable) units with different unit managers within a particular organisation enables a better understanding of why subordinate managers differ in their perception of evaluative styles, both within and across leaders. Are differences in perceptions based on actual differences between leaders? Do leaders evaluate all their subordinate managers in the same way or do they evaluate different subordinate managers in different ways? If so, why? Scientifically, if evaluative style is (partly) a group level phenomenon, i.e. based on differences between leaders, the results of data analysis at the level of individual managers - if the data contains two or more managers who report to the same leader- could be seriously distorted by unmeasured group effects (Shriesheim et al., 2001). In the theoretical part of this study, this topic was identified as an important topic for future research. Although some propositions were developed to explore this topic

⁶⁵ By using the word "response", I do not intend to suggest that evaluative style is necessarily a conscious choice of the superior, i.e. intended behaviour. This in fact is one of the issues that has remained largely unexplored despite decades of research on evaluative style, but that future research on evaluative style needs to address,

empirically, and although the topic was addressed empirically, it should be noted that in the empirical part of the study this topic has not received the attention that it deserves. Instead, this study has mainly focused on the perception of subordinate managers, without considering the relationship between the perception of the subordinate and actual behaviours of the superior. There is certainly a need for studies on evaluative style at the dyadic level, and for studies using a multi-level approach.

Third, this research project contributes to a better understanding of the relationship between evaluative style and managerial behaviour by giving explicit attention to the organisational context. While survey research has been the dominant research method in the existing RAPM literature, sampling managers from different organisations, findings from these studies are hard to interpret -apart from other problems- because differences in evaluative style are not distinguishable from differences in evaluative system. There is a significant difference between studying the behavioural effects of management control system's **design** and the behavioural effects of management control system's **use**. In investigations of the effects of evaluative style we should control for the adequacy or inadequacy of a performance evaluation system. This way, conceptualisations of evaluative style refer to differences in how superiors use such a performance evaluation system and the information from this system. In fact, this has been the major contribution and implication of the earlier studies on evaluative style. Hopwood (1973, p. 195) for example concludes: "Accounting information does not in and of itself pose a threat to the members of an organisation. Rather, the perception of threat and the consequent defensive behaviours which may be dysfunctional to the organisation as a whole are dependent upon the manner in which the information is seen as being used." This study shows theoretically and empirically that to understand the relationship between evaluative style and managerial behaviour, the design of the evaluative system is one of the most important contextual variables to consider. Conducting a field study within a single organisation enables a more in-depth understanding of this relationship than through cross-sectional surveys, because it controls findings from the influence of differences in evaluative system, while it simultaneously includes the evaluative system as an important contextual variable by enabling a detailed description of the evaluative system. This has been illustrated in the previous section where I described how in the particular organisation studied the organisational context helps to understand the empirically found relationships between evaluative style and several outcome variables.

Fourth, this study is one of the few studies in the RAPM field that includes superiors and subordinates, one of the few studies that uses qualitative data and quantitative data, and one of the few studies that addresses both consequences and antecedents of evaluative style. As such, this study provides a useful perspective, both theoretically and empirically,

as I have argued earlier in this chapter.

for further research on any of the six topics that were identified from the literature review. In this way, the shortcomings of the RAPM-literature can be avoided, while at the same time the relevance of the contribution from earlier studies is recognised.

From a more practical point of view, the topic of this research project is important as it addresses the effectiveness of performance evaluation. Many organisations have staff departments where people are being paid to develop, implement and maintain all kinds of systems that are intended to influence and motivate participants to behave in the best interest of the organisation. High investments are made in sophisticated management control systems, performance evaluation systems, accounting information systems, etc. The people that design and implement the system are usually not the people that will actually use the system. The question is whether these systems really affect people in the way that they are intended to, and if not, how the effectiveness of these systems could be enhanced. This study shows that designers and users of systems should work closely together in order to prevent dysfunctional consequences, and that solid knowledge of operations of the organisation would benefit the design. For example, human resource departments and accounting departments may need to work together in order to implement performance measurement systems and performance evaluation systems that are aligned with the strategic objectives of the organisation.

9.4 Limitations

The findings of this study should be interpreted in the light of its limitations. There are several limitations that are important to mention.

First, in this study I have used the within-group agreement coefficient to assess the level of agreement on perceived evaluative style among subordinates within groups. It should be noted, however, that the within-group agreement coefficient has been criticised on a number of aspects. Schriesheim et al. (1995), for example, indicate that it is problematic to determine the statistical significance of obtained values, and that it is not clear how r_{wg} should be summarised. James et al. (1993) are well aware of some of the "faults" in their coefficient, and they state that there is a need to clearly specify the conditions in which in interrater agreement index is not useful. Furthermore, James et al. (1993) have specified a number of requirements that must be met to use r_{wg} (cf. Schriesheim et al., 1995), such as acceptable psychometric properties, equal-interval measurement, and empirical evidence for the null distribution that is used. Unfortunately, these requirements have not been (fully) met in this research.

Additionally, although this study has addressed the within-group agreement, no effort has been made to assess whether perceived evaluative style is a group-level and/or an individual-

level phenomenon. In the psychological literature on leadership there has been considerable debate on level of analysis issues that may also be relevant for the future study of perceived evaluative style. Several methods have been developed within the leadership literature to assess the (most appropriate) level of analysis from a statistical and/or practical point of view (see Schriesheim et al., 1995). However, the theoretical underpinning of the level at which certain relationships between variables hold, or at which certain phenomena are expected to occur, remains the most important for researchers. There is certainly a need for the development of theory on perceived evaluative style at the group level of analysis, based around concepts of uncertainty, personality of the leader, function, etc. Studying perceived evaluative style as a group-level phenomenon could also help to overcome some of the problems with using performance as an independent variable. For at the level of group, the group performance would be an appropriate measure of performance rather than the more personal and subjective measures of managerial performance.

Indeed, although the framework that was developed and tested in this study does seem to be helpful for future research, I have some doubt about the appropriateness and relevance of the dependent variables used: job satisfaction and job related tension. In the organisational behaviour literature, for example, citizenship behaviours are often used as dependent variables, especially in relation to fairness. Another more important and useful dependent variable would be individual performance, although the measurement of individual performance has always been problematic.

As to fairness of evaluation, Blader & Tyler (2003, p. 111) distinguish different types of procedural justice. They emphasise that 'the control⁶⁶ and group value⁶⁷ models of procedural justice emphasize a different set of concerns with procedures and, consequently, they predict a different set of criteria employees may use for evaluating procedures. This makes distinguishing between them important for attempts by both researchers and practitioners to understand how employees judge fairness'. Improvements in this field on different models of fairness may further our understanding of the behavioural impact of perceived evaluative style. Furthermore, I have focused on what in the organisational justice literature is known as procedural justice. However, distributive justice may be important as well, i.e. not just a focus on whether procedures (i.e. how managers are evaluated) is fair, but also a more explicit focus on outcomes (pay, bonus) could be appropriate in future research.

⁶⁶ According to Blader & Tyler (2003, p. 109), the control model states that 'assessments of how fairly decisions are made are an important determinant of reactions to the outcomes of those processes. In other words, ... people accept outcomes that may be less than what they wanted because they regard the process that led to those outcomes as fair'.

⁶⁷ According to Blader & Tyler (2003, p. 110) the group value model 'stress the relational importance of procedures'.

The results in this study also suggest that there are differences caused by function. This has posed some limitations on the analyses in this study, which are even more severe because of the small overall sample. With larger samples, before conducting principal components analysis, it would be wise to test whether there are similar patterns of intercorrelations between all items for each of the separate function groups. For this purpose, one should compare the different matrices of the intercorrelations, and assess the fit using an index such as NNFI (non-normed fit index) that is available in most statistical packages (LISREL, etc.). (See Den Hartog (1997, p. 153)). This however would require quite large samples, which are quite difficult to realise when limiting the research to single organisations. In any case, this would require doing the field study in rather large organisations, which is a limitation in itself. It could be interesting to study small organizations as well. This is a trade-off that has to be made in individual cases.

A further limitation was that no production heads have been interviewed. This was a mistake, because for the production heads a slightly different evaluative system was in use, thereby providing a different context. Since I have argued for the need to separate design of the evaluative system from the use of the system throughout this thesis, interviews could have provided a description of the differences in evaluation and contexts, which could result in a better understanding of the results from this study.

Finally, the interview data and the data from the written questionnaire were collected simultaneously. With the benefit of hindsight, it would have been better to collect and analyse data from one source first, and then, based on the preliminary findings, focus on a deeper understanding of the findings. With the current research design, it was hard to compare the quantitative and qualitative evidence from the two sources. Additionally, the findings of this study are based on data that was collected on a single moment in time. To understand the underlying processes, it would have been better to conduct interviews over a longer period of time.

9.5 Possibilities for further research

In conclusion, this study has identified important topics which provide promising avenues for future research on evaluative style and its behavioural consequences. Further research will be needed in different organizations. I have pointed out some possibilities for further research earlier in this thesis. In this final section I will limit myself to a brief description of four additional research opportunities emerging from (the findings of) this study that in my opinion are worthwhile to investigate.

First, further research on the topics identified in this study will be needed amongst managers that have unit responsibility. That allows exploring the question superiors in performance

evaluation differentiate the performance of the unit (economic performance) from the performance of the manager responsible for the unit (managerial performance), and if so, how and why. Additionally, it is likely that studying subordinate managers with unit responsibility will provide more insight into the role of accounting performance measures in performance evaluation than at lower levels in the hierarchy (see also § 4.3). These topics could not be addressed in this study, because the subordinate managers in my study bore no unit responsibility.

Second, there is a need for longitudinal research. This study, although using different research methods, and different sources of data, was conducted at a single moment in time. As such, it does not address the processes through which performance evaluation and performance evaluation style affect subordinate's behaviour. The study suggests that fairness of evaluation and trust in superior will play a role in these processes. However, how these variables evolve over time has not been addressed. Longitudinal research could add relevant insights to the framework offered in this study and could give more account to changes in organisational context in more detail. For example, organisational level variables such as strategy and organizational climate may not be static concepts. Rather, strategy and organizational climate may change in time. In fact, they may even change as an effect of performance appraisal. Longitudinal research is still very rare in management accounting, probably because it is very time consuming. However, it would greatly benefit our knowledge of how and why accounting and management control in general affect people within organisations, and as part of it, what the role of evaluative style is.

Third, future research is needed to address the role of trust in superior in the framework. In this study trust in superior was treated as a consequence of evaluative style. An alternative would be to consider trust in superior as a more independent variable that affects the effectiveness of evaluative style. The hypothesis to be tested could be as follows: Subordinates who report high levels of trust in superior will report lower levels of dysfunctional and higher levels of functional behaviours associated with effective evaluative styles than subordinates who report low levels of trust in superior. Thus, trust in superior could be a moderator variable rather than the consequence of evaluative style. Yet, if trust in superior is a moderator variable, this assumes no (or a low) correlation between evaluative style and trust in superior. In my study, the correlation was high, and therefore in this study treating trust in superior as a moderator variable is not a likely alternative. However, this could be different in other organisations. Additionally, further research is needed to understand how trust in superior evolves over time. Is it really dependent on evaluative style, or is it dependent upon, for example, the time a subordinate has been working under his current superior? So, does tenure matter in explaining the relationship between evaluative style and trust in superior? Furthermore, does trust in superior increase because a subordinate feels his performance evaluation is fair? Or does he report that he feels his performance evaluation is fair because he trusts his superior? Again, longitudinal research

could help in addressing some of the processes through subordinates' trust in superior is built in organisations.

Fourth, future research is needed in organisations in which it is more likely that dysfunctional behaviours appear due to the manner in which performance is being evaluated. Especially studying not-for-profit organisations and organisations in the public sector could be interesting given the developments in practice in the past few years. Examples could include universities, with their increasing emphasis on performance evaluation in terms of narrow definitions of quality and quantity of publications of researchers, or public sector organisations, such as municipalities and provinces, which are faced with increasing pressure to account for and report performance with just a few, simple, SMART measures. In these type of organisations, there are often multiple stakeholders and there are multiple goals, not just one overall goal such as profit. Therefore, in organisations like these, measuring and assessing performance based on just a few indicators is likely to provide a too narrow focus on performance, and, accordingly, it may lead to dysfunctional behaviours if individuals within these organisations are evaluated on these same measures. In these circumstances, the manner in which superiors use these measures, given the apparent inappropriateness of the formal design of the system, may be an important factor in explaining the functional or dysfunctional responses of subordinates to performance evaluation. On the other hand, over time, the system used to measure performance may also serve as a selection tool, attracting people with specific competencies, skills, and personalities, and discouraging others. It is only in the long run that we can assess whether these systems really affect people in the way that they are intended to. These developments show that exploring the consequences and effectiveness of design and use of evaluative system is as important today as it has been in the past decades, and that it will remain important in the future.

References

- Anthony, R. N. (1965), *Planning and Control Systems: A Framework for Analysis*, Boston: Harvard University Division of Research, Graduate School of Business Administration.
- Anthony, R. N. (1988), *The Management Control Function*, Boston, MA: Harvard Business School Press.
- Anthony, R.N., J. Dearden & V. Govindarajan (1992), *Management Control Systems*, 7th ed., Homewood: Irwin.
- Anthony, R.N. & V. Govindarajan (1995), *Management Control Systems*, 8th ed., Chicago: Irwin.
- Argyris, C. (1952), *The Impact of Budgets on People*. New York: Controllershship Foundation.
- Aryee, S., P. S. Budhwar and Z. X. Chen (2002), Trust as a Mediator of the Relationship between Organizational Justice and Work Outcomes: Test of a Social Exchange Model, *Journal of Organizational Behavior* 23 (3): 267-285.
- Bandura, A. and D. H. Schunk (1981), Cultivating Competence, Self-Efficacy, and Intrinsic Interest through Proximal Self-Motivation, *Journal of Personality & Social Psychology* 41 (3): 586-598.
- Baron, R. M. and D. A. Kenny (1986), The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations, *Journal of Personality and Social Psychology* 51 (6): 1173-1182.
- Bass, B. M. and R. M. Stogdill (1990), *Bass & Stogdill's Handbook of Leadership: Theory, Research, and Managerial Applications*. New York: Free Press.
- Birnberg, J. G., L. Turopolec and S. M. Young (1983), The Organizational Context of Accounting, *Accounting, Organizations and Society* 8 (2-3): 111-129.
- Blader, S. L. and T. R. Tyler (2003), What Constitutes Fairness in Work Settings: A Four-Component Model of Procedural Justice, *Human Resource Management Review* 13: 107-126.
- Brah, S. A. and W. K. Chong (2004), Relationship between Total Productive Maintenance and Performance, *International Journal of Production Research* 42 (12): 2383-2401.
- Briers, M. and M. Hirst (1990), The Role of Budgetary Information in Performance Evaluation, *Accounting, Organizations and Society* 15 (4): 373-398.
- Brownell, P. (1981), Participation in Budgeting, Locus of Control and Organizational Effectiveness, *The Accounting Review* 56 (4): 844-860.
- Brownell, P. (1982a), A Field Study Examination of Budgetary Participation and Locus of Control, *The Accounting Review* 57 (4): 766-777.
- Brownell, P. (1982b), The Role of Accounting Data in Performance Evaluation, Budgetary Participation, and Organizational Effectiveness, *Journal of Accounting Research* 20 (1): 12-27.

- Brownell, P. (1985), Budgetary Systems and the Control of Functionally Differentiated Organizational Activities, *Journal of Accounting Research* 23 (2): 502-512.
- Brownell, P. (1987), The Role of Accounting Information, Environment and Management Control in Multi-National Organizations, *Accounting and Finance* 27 (1): 1-16.
- Brownell, P. and A. S. Dunk (1991), Task Uncertainty and Its Interaction with Budgetary Participation and Budget Emphasis: Some Methodological Issues and Empirical Investigation, *Accounting, Organizations and Society* 16 (8): 693-703.
- Brownell, P. and M. Hirst (1986), Reliance on Accounting Information, Budgetary Participation, and Task Uncertainty: Tests of a Three-Way Interaction, *Journal of Accounting Research* 24 (2): 241-249.
- Brownell, P. and M. McInnes (1986), Budgetary Participation, Motivation, and Managerial Performance, *The Accounting Review* 61 (4): 587-600.
- Bruns, W. J., Jr. and J. H. Waterhouse (1975), Budgetary Control and Organization Structure, *Journal of Accounting Research* 13 (2): 177-203.
- Buja, A. and N. Eyuboglu (1992), Remarks on Parallel Analysis, *Multivariate Behavioral Research* 27 (Issue 4): 509-540.
- Caplan, E.H. (1966), The Behavioral Implications of Management Accounting, *Management International Review* 32, Special Issue (1), p. 92-102, 1992.
- Chapman, C. S. (1998), Accountants in Organisational Networks, *Accounting, Organizations and Society* 23 (8): 737-766.
- Chenhall, R. H. (2003), Management Control Systems Design within Its Organizational Context: Findings from Contingency-Based Research and Directions for the Future, *Accounting, Organizations and Society* 28 (2,3): 127-168.
- Chenhall, R. H. and K. Langfield-Smith (1998), The Relationship between Strategic Priorities, Management Techniques and Management Accounting: An Empirical Investigation Using a Systems Approach, *Accounting, Organizations and Society* 23 (3): 243-264.
- Covaleski, M. A., M. W. Dirmsmith and S. Samuel (2003), Changes in the Institutional Environment and the Institutions of Governance: Extending the Contributions of Transaction Cost Economics within the Management Control Literature, *Accounting, Organizations and Society* 28 (5): 417-441.
- Dansereau, F., F. J. Yammarino and S. E. Markham (1995), Leadership: The Multiple-Level Approaches, *The Leadership Quarterly* 6 (2): 97-109.
- DeCoster, D. T. and J. P. Fertakis (1968), Budget-Induced Pressure and Its Relationship to Supervisory Behavior, *Journal of Accounting Research* 6: 237-246.
- Dillman, D. A. (1978), *Mail and Telephone Surveys: The Total Design Method*, New York: John Wiley & Sons.
- Dillman, D. A. (2000), *Mail and Internet Surveys: The Tailored Design Method*, New York: John Wiley & Sons.
- Drazin, R. and A. H. Van de Ven (1985), Alternative Forms of Fit in Contingency Theory, *Administrative Science Quarterly* 30 (4): 514-539.

- Drury, J.C. (1992), *Management and Cost Accounting*, 3rd ed., London: Chapman & Hall.
- Duncan, R. B. (1972), Characteristics of Organizational Environments and Perceived Environmental Uncertainty, *Administrative Science Quarterly*, 17: 313-327.
- Dunk, A. S. (1989), Budget Emphasis, Budgetary Participation and Managerial Performance: A Note, *Accounting, Organizations and Society* 14 (4): 321-324.
- Dunk, A. S. (1990), Budgetary Participation, Agreement on Evaluation Criteria and Managerial Performance: A Research Note, *Accounting, Organizations and Society* 15 (3): 171-178.
- Dunk, A. S. (1992), Reliance on Budgetary Control, Manufacturing Process Automation and Production Subunit Performance: A Research Note, *Accounting, Organizations and Society* 17 (3,4): 195-203.
- Dunk, A. S. (1993a), The Effect of Budget Emphasis and Information Asymmetry on the Relation between Budgetary Participation and Slack, *The Accounting Review* 68 (2): 400-410.
- Dunk, A. S. (1993b), The Effects of Job-Related Tension on Managerial Performance in Participative Budgetary Settings, *Accounting, Organizations and Society* 18 (7,8): 575-585.
- Dunk, A. S. (2003), Moderated Regression, Constructs and Measurement in Management Accounting: A Reflection, *Accounting, Organizations and Society* 28 (7-8): 793-802.
- Dijk, M.P. van (1992), Participatie van Managers in het Budgetproces: Een Evaluatie van Empirisch Onderzoek, *Maandblad voor Accountancy en Bedrijfseconomie*, 66 (3): 108-115.
- Emsley, D. (2001), Budget-Emphasis in Performance Evaluation and Managers' Job Related Tension: The Moderating Effect of Information Completeness, *The British Accounting Review* 33 (3): 399-419.
- Erdogan, B. (2002), Antecedents and Consequences of Justice Perceptions in Performance Appraisals, *Human Resource Management Review* 12 (4): 555-578.
- Ezzamel, M. (1990), The Impact of Environmental Uncertainty, Managerial Autonomy and Size on Budget Characteristics, *Management Accounting Research* 1 (1): 181-197.
- Ezzammel, M. & H. Hart (1987), *Advanced Management Accounting: An Organisational Emphasis*, London: Cassell Educational Limited.
- Farh, J.-L., P. C. Earley and S.-C. Lin (1997a), Impetus for Action: A Cultural Analysis of Justice and Organizational Citizenship Behavior in Chinese Society, *Administrative Science Quarterly* 42 (3): 421-444.
- Fisher, J. (1995), Contingency-Based Research on Management Control Systems: Categorization by Level of Complexity, *Journal of Accounting Literature* 14: 24-53.
- Flamholtz, E. G. (1983), Accounting, Budgeting and Control Systems in Their Organizational Context: Theoretical and Empirical Perspectives, *Accounting, Organizations and Society* 8 (2,3): 153-169.

- Fleishman, E. A. (1973), Twenty Years of Consideration and Structure, In: Fleishman, E.A. and J. G. Hunt (eds), *Current Developments in the Study of Leadership*, Carbondale: Southern Illinois University Press.
- Folger, R. and M. A. Konovsky (1989), Effects of Procedural and Distributive Justice on Reactions, *Academy of Management Journal* 32 (1): 115-130.
- Ford, M. E. (1992), *Motivating Humans: Goals, Emotions, and Personal Agency Beliefs*, Newbury Park: Sage.
- Frucot, V. and W. T. Shearon (1991), Budgetary Participation, Locus of Control, and Mexican Managerial Performance and Job Satisfaction, *The Accounting Review* 66 (1): 80-99.
- Fulk, J., A. P. Brief and S. H. Barr (1985a), Trust-in-Supervisor and Perceived Fairness and Accuracy of Performance Evaluations, *Journal of Business Research* 13 (4): 301-313.
- George, J. M. (1990), Personality, Affect, and Behavior in Groups, *Journal of Applied Psychology* 75 (2): 107-116.
- George, J. M. and L. R. James (1993), Personality, Affect, and Behavior in Groups Revisited: Comment on Aggregation, Levels of Analysis, and a Recent Application of within and between Analysis, *Journal of Applied Psychology* 78 (5): 798-804.
- Gerdin, J. and J. Greve (2004), Forms of Contingency Fit in Management Accounting Research--a Critical Review, *Accounting, Organizations and Society* 29 (3-4): 303-326.
- Govindarajan, V. (1984), Appropriateness of Accounting Data in Performance Evaluation: An Empirical Examination of Environmental Uncertainty as an Intervening Variable, *Accounting, Organizations and Society* 9 (2): 125-135.
- Govindarajan, V. (1988), A Contingency Approach to Strategy Implementation at the Business-Unit Level: Integrating Administrative Mechanisms with Strategy, *Academy of Management Journal* 31 (4): 828-853.
- Govindarajan, V. and A. K. Gupta (1985), Linking Control Systems to Business Unit Strategy: Impact on Performance, *Accounting, Organizations and Society* 10 (1): 51-66.
- Greenberg, P.S., R.H. Greenberg & H. Nouri (1994), Participative budgeting: a meta-analytic examination of methodological moderators, *Journal of Accounting Literature* 13: 117-141.
- Groot, T.L.M.C. (1994), *Besturen is vooruitzien: de opkomst van management Accounting in wetenschap en praktijk*, Groningen: Wolters-Noordhoff.
- Gupta, A. K. (1987), SBU Strategies, Corporate-SBU Relations, and SBU Effectiveness in Strategy Implementation, *Academy of Management Journal* 30 (3): 477-500.
- Hansen, D.R. & M.M. Mowen (1994), *Management Accounting*, 3rd ed., Cincinnati: South-Western.
- Hansen, S., D. Otley, and W. Van der Stede (2003), Practice Development in Budgeting: An Overview and Research Perspective, *Journal of Management Accounting Research* 15: 95-116.

- Harrison, G. L. (1992), The Cross-Cultural Generalizability of the Relation between Participation, Budget Emphasis and Job Related Attitudes, *Accounting, Organizations and Society* 17 (1): 1-15.
- Harrison, G. L. (1993), Reliance on Accounting Performance Measures in Superior Evaluative Style - the Influence of National Culture and Personality, *Accounting, Organizations and Society* 18 (4): 319-339.
- Hartmann, F. G. H. (1997), *Accounting for Performance Evaluation: Effects of Uncertainty on the Appropriateness of Accounting Performance Measures*, Maastricht: Universiteit Maastricht.
- Hartmann, F. G. H. (2000), The Appropriateness of RAPM: Toward the Further Development of Theory, *Accounting, Organizations and Society* 25 (4-5): 451-482.
- Hartmann, F. G. H. and F. Moers (1999), Testing Contingency Hypotheses in Budgetary Research: An Evaluation of the Use of Moderated Regression Analysis, *Accounting, Organizations and Society* 24 (4): 291-315.
- Hartog, D. N. den, (1997), *Inspirational Leadership*, Enschede: Ipskamp.
- Hersey, P. and K. H. Blanchard (1993), *Management of Organizational Behavior: Utilizing Human Resources*. Englewood Cliffs: Prentice Hall.
- Hirst, M. K. (1981), Accounting Information and the Evaluation of Subordinate Performance: A Situational Approach, *The Accounting Review* 56 (4): 771-784.
- Hirst, M. K. (1983), Reliance on Accounting Performance Measures, Task Uncertainty, and Dysfunctional Behavior: Some Extensions, *Journal of Accounting Research* 21 (2): 596-605.
- Hirst, M. K. (1987), The Effects of Setting Budget Goals and Task Uncertainty on Performance: A Theoretical Analysis, *The Accounting Review* 62 (4): 774-784.
- Hirst, M. K. and P. Yetton (1984), Influence of Reliance on Accounting Performance Measures and Job Structure on Role Ambiguity for Production and Non-Production Jobs, *Australian Journal of Management* 9 (1), 53-61.
- Hofstede, G. (1980), *Culture's Consequences: International Differences in Work-related Values*, Beverly Hills, CA and London: Sage.
- Hofstede, G. H. (1968), *The Game of Budget Control*, Assen: Van Gorcum.
- Hopwood, A. G. (1972a), An Empirical Study of the Role of Accounting Data in Performance Evaluation, *Journal of Accounting Research* 10: 156-182.
- Hopwood, A. G. (1972b), [Discussion of an Empirical Study of the Role of Accounting Data in Performance Evaluation]: A Reply, *Journal of Accounting Research* 10: 189-193.
- Hopwood, A. G. (1973), *An Accounting System and Managerial Behaviour*, Farnborough: Saxon House.
- Hopwood, A. G. (1974), Leadership Climate and the Use of Accounting Data in Performance Evaluation, *The Accounting Review* 49 (3): 485-495.
- Horn, J. L. (1965), A Rationale and Test for the Number of Factors in Factor Analysis, *Psychometrika* 30: 179-185.

- Horngren, C.T., G. Foster & S.M. Datar (1994), *Cost Accounting: a managerial emphasis*, 8th ed., Englewood Cliffs: Prentice-Hall.
- Horgren, C.T. & G.L. Sundem (1987), *Introduction to Management Accounting*, 7th ed., Englewood Cliffs: Prentice-Hall.
- Imoisili, O. A. (1989), The Role of Budget Data in the Evaluation of Managerial Performance, *Accounting, Organizations and Society* 14 (4): 325-335.
- James, L. R., R. G. Demaree and G. Wolf (1984), Estimating within-Group Interrater Reliability with and without Response Bias, *Journal of Applied Psychology* 69 (1): 85-98.
- James, L. R., R. G. Demaree and G. Wolf (1993), R_{wg} : An Assessment of within-Group Interrater Agreement, *Journal of Applied Psychology* 78 (2): 306-309.
- Johnson, H.T. & R.S. Kaplan (1987), *Relevance Lost: the rise and fall of management accounting*, Boston: Harvard Business School Press.
- Judge, T. A., R. F. Piccolo and R. Ilies (2004), The Forgotten Ones? The Validity of Consideration and Initiating Structure in Leadership Research, *Journal of Applied Psychology* 89 (1): 36-51.
- Kahn, R. L. (1972), Discussion of an Empirical Study of the Role of Accounting Data in Performance Evaluation, *Journal of Accounting Research* 10: 183-186.
- Kahn, R. L., D. M. Wolfe, R. P. Quinn, J. D. Snoek and R. A. Rosenthal (1964). *Organizational Stress: Studies in Role Conflict and Ambiguity*, New York: Wiley.
- Kaplan, R. S. and A. A. Atkinson (1989), *Advanced Management Accounting*, New Jersey: Prentice-Hall.
- Kaplan, R. S. and D. P. Norton (1992), The Balanced Scorecard--Measures That Drive Performance, *Harvard Business Review* 70 (1): 71-79.
- Kaplan, R.S. and D.P. Norton (1993), Putting the Balanced Scorecard to Work, *Harvard Business Review* 71 (September-October): 134-142.
- Kaplan, R. S. and D. P. Norton (1996a), Using the Balanced Scorecard as a Strategic Management System, *Harvard Business Review* (1): 75-85.
- Kaplan, R. S. and D. P. Norton (1996b), *The Balanced Scorecard: Translating Strategy into Action*, Boston: Harvard Business School Press.
- Kaplan, R.S. and D.P. Norton (2001a), Transforming the Balanced Scorecard from Performance Measurement to Strategic Management: Part I, *Accounting Horizons* 15: 87-104.
- Kaplan, R.S. and D.P. Norton (2001b), Transforming the Balanced Scorecard from Performance Measurement to Strategic Management: Part II, *Accounting Horizons* 15: 147-160.
- Kenis, I. (1979), Effects of Budgetary Goal Characteristics on Managerial Attitudes and Performance, *The Accounting Review* 54: 707-721.
- Kerr, S. (1983), "Accounting, Budgeting and Control Systems in Their Organizational Context: Theoretical and Empirical Perspectives": Comments by the Discussant, *Accounting, Organizations and Society* 8 (2,3): 171-174.

- Kim, J.-o. and C. W. Mueller (1978), *Factor Analysis: Statistical Methods and Practical Issues*. Beverly Hills: Sage Publications.
- Kozlowski, S. W. J. and K. Hattrup (1992), A Disagreement About within-Group Agreement, *Journal of Applied Psychology* 77 (2): 161-167.
- Kren, L. (1992), Budgetary Participation and Managerial Performance: The Impact of Information and Environmental Volatility, *The Accounting Review* 67 (3): 511-526.
- Landy, F. J., J. L. Barnes and K. R. Murphy (1978), Correlates of Perceived Fairness and Accuracy of Performance Evaluation, *Journal of Applied Psychology*, 63: 751-754.
- Landy, F. J. and J. L. Farr (1980), Performance Rating, *Psychological Bulletin*, 87: 72-107.
- Langfield-Smith, K. (1997), Management Control Systems and Strategy: A Critical Review, *Accounting, Organizations and Society* 22 (2): 207-232.
- Lau, C. M. and C. Buckland (2001), Budgeting-Role of Trust and Participation: A Research Note, *Abacus* 37 (3): 369-388.
- Lau, C. M., L. C. Low and I. R. C. Eggleton (1995), The Impact of Reliance on Accounting Performance Measures on Job-Related Tension and Managerial Performance: Additional Evidence, *Accounting, Organizations and Society* 20 (5): 359-381.
- Lau, C. M., L. C. Low and I. R. C. Eggleton (1997), The Interactive Effect of Budget Emphasis, Participation and Task Difficulty on Managerial Performance: A Cross-Cultural Study, *Accounting, Auditing & Accountability Journal* 10 (2): 175-197.
- Lau, C. M. and J. J. Tan (1998), The Impact of Budget Emphasis, Participation and Task Difficulty on Managerial Performance: A Cross-Cultural Study of the Financial Services Sector, *Management Accounting Research* 9 (2): 163-183.
- Lau, C. M. and S. L. C. Tan (2006), The Effects of Procedural Fairness and Interpersonal Trust on Job Tension in Budgeting, *Management Accounting Research* 17 (2): 171-186.
- Linting, M., P. J. F. Groenen and J. J. Meulman (2004). Stability of Categorical Principal Components Analysis: An Empirical Study: Working paper, Data Theory Group, Leiden University, The Netherlands.
- Locke, E. A., G. P. Latham and M. Erez (1988), The Determinants of Goal Commitment, *Academy Of Management Review* 13 (1): 23-39.
- Lynch, R.L. & K.F. Cross, (1991), *Measure Up! Yardsticks for Continuous Improvement*, Basil Blackwell: Cambridge.
- Lycke, L. (2003), Team Development When Implementing TPM, *Total Quality Management & Business Excellence* 14 (Issue 2): 205-213.
- Machin, J.L.J. (1983), Management Control Systems: Whence and Whither?, In: *Readings in Accounting for Management Control*, Emmanuel, C., D.T. Otley & K. Merchant (editors) (1992), London: Chapman & Hall, pp. 9-27.
- Macintosh, N. (1985), Human Relations and Budgeting Systems, In: *Readings in Accounting for Management Control*, Emmanuel, C., D.T. Otley & K. Merchant (editors) (1992), London: Chapman & Hall, pp. 215-231.

- Macintosh, N. B. (1994), *Management Accounting and Control Systems: An Organisational and Behavioural Approach*. Chichester: Wiley.
- Macintosh, N. B. and R. L. Daft (1987), Management Control Systems and Departmental Interdependencies: An Empirical Study, *Accounting, Organizations and Society* 12 (1): 49-61.
- Marginson, D. E. W. (1999), Beyond the Budgetary Control System: Towards a Two-Tiered Process of Management Control, *Management Accounting Research* 10 (3): 203-230.
- Meng, X.-L., Rosenthal, R. and Rubin, D. B. (1992), Comparing Correlated Correlation Coefficients, *Psychological Bulletin* 111 (1): 172-175.
- Merchant, K. A. (1981), The Design of the Corporate Budgeting System: Influences on Managerial Behavior and Performance, *The Accounting Review* 56 (4): 813-829.
- Merchant, K. A. (1984), Influences on Departmental Budgeting: An Empirical Examination of a Contingency Model, *Accounting, Organizations and Society* 9 (3-4): 291-307.
- Merchant, K. A. (1990), The Effects of Financial Controls on Data Manipulation and Management Myopia, *Accounting, Organizations and Society* 15 (4): 297-313.
- Merchant, K. A. (1998), *Modern Management Control Systems: Text and Cases*. Upper Saddle River: Prentice Hall.
- Merchant, K. A. and W. A. Van der Stede (2003), *Management Control Systems: Performance Measurement, Evaluation, and Incentives*. New York: FT Prentice Hall.
- Meulman, J. J., W. J. Heiser and S. Inc. (2001). SPSS Categories 11.0. Chicago: SPSS, Inc.
- Mia, L. and R. H. Chenhall (1994), The Usefulness of Management Accounting Systems, Functional Differentiation and Managerial Effectiveness, *Accounting, Organizations and Society* 19 (1): 1-13.
- Miles, M. B. and A. M. Huberman (1994). *Qualitative Data Analysis: An Expanded Sourcebook*, Thousand Oaks: Sage.
- Nakajima, S. (1989), *TPM Development Program: Implementing Total Productive Maintenance*, Cambridge, Mass.: Productivity Press.
- NAA - The National Association of Accountants (1990), *Statements on Management Accounting*, Englewood Cliffs: Prentice Hall.
- Nunnally, J. C. (1978), *Psychometric Theory*. New York: McGraw-Hill.
- O Connor, N. G. (1995), The Influence of Organizational Culture on the Usefulness of Budget Participation by Singaporean-Chinese Managers, *Accounting, Organizations and Society* 20 (5): 383-403.
- Otley, D.T. & A.J. Berry (1980), Control, organization and accounting, In: *Readings in Accounting for Management Control*, Emmanuel, C., D.T. Otley & K. Merchant (editors) (1992), London: Chapman & Hall, pp. 28-48.
- Otley, D. and A. Fakiolas (2000), Reliance on Accounting Performance Measures: Dead End or New Beginning? *Accounting, Organizations and Society* 25 (4,5): 497-510.
- Otley, D. and R. M. Pollanen (2000), Budgetary Criteria in Performance Evaluation: A Critical Appraisal Using New Evidence, *Accounting, Organizations and Society* 25 (4,5): 483-496.

- Otley, D. T. (1978), Budget Use and Managerial Performance, *Journal of Accounting Research* 16 (1): 122-149.
- Otley, D. T. (1994), Management Control in Contemporary Organizations: Towards a Wider Framework, *Management Accounting Research* 5 (3-4): 289-299.
- Otley, D. T. and B. J. Pierce (1995), The Control Problem in Public Accounting Firms: An Empirical Study of the Impact of Leadership Style, *Accounting, Organizations and Society* 20 (5): 405-420.
- Parker, L.D., K.R. Ferris & D.T. Otley (1989), *Accounting for the Human Factor*, Sydney: Prentice Hall.
- Pelz, D. C. and F. M. Andrews (1976), *Scientists in Organizations: Productive Climates for Research and Development*, Ann Arbor: Institute for Social Research University of Michigan.
- Pillai, R., C. A. Schriesheim and E. S. Williams (1999), Fairness Perceptions and Trust as Mediators for Transformational and Transactional Leadership: A Two-Sample Study, *Journal of Management* 25 (6): 897-933.
- Pinder, C. C. (1984), *Work Motivation: Theory, Issues, and Applications*, Glenview: Scott Foresman.
- Pintrich, P. R. (1995), Motivation: Ford's Motivational Systems Theory, *Contemporary Psychology* 40 (10): 999-1001.
- Poe, C. D., J. Shearon, Winston T. and R. H. Strawser (1991), Accounting Evaluative Styles and the Contagion Effect in Middle-Managers: An Empirical Study, *Journal of Management Accounting Research* 3: 169-193.
- Pratt, J. and J. Jiambalvo (1981), Relationships between Leader Behaviors and Audit Team Performance, *Accounting, Organizations and Society* 6 (2): 133-142.
- Pratt, J. and J. Jiambalvo (1982), Determinants of Leader Behavior in an Audit Environment, *Accounting, Organizations and Society* 7 (4): 369-379.
- Preacher, K. J. and A. F. Hayes (2004), Spss and Sas Procedures for Estimating Indirect Effects in Simple Mediation Models, *Behavior Research Methods, Instruments, and Computers* 36: 717-731.
- Preacher, K.J. and A.F. Hayes (2006), Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Simple and Multiple Mediator Models, Working paper, downloaded 30 August, 2006 at: <http://www.comm.ohio-state.edu/ahayes/indirect2.pdf>
- Read, W. H. (1962), Upward Communication in Industrial Hierarchies, *Human Relations*, 15: 3-15.
- Ridgway, V. F. (1956), Dysfunctional Consequences of Performance Measurements, *Administrative Science Quarterly* 1 (2): 240-247.
- Ross, A. (1994), Trust as a Moderator of the Effect of Performance Evaluation Style on Job-Related Tension: A Research Note, *Accounting, Organizations and Society* 19 (7): 629-635.

- Ross, A. (1995), Job Related Tension, Budget Emphasis and Uncertainty: A Research Note, *Management Accounting Research* 6 (1): 1-11.
- Schriesheim, C. A., C. C. Coglisier and L. L. Neider (1995), Is It "Trustworthy"? A Multiple-Levels-of-Analysis Reexamination of an Ohio State Leadership Study, with Implications for Future Research, *The Leadership Quarterly* 6 (2): 111-145.
- Shrout, P. E. and N. Bolger (2002), Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations., *Psychological Methods* 7 (4): 422-445.
- Simons, R. (1987), Accounting Control Systems and Business Strategy: An Empirical Analysis, *Accounting, Organizations and Society* 12 (4): 357-374.
- Simons, R. (1990), The Role of Management Control Systems in Creating Competitive Advantage: New Perspectives, *Accounting, Organizations and Society* 15 (1-2): 127-143.
- Simons, R. (1991), Strategic Orientation and Top Management Attention to Control Systems, *Strategic Management Journal* 12 (1): 49-62.
- Simons, R. (1994), How New Top Managers Use Control Systems as Levers of Strategic Renewal, *Strategic Management Journal* 15 (3): 169-189.
- Simons, R. (1995), *Levers of Control: How Managers Use Innovative Control Systems to Drive Strategic Renewal*, Boston: Harvard Business School Press.
- Smith, P. C., L. M. Kendall and C. L. Hulin (1969), *The Measurement of Satisfaction in Work and Retirement*, Chicago: Rand McNally.
- Sobel, M. E. (1982), Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models, *Sociological Methodology* 13: 290-312.
- Swieringa, R. J. and R. H. Moncur (1975), *Some Effects of Participative Budgeting on Managerial Behavior*, New York: National Association of Accountants.
- Thompson, J. D. (1967), *Organizations in Action: Social Science Bases of Administrative Theory*. New York: McGraw-Hill.
- Vagneur, K. and M. Peiperl (2000), Reconsidering Performance Evaluative Style, *Accounting, Organizations and Society* 25 (4,5): 511-525.
- Van de Ven, A. H. and A. L. Delbecq (1974), A Task Contingent Model of Work-Unit Structure, *Administrative Science Quarterly* 19 (2): 183-197.
- Van der Stede, W. A. (2000), The Relationship between Two Consequences of Budgetary Controls: Budgetary Slack Creation and Managerial Short-Term Orientation, *Accounting, Organizations and Society* 25 (6): 609-622.
- Van der Stede, W. A. (2001), Measuring 'Tight Budgetary Control', *Management Accounting Research* 12 (1): 119-137.
- Weiss, D. J., R. V. Dawis, G. W. England and L. H. Lofquist (1967), *Manual for the Minnesota Satisfaction Questionnaire*, Minneapolis: Work Adjustment Project, Industrial Relations Center, University of Minnesota.
- Withey, M., R. L. Daft and W. H. Cooper (1983a), Measures of Perrow's Work Unit Technology: An Empirical Assessment and a New Scale, *The Academy of Management Journal* 26 (1): 45-63.

Wood, R. and A. Bandura (1989), Social Cognitive Theory of Organizational Management, *Academy Of Management Review* 14 (3): 361-384.

Young, D.W. (1994), *Introduction to Financial & Management Accounting: a user perspective*, Boston: South-Western.

Zwick, W. R. and W. F. Velicer (1986), Comparison of Five Rules for Determining the Number of Components to Retain, *Psychological Bulletin* 99 (3): 432-442.

Appendix A: Interview protocols used within Service Group (see § 3.5)

1. Interview protocol for interviewing the branch manager (translated)

1. How long have you been employed by <name of the organisation>?
2. How long have you been employed in your current occupation?
3. What other positions did you hold within this organisation?
4. What is your educational background?
5. How many employees work under your direct supervision?
6. How would you describe your own position (in terms of goals, responsibilities, means)?
7. By whom is your performance evaluated?
8. What are your agreements for the current year?⁶⁸
9. Have specific agreements been made on how you should achieve these agreements (as described under 8)?⁶⁹
10. On what information are the agreements made with you founded?
11. What is your own contribution in making the agreements?
12. To what extent do agreements reflect all relevant aspects of your task as a branch manager?
13. In your opinion, what should be the most important agreements in your performance evaluation?
14. To what agreements do you think your area manager attaches the greatest value when evaluating your performance?
15. And the former area manager?⁷⁰
16. What was the final overall evaluation mark you received last year?⁷¹
17. Are you satisfied with the agreements in your contract and with the manner in which your performance is evaluated?
18. What sorts of information do you receive during the year that enables you to compare your actual performance with agreements? How often do you receive this information? Who provides it to you?

⁶⁸ In this organisation, at the beginning of the year the superior sits down with each of his subordinates to enter a contract that contains specific goals that have to be achieved in the coming year. Both the superior and the subordinate sign the contract and it then serves as the basis for the formal performance evaluation at the end of the year. The goals specified in the contract are called 'agreements'.

⁶⁹ This question refers to a section in the contract called: manner of functioning. In this section, according to a brochure about the performance evaluation system, agreements can be inserted about specific shortcomings in the manager's functioning that he should work on, training-courses that the subordinate or his subordinates should follow, etc.

⁷⁰ This question was inserted in the interviews with the branch managers in area X since the Area Manager had only been employed in his current position since January 1, 1997.

⁷¹ The same brochure as mentioned in note 2 above talked about four possible overall evaluation marks for a

19. Is there any information you would like to receive that you currently do not receive?
20. Is there any information you currently receive that you think of as being unimportant?
21. How frequent do you have formal conversations with the area manager? And informal conversations? Are these conversations usually held at fixed moments, or if something gives rise to it? What are possible occasions for having a conversation?
22. What happens when performance is not as it should be?
23. Do you feel free to discuss problems in your work with the area manager?
24. In providing you with feedback, does the area manager mainly refer to the numbers or does he invite you for a personal conversation to discuss what caused these results?
25. When your performance falls short, does the area manager contribute to a solution, give you advice on how to improve, or does he signal the problem and leave it up to you what to do about it?
26. Do you think the area manager has enough information to enable him to evaluate your performance?
27. In general terms, when is a branch manager performing well in your opinion? And when is he performing poor?
28. Do you usually agree with your supervisor on the final performance evaluation?
29. In what way does the performance evaluation cycle contribute to your own functioning (in terms of performance, motivation, stress, communication with colleagues, team spirit, etc.)?
30. Do you see any positive or negative effects that the performance evaluation cycle has on yourself? And on the organisation as a whole?

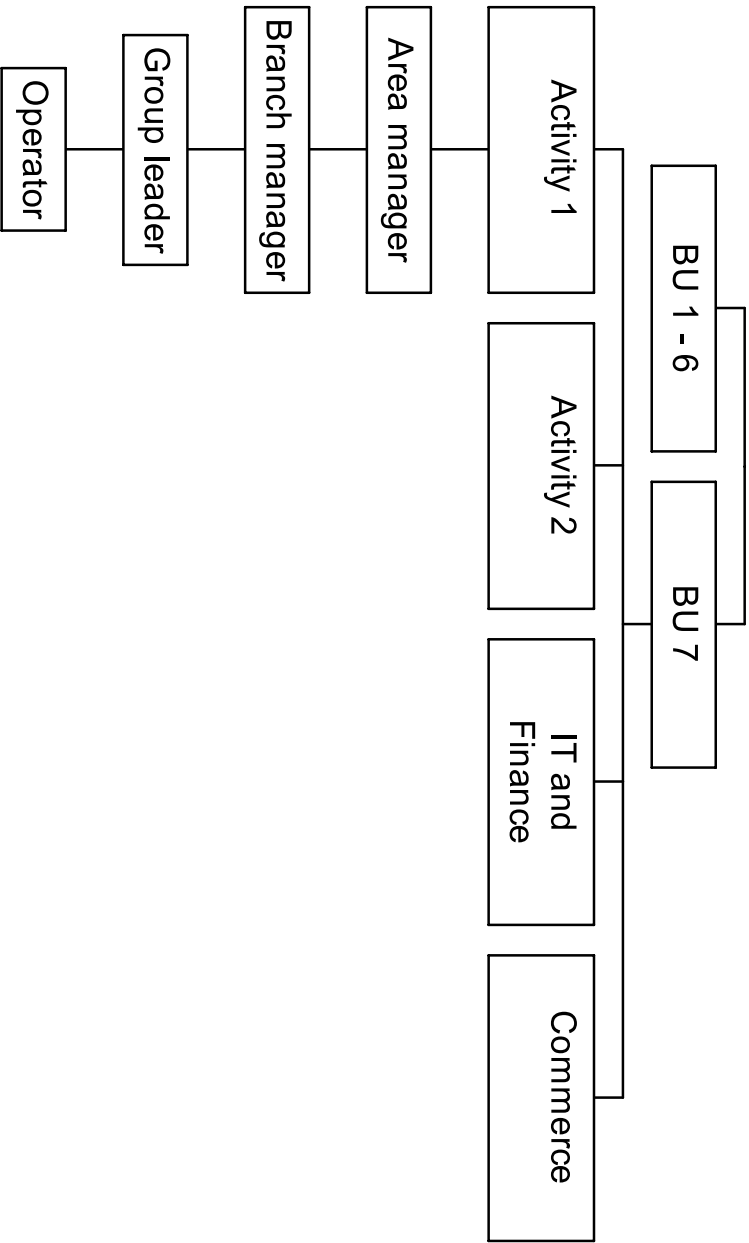
2. Interview protocol for interviewing the area manager (translated)

1. How long have you been employed by <name of the organisation>?
2. How long have you been employed in your current occupation?
3. What other positions did you hold within this organisation?
4. What is your educational background?
5. How many employees work under your direct supervision and are evaluated by you?
6. How would you describe your own position (in terms of goals, responsibilities, means)?
7. By whom is your performance evaluated?
8. What are your agreements for the current year?
9. Have specific agreements been made on how you should achieve these agreements (as described under 8)?
10. Which agreements are especially important when your performance is being evaluated?

superior to choose from. These marks were: poor, passable, good, and excellent.

11. Are you satisfied with the agreements in your contract and with the manner in which your performance is evaluated?
12. What was the final overall evaluation mark you received last year?
13. How would you describe the branch manager's position (in terms of goals, responsibilities, means)?
14. What do you consider the most important result areas for the branch managers?
15. On what information are the agreements that you make with the branch managers founded?
16. What are the agreements made with the branch managers for the current year?
17. Do you also include agreements about how branch managers should achieve these agreements?
18. What is the contribution of branch managers in making the agreements?
19. To what extent do agreements reflect all relevant aspects of a branch manager's task?
20. Are there any other performance measures used in the organisation besides the agreements?
21. Do the items about which agreements are made change in the course of years?
22. What sorts of information do you receive during the year that enables you to compare actual performance with agreements? How often do you receive this information? Who provides it to you?
23. Is there any information you would like to receive that you currently do not receive?
24. Is there any information you currently receive that you think of as being unimportant?
25. How frequent do you have formal conversations with the branch managers? And informal conversations? Are these conversations usually held at fixed moments, or if something gives rise to it? What are possible occasions for having a conversation?
26. What actions do you undertake when branch manager's performance is not as it should be?
27. In interpreting the information you get and when you give feedback to branch managers, do you mainly refer to the numbers or do you invite him for a personal conversation to discuss what caused these results?
28. Do you think you have enough information to evaluate branch managers' performance?
29. In general terms, when is a branch manager performing well in your opinion? And when is he performing poor?
30. Do branch managers usually agree with you on the final performance evaluation?
31. If branch managers have problems in achieving the agreements, do they come to you or will they wait until you signalise it?
32. In what way does the performance evaluation cycle contribute to your own functioning (in terms of performance, motivation, stress, communication with colleagues, team spirit, etc.)? And to the branch managers' functioning?
33. Do you see any positive or negative effects that the performance evaluation cycle has on yourself? And on the branch managers?

Appendix B. Organisational structure of business unit of Service Group



Appendix C. Interview protocols used at VDBN

1. Interview protocol subordinate

1. Could you please describe your educational and professional background? (*Hoe zag uw loopbaan er tot nu toe uit?*)
2. How many employees do you evaluate? (*Hoeveel medewerkers worden door u beoordeeld?*)
3. By whom is your performance evaluated? (*Door wie wordt u beoordeeld, aan wie legt u verantwoording af?*)
4. How would you describe your own position/ function (in terms of goals, expected outcomes, means, responsibility)? (*Hoe zou u uw eigen functie omschrijven (in termen van doel, verwachte resultaten, middelen, verantwoordelijkheid)?*)
5. How would you describe the position/ function of your superior (in the same terms)? (*Hoe zou u de functie(s) van uw leidinggevende (in deze zelfde termen) omschrijven?*)
6. Is there agreement on the requirements of these different positions/ functions? (*Bestaat er overeenstemming over de inhoud van deze verschillende functies?*)
 - Are there any formal descriptions of positions/ functions? (*Zijn er van alle functies functieomschrijvingen?*)
7. What are the most important tools available within the company to assess whether the expectations and goals that belong to a certain function/ position have been met? (For example performance measures, periodic reports, meetings, etc.) (*Wat zijn de belangrijkste middelen die er binnen de onderneming zijn om vast te stellen of aan de verwachtingen en doelen die bij een bepaalde functie horen wordt voldaan? (Ik denk hierbij bijvoorbeeld aan prestatie maatstaven, periodieke rapportages, werkoverleg, etc.)*)
8. What is the purpose and meaning of the PDP-system? (*Wat is het doel en de betekenis van het PDP-systeem?*)
9. Apparently there is a lot of information available on your performance and the performance of your unit. Which information and which performance measures in particular does your superior use when evaluating your performance with regards to PDP? (*Er is dus veel informatie beschikbaar over uw prestaties en de prestaties van uw eenheid als geheel. Welke informatie en welke prestatie maatstaven gebruikt uw leidinggevende bij de beoordeling van uw prestaties in het kader van PDP?*)
10. Are there any performance measures or targets that your superior thinks are particularly important? (Why do you think that?) (*Zijn er bepaalde maatstaven of doelstellingen die uw leidinggevende belangrijker vindt dan andere? (Waarom denkt u dat?)*)
11. What is your own influence and that of your colleagues in determining these

measures and targets? *(Wat is uw eigen inbreng en die van uw collega's bij het vaststellen van deze maatstaven en de doelstellingen?)*

- In your opinion, to what extent are targets that your superior has to meet important in determining these targets? *(In hoeverre spelen de doelstellingen die uw leidinggevende zelf moet halen volgens u een rol bij het vaststellen van deze doelstellingen?)*
12. How does your superior use these performance measures for evaluating your performance? For example, suppose you do not meet the target of <mentioned in question 9> on <the measures as described in question 9>, how will your superior deal with that when evaluating your performance? *(Hoe gebruikt uw leidinggevende deze prestatie maatstaven bij uw beoordeling? Stel bijvoorbeeld dat u het gestelde doel van <bij vraag 9 genoemd> op <een prestatie maatstaf genoemd bij vraag 9> niet gehaald heeft, hoe gaat uw leidinggevende daarmee om in de prestatiebeoordeling?)*
 13. And if you exceed the target <as mentioned in question 11>, how will your superior deal with that? *(En als het gestelde doel van <vraag 11> ruim gehaald wordt, hoe gaat uw leidinggevende er dan mee om?)*
 14. Next, I will ask whether the manner in which the superior evaluates his subordinates, with more or less emphasis on particular performance measures, is related to certain characteristics of these measures ((in)completeness, (in)validity, (un)controllability) <Explain further during interview, depending on answer on questions 11 and 12> *(Vervolgens vraag ik of de manier waarop de leidinggevende zijn medewerkers beoordeeld, met meer of minder aandacht voor bepaalde prestatie maatstaven, te maken heeft met bepaalde karakteristieken van de maatstaven ((on)volledig, (on)betrouwbaar, (on)beheersbaar). <Nader toelichten en formuleren in het gesprek, afhankelijk van het antwoord op vraag 11 en 12>)*
 15. You have indicated that accounting performance measures are relatively unimportant for the evaluation of your performance. Can you explain why? (Is this related to the characteristics of the measures too?) *(U heeft aangegeven dat accounting prestatie maatstaven relatief onbelangrijk zijn voor de beoordeling van uw prestaties. Kunt u daar een reden voor geven? (Heeft dat weer te maken met bepaalde karakteristieken van die maatstaven?))*
 16. There is a lot of accounting information available. For what purposes does your superior use this information? What is the use, the purpose of this information? *(Er is wel veel informatie beschikbaar. Waarvoor gebruikt u en uw leidinggevende die informatie dan wel? Wat is het nut, de functie, het doel van die informatie?)*
 17. Do you agree with the evaluation criteria that are applied? *(Bent u het eens met de gehanteerde beoordelingscriteria?)*
 18. Do you agree with the manner in which your superior evaluates your performance, or do you think (s)he should do it differently? *(Bent u het eens met de manier waarop uw baas uw prestaties beoordeeld, of vindt u dat hij/zij het anders zou moeten)*

doen?)

19. Do you agree or did you reach consensus on your final evaluation? *(Bent u het eens of het eens geworden met uw uiteindelijke beoordeling?)*
20. Do you think the PDP-system is a good performance evaluation system? (Why (not)?) *(Vindt u het PDP-systeem een goed beoordelingssysteem? (Waarom (niet))?)*
21. What has changed in your performance evaluation now that PDP has been implemented (as compared to the old (previous) system)? *(Wat is er veranderd in uw beoordeling nu het PDP is ingevoerd (ten opzichte van het oude systeem)?)*
22. Are there any changes in performance measures that are important in your performance evaluation over time? *(Veranderen prestatiemaatstaven die een rol spelen in de beoordeling van tijd tot tijd?)*
23. Does your superior use different measures for different subordinates (if so, which measures and why), or does (s)he evaluate all subordinates on the same aspects (why)? *(Hanteert uw leidinggevende verschillende maatstaven voor verschillende medewerkers (zo ja, welke maatstaven en waarom), of beoordeelt hij/zij alle medewerkers op dezelfde aspecten (waarom)?)*
24. In your opinion, what is the purpose of PDP? (of performance evaluation) *Wat is volgens u het doel van PDP? (van prestatiebeoordeling)*
25. Are these purposes achieved? If so, how? Can you provide some examples? *(Worden die doelen bereikt? Zo ja, hoe? Kunt u voorbeelden geven?)*
26. Do you see any examples of learning and innovation within your own unit? Can you provide some examples? *(Ziet u voorbeelden van leren en innovatie in uw eigen unit? Kunt u voorbeelden geven?)*
27. Both academic and popular/ professional management literature provide examples of dysfunctional behaviours as a consequence of using certain performance measures. A well-known example is postponing necessary maintenance on machinery till the next period which could result in [storingen en vastlopen] of the machinery. Do you know examples of such type of behaviours within VDBN? *(In de wetenschappelijke en populaire management literatuur staan verschillende voorbeelden van disfunctionele gedragingen die het gevolg zijn van het hanteren van bepaalde prestatie maatstaven. Bijvoorbeeld het uitstellen van noodzakelijk onderhoud aan machines tot een volgende periode met het risico op storingen en vastlopen van de machines. Kent u zelf voorbeelden van zulk soort gedragingen binnen VDBN?)*
28. Do you think there are any other positive and/ or negative aspects of the PDP-system that have not been mentioned before? *(Ziet u verder nog positieve en/of negatieve kanten aan het PDP-systeem?)*

2. Interview protocol leader

1. Could you please describe your educational and professional background? (*Hoe zag uw loopbaan er tot nu toe uit?*)
2. How many employees do you evaluate? (*Hoeveel medewerkers worden door u beoordeeld?*)
3. By whom is your performance evaluated? (*Door wie wordt u beoordeeld, aan wie legt u verantwoording af?*)
4. How would you describe your own position/ function (in terms of goals, expected outcomes, means, responsibility)? (*Hoe zou u uw eigen functie omschrijven (in termen van doel, verwachte resultaten, middelen, verantwoordelijkheid)?*)
5. How would you describe the position/ function of your subordinates (in the same terms)? (*Hoe zou u de functie(s) van uw leidinggevende (in deze zelfde termen) omschrijven?*)
6. Is there agreement on the requirements of these different positions/ functions? (*Bestaat er overeenstemming over de inhoud van deze verschillende functies?*)
 - Are there any formal descriptions of positions/ functions? (*Zijn er van alle functies functieomschrijvingen?*)
7. What are the most important tools that you have to assess whether the expectations and goals that belong to a certain function/ position have been met? (For example performance measures, periodic reports, meetings, etc.) (*Wat zijn de belangrijkste middelen die u heeft om vast te stellen of aan de verwachtingen en doelen die bij een bepaalde functie horen wordt voldaan? (Ik denk hierbij bijvoorbeeld aan prestatie maatstaven, periodieke rapportages, werkoverleg, etc.)*)
8. What is the purpose and meaning of the PDP-system? (*Wat is het doel en de betekenis van het PDP-systeem?*)
9. Apparently there is a lot of information available on the performance of your subordinates and the performance of your unit. Which information and which performance measures do you use for evaluating the performance of your subordinates with regards to PDP? (*Er is dus veel informatie beschikbaar over de prestaties van uw medewerkers en de prestaties van eenheid als geheel. Welke informatie en welke prestatie maatstaven gebruikt u voor de beoordeling van uw medewerkers in het kader van PDP?*)
10. Do you have any discretion to choose these measures or are they fixed? (*Bent u vrij om deze maatstaven zelf te kiezen of liggen deze vast?*)
 - What is the influence of your subordinates in determining these measures and targets? (*Wat is de inbreng van medewerkers bij het vaststellen van de maatstaven en de doelstellingen?*)
11. How do you use these performance measures for evaluating the performance of your subordinates? For example, suppose a subordinate does not meet the target of

- <mentioned in question 9> on <the measures as described in question 9>, how will you deal with that when evaluating his performance? (*Hóe gebruikt u deze prestatiemaatstaven bij de beoordeling van uw medewerkers? Stel bijvoorbeeld dat een medewerker het gestelde doel van <bij vraag 9 genoemd> op <een prestatiemaatstaf genoemd bij vraag 9> niet gehaald heeft, hoe gaat u daarmee om in de prestatiebeoordeling?*)
12. And if the target <as mentioned in question 11> is well exceeded, how will you deal with that? (*En als het gestelde doel van <vraag 11> ruim gehaald wordt, hoe gaat u er dan mee om?*)
 13. Next, I will ask whether the manner in which the superior evaluates his subordinates, with more or less emphasis on particular performance measures, is related to certain characteristics of these measures ((in)completeness, (in)validity, (un)controllability) <Explain further during interview, depending on answer on questions 11 and 12> (*Vervolgens vraag ik of de manier waarop de leidinggevende zijn medewerkers beoordeeld, met meer of minder aandacht voor bepaalde prestatiemaatstaven, te maken heeft met bepaalde karakteristieken van de maatstaven ((on)volledig, (on)betrouwbaar, (on)beheersbaar). <Nader toelichten en formuleren in het gesprek, afhankelijk van het antwoord op vraag 11 en 12>*)
 14. You have indicated that accounting performance measures are relatively unimportant for performance evaluation. Why do you consider them unimportant? (Is this related to the characteristics of the measures too?) (*U heeft aangegeven dat accounting prestatiemaatstaven relatief onbelangrijk zijn voor de prestatiebeoordeling. Waarom vindt u ze onbelangrijk? (Heeft dat weer te maken met bepaalde karakteristieken van die maatstaven?)*)
 15. For what purposes do you use this information? What is the use, the purpose of this information? (*Waarvoor gebruikt u die informatie dan wel? Wat is het nut, de functie, het doel van die informatie?*)
 16. Until now we have discussed the evaluation of your subordinates. But of course you yourself are evaluated also by your superior. Which performance measures are important in the evaluation of your own performance, and how does your superior evaluate you on these measures? (*We hebben het tot nu toe gehad over de beoordeling van uw medewerkers. Maar u wordt zelf natuurlijk ook beoordeeld door uw leidinggevende. Welke prestatiemaatstaven zijn bij uw eigen beoordeling belangrijk, en hoe beoordeelt uw baas u hierop?*)
 17. Do you agree with the manner in which your performance is evaluated? (*Bent u het eens met de manier waarop u beoordeeld wordt?*)
 18. Do you and your subordinates usually agree or do you usually reach consensus on the final evaluation? (*Zijn u en uw medewerkers het meestal eens of het eens geworden over de uiteindelijke beoordeling?*)
 19. Do you think the PDP-system is a good performance evaluation system? (Why (not)?)

- (Vindt u het PDP-systeem een goed beoordelingssysteem? (Waarom (niet))?)*
20. What has changed in the performance evaluation of your subordinates now that PDP has been implemented as compared to the old (previous) system? *(Wat is er veranderd in de beoordeling van uw medewerkers nu het PDP is ingevoerd ten opzichte van het oude systeem?)*
21. Are there any changes in performance measures that are important in your performance evaluation over time? *(Veranderen prestatiemaatstaven die een rol spelen in de beoordeling van tijd tot tijd?)*
22. Do you use different measures for different subordinates (if so, which measures and why), or do you evaluate all subordinates on the same aspects (why?)? *(Hanteert u verschillende maatstaven voor verschillende medewerkers (zo ja, welke maatstaven en waarom), of beoordeelt u alle medewerkers op dezelfde aspecten (waarom)?)*
23. In your opinion, what is the purpose of PDP? (of performance evaluation) *(Wat is volgens u het doel van PDP? (van prestatiebeoordeling))*
24. Are these purposes achieved? If so, how? Can you provide some examples? *(Worden die doelen bereikt? Zo ja, hoe? Kunt u voorbeelden geven?)*
25. Do you see any examples of learning and innovation within your own unit? Can you provide some examples? *(Ziet u voorbeelden van leren en innovatie in uw eigen unit? Kunt u voorbeelden geven?)*
26. Both academic and popular/ professional management literature provide examples of dysfunctional behaviours as a consequence of using certain performance measures. A well-known example is postponing necessary maintenance on machinery till the next period which could result in [storingen en vastlopen] of the machinery. Do you know examples of such type of behaviours within VDBN? *(In de wetenschappelijke en populaire management literatuur staan verschillende voorbeelden van disfunctionele gedragingen die het gevolg zijn van het hanteren van bepaalde prestatie maatstaven. Bijvoorbeeld het uitstellen van noodzakelijk onderhoud aan machines tot een volgende periode met het risico op storingen en vastlopen van de machines. Kent u zelf voorbeelden van zulk soort gedragingen binnen VDBN?)*
27. Do you think there are any other positive and/ or negative aspects of the PDP-system that have not been mentioned before? *(Ziet u verder nog positieve en/of negatieve kanten aan het PDP-systeem?)*

Appendix D. Overview of variable measurement in the questionnaire

Table 1. Perceived evaluative style / Preferred evaluative style

1.	Korte termijn doelen - <i>Short-term goals</i>
2.	Lange termijn doelen - <i>Long-term goals</i>
3.	Financiële informatie - <i>Financial information</i>
4.	Niet-financiële informatie - <i>Non-financial information</i>
5.	Afwijkingen van afgesproken prestaties - <i>Deviations from agreed performance</i>
6.	Uw uitleg van uw prestaties - <i>Your explanation of your performance</i>
7.	Objectief meetbare prestaties - <i>Objectively measurable performance</i>
8.	Zijn/haar eigen intuïtie - <i>His/her own intuition</i>
9.	In het afgelopen jaar geleverde prestaties - <i>Performance delivered in the past year</i>
10.	In de afgelopen jaren geleverde prestaties - <i>Performance delivered over the past few years</i>
11.	Uw plannen voor het komende jaar - <i>Your plans for the next year</i>
12.	Positieve aspecten van uw presteren - <i>Positive aspects of your performance</i>
13.	Negatieve aspecten van uw presteren - <i>Negative aspects of your performance</i>
14.	Prestaties van de eenheid waarvoor u verantwoordelijk bent - <i>Performances of the unit that you are responsible for</i>
15.	Persoonlijke, individuele prestaties - <i>Personal, individual performance</i>
16.	Oorzaken voor het achterblijven van prestaties - <i>Causes of failings in performance</i>
17.	Mogelijkheden tot verbetering/ontwikkeling van uw prestaties - <i>Possibilities to improve/develop your performance</i>
18.	Informatie die hij/zij ontvangt uit het gesprek met u - <i>Information he/she gets from the conversation with you</i>
Source: developed for this study (in Dutch)	
17 items, item 10 dropped	
For further statistics and underlying dimensions: see Chapter 7, Table 7.4 and Table 7.5.	

Table 2. Perceived evaluative style: Interpersonal evaluation

1.	Mijn baas is grondig bekend met mijn werk-prestatie (job performance) - <i>My supervisor is thoroughly familiar with my job performance</i>
2.	Mijn baas geeft me ruimte mijn kant van het verhaal te vertellen bij de prestatiebeoordeling - <i>My supervisor allows me to tell my side of the story in performance evaluation</i>
3.	Mijn baas laat me mijn beoordelingsuitkomsten weten en geeft er een rechtvaardiging voor - <i>My supervisor lets me know my appraisal outcomes and provides justification</i>
4.	Mijn baas laat me mijn salarisverhoging en jaarlijkse bonus weten en geeft er een rechtvaardiging voor - <i>My supervisor lets me know my pay raise and annual bonuses and provides justification</i>
5.	Mijn baas bespreekt mijn geleverde prestaties met me alsmede plannen of doelstellingen om mijn prestaties te verbeteren - <i>My supervisor reviews my performance with me and discusses plans or objectives to improve my performance</i>
Source: Farh, Early & Lin, 1996	
5 items	
Cronbach's α = .73	

Table 3. Leadership style

3A. Active management-by-exception	
1.	Verwacht veel van ons - <i>Expects a lot from us</i>
2.	Wijst mij erop wanneer mijn werk onder de maat is - <i>Points it out to me when my work is not up to par</i>
3.	Let op fouten in de prestaties die correctie behoeven - <i>Monitors performance for errors needing correction</i>
4.	Toont zijn/haar ongenoegen wanneer mijn werk onder aanvaardbare normen is - <i>Shows his or her displeasure when my work is below acceptable standards</i>
5.	Vestigt de aandacht op onregelmatigheden, uitzonderingen en afwijkingen van wat van mij verwacht wordt - <i>Focuses attention on irregularities, exceptions and deviations from what is expected of me</i>
6.	Zou afkeuring laten blijken als ik op een laag niveau zou presteren - <i>Would indicate disapproval if I performed at a low level</i>
7.	Vestigt de aandacht op vergissingen en afwijkingen die ik bega - <i>Focuses attention on errors I make</i>
8.	Houdt fouten goed in de gaten - <i>Keeps careful track of mistakes</i>
Source: Den Hartog (1998)	
7 items, item 1 dropped	
Cronbach's $\alpha = .87$	
average inter-item correlation .48	

3B. Supportive	
1.	Behandelt mij als een individu, in plaats van als zomaar een lid van de groep - <i>Treats me as an individual rather than just a member of the group</i>
2.	Luistert naar zaken die voor mij van belang zijn - <i>Listens to my concerns</i>
3.	Geeft mij veel zeggenschap in het formuleren van mijn eigen (prestatie) doelen - <i>Allows me a strong hand in setting my own performance goals</i>
4.	Toont vertrouwen in mijn vermogen bij te dragen aan de doelen van deze eenheid - <i>Shows confidence in my ability to contribute to the goals of this unit</i>
5.	Toont een volledig vertrouwen in mij - <i>Demonstrates total confidence in me</i>
6.	Houdt rekening met mijn persoonlijke welzijn - <i>Looks out for my personal welfare</i>
7.	Is oprecht geïnteresseerd in de ontwikkeling van zijn/haar medewerkers - <i>Is genuinely concerned about the growth and the development of subordinates</i>
8.	Geeft advies wanneer dat nodig is -
Source: Den Hartog (1998) (Adapted; see § 5.3.2)	
7 items, item 8 dropped	
Cronbach's $\alpha = .76$	
average inter-item correlation .34	

3C. Team-building

1.	Spoort medewerkers aan om 'teamspelers' te zijn - <i>Encourages employees to be 'team players'</i>
2.	Ontwikkelt teamgeest bij medewerkers - <i>Develops teamspirit among employees</i>
3.	Krijgt de groep tot samenwerking voor hetzelfde doel - <i>Gets the group to work together for the same goal</i>
4.	Tracht een klimaat van vertrouwen te creëren onder de leden van het management-team - <i>Works at creating a climate of trust among members of the management team</i>
5.	Doorbreekt communicatie-barrières tussen werkgroepen - <i>Breaks down barriers to communication between work groups</i>
Source: Den Hartog (1998)	
5 items	
Cronbach's α = .93	
average inter-item correlation .72	

Table 4. Perceived fairness of evaluation

1.	De beoordeling die ik ontvang is gebaseerd op factoren die ik volledig onder controle heb - <i>The evaluation I receive is based on factors over which I have full control</i>
2.	Het gebeurt vaak dat mijn baas mij bepaalde (tegenvallende) resultaten aanrekent waar ik echt niets aan kan doen - <i>It often happens that my superior holds me accountable for certain (negative) results that I really cannot help (R)</i>
3.	De beoordeling die ik ontvang is gebaseerd op factoren die ook ik relevant vind voor mijn functioneren - <i>The evaluation I receive is based on factors that also I find relevant for my functioning</i>
4.	Bij de beoordeling van mijn functioneren legt mijn baas vaak de nadruk op aspecten van mijn werk die ik niet relevant vind - <i>When evaluating my functioning my superior often emphasises aspects of my work which I think are irrelevant (R)</i>
5.	De beoordeling die ik ontvang is gebaseerd op een compleet beeld van mijn werkelijke prestaties - <i>The evaluation I receive is based on a complete picture of my true performance</i>
6.	Bepaalde prestaties en acties die ik belangrijk vind in mijn functioneren worden door mijn baas vergeten als hij/zij mij beoordeelt - <i>Certain achievements and actions which I think are important in my functioning are overlooked by my superior when he evaluates me (R)</i>
7.	In het algemeen vind ik dat mijn functioneren en mijn prestaties op een eerlijke en 'faire' manier worden beoordeeld - <i>In general I think that my functioning and performance is evaluated in an honest and fair way</i>
8.	In het algemeen vind ik de criteria waarop mijn baas mij beoordeelt eerlijk en 'fair' - <i>Inn general I think that the criteria my superior uses to evaluate me are honest and fair</i>
9.	Ik ben zeer tevreden over de manier waarop ik word beoordeeld - <i>I am very satisfied with the way in which I am evaluated</i>
Source: Hartmann (1997)	
9 items	
Cronbach's α = .88	

Table 5. Felt appropriateness of accounting performance measures

1.	De afdeling waaraan ik leiding geef heeft bovenal een financiële doelstelling (zoals 'het maximaliseren van de winst') - <i>The department I supervise has above all a financial goal (like 'maximising profit')</i>
2.	Of mijn afdeling goed functioneert of niet is nauwkeurig in financiële cijfers uit te drukken - <i>Whether my department functions well or not can be expressed accurately in financial figures</i>
3.	Ik vind dat mijn afdeling pas goed heeft gefunctioneerd als er positieve budgetresultaten ontstaan - <i>I feel that my department has performed well only if it shows positive budget variances</i>

4.	Budgetresultaten geven een volledig overzicht van het functioneren en van de resultaten van mijn afdeling - <i>Budget variances provide a complete overview of the functioning and performance of my department</i>
5.	Veel van de activiteiten en taken die mijn afdeling verricht komen niet in de budgetresultaten tot uitdrukking - <i>Many of the activities and tasks my department performs do not show up in the budget variances</i> (R)
6.	Voor het bepalen of mijn afdeling goed heeft gepresteerd of niet, zijn budgetresultaten niet relevant - <i>Budget variances are irrelevant for determining whether or not my department has performed well</i> (R)
Source: Hartmann (1997)	
4 items, items 2 and 5 dropped	
Cronbach's α = .76	

Table 6. Trust in superior

1.	Benut uw baas mogelijkheden die zich voordoen om uw belangen te bevorderen door middel van zijn/haar daden en beslissingen? - <i>Does your superior take advantage of opportunities that come up to further your interests by his actions and decisions?</i> (R)
2.	Hoe vrij voelt u zich om met uw baas de problemen en moeilijkheden die u in uw werk heeft te bespreken, zonder dat het uw positie in gevaar brengt of dat het later "tegen u gebruikt" wordt? - <i>How free do you feel to discuss with your superior the problems and difficulties you have in your job without jeopardizing your position or having it 'held against' you later on?</i> (R)
3.	Hoeveel vertrouwen heeft u erin dat uw baas u volledig en openhartig op de hoogte houdt van zaken die u aan kunnen gaan? - <i>How confident do you feel that your superior keeps you fully and frankly informed about things that might concern you?</i> (R)
4.	Leidinggevendens moeten wel eens beslissingen nemen die lijken in te gaan tegen de belangen van hun medewerkers. Als dit u als medewerker overkomt, hoeveel vertrouwen heeft u er dan in dat de beslissing van uw baas gerechtvaardigd wordt door andere overwegingen? - <i>Superiors at times must make decisions which seem to be against the interests of their subordinates. When this happens to you as a subordinate, how much trust do you have that your superior's decision is justified by other considerations?</i> (R)
5.	Stel dat u een probleem in uw werk zou hebben. In welke mate denkt u dat uw baas bereid zou zijn om moeite te doen om u te helpen als u daarom zou vragen? - <i>If you were having some difficulty in your job, to what extent do you feel your departmental supervisor would be willing to go out of his way to help you if you asked for it?</i> (R)
Source: Read (1962) and Hopwood (1973)	
5 items	
Cronbach's α = .78	

Table 7. Job satisfaction ('satisfaction with organisation': see § 5.4.7)

1.	Ik zou liever ander werk hebben dan ik op dit moment doe - <i>I would rather have some other work (R)</i>
2.	Ik heb al veel vooruitgang gemaakt in deze organisatie - <i>I have made a great deal of progress in this organisation</i>
3.	Los van mijn precieze baan en mijn precieze functie houd ik bijzonder veel van het <u>soort werk</u> dat ik doe - <i>Not counting all the other aspects of my job, I like the kind of work that I do very much</i>
4.	Mijn functie geeft mij uitstekend de gelegenheid die dingen te doen waarin ik het beste ben - <i>My job gives me the chance to do the things I am best at</i>
5.	Het is zeer prettig voor deze organisatie te werken - <i>I like working for this company very much</i>
6.	Ik zou een vriend of vriendin zeker adviseren in deze organisatie te komen werken - <i>I would certainly advise a friend to come and work for this organisation</i>
7.	Als ik mijn huidige werk, tegen hetzelfde salaris, binnen een andere organisatie zou kunnen doen, dan zou ik zeker hier blijven - <i>If I had a chance to do the same kind of work for the same pay, but in another company, I would certainly stay here</i>
Source: Hartmann (1997)	
3 items, items 1, 2, 3 and 4 dropped	
Cronbach's $\alpha = .68$	

Table 8. Job-related tension

1.	Ik ervaar (te) veel stress in mijn werk - <i>I experience (too) much tension in my work</i>
2.	Tijdens beoordelingsgesprekken voel ik me gespannen - <i>I experience job-tension during performance evaluations</i>
3.	Als ik de doelen in mijn werk niet haal voel ik me gespannen - <i>If I don't attain my performance goals, I feel tense</i>
4.	Ik maak me vaak zorgen over de hoeveelheid werk die ik heb - <i>I often worry about the amount of work I have to do</i>
5.	Ik maak me vaak zorgen dat de hoeveelheid werk die ik verricht ten koste gaat van de kwaliteit ervan - <i>I often worry that the amount of work interferes with how well it gets done</i>
6.	Mijn baan heeft directe negatieve invloed op mijn gezondheid - <i>My job tends to directly affect my health</i>
7.	Problemen in het werk hebben mij 's nachts weleens wakker gehouden - <i>Problems associated with work have kept me awake at night</i>
8.	Ik neem vaak werk 'mee naar huis' in de zin dat ik eraan denk terwijl ik met iets anders bezig ben - <i>I often 'take my job home with me' in the sense that I think about it when doing other things</i>
Source: Hartmann (1997)	
5 items, items 2, 3 and 8 dropped	
Cronbach's $\alpha = .78$	

Table 9. Task uncertainty

1.	Ik kan in mijn werk gebruik maken van bekende stappen, procedures en ervaringen - <i>To do my work, I can rely on established procedures and practices</i>
2.	Ik kan bij het uitvoeren van mijn taken goed terugvallen op bekende kennis (handboeken, procedures, advies van anderen etc.) - <i>There is a clearly defined body of knowledge of subject matter which can guide me in doing my work</i>
3.	Mijn eigen taken kennen veel herhalingen - <i>My duties are repetitious</i>
4.	Mijn taken zijn van dag tot dag hetzelfde - <i>My tasks are the same from day-to-day</i>
5.	Voor het uitvoeren van veel van mijn taken is er een 'bekende weg' - <i>There is a clearly known way to do the major types of work I normally encounter</i>
6.	In het algemeen doen de medewerkers binnen mijn afdeling dagelijks ongeveer hetzelfde werk, op ongeveer dezelfde manier - <i>People in this unit do about the same job in the same way most of the time</i>
7.	In het algemeen zou ik mijn werk als 'routinematig' willen kenschetsen - <i>In general I would say that my work is fairly routine</i>
8.	De uitvoering van mijn taken geschiedt volgens een duidelijke en begrijpelijke volgorde van stappen - <i>There is an understandable sequence of steps that can be followed in doing my work</i>
9.	In essentie verrichten de medewerkers binnen mijn afdeling veel zich herhalende werkzaamheden in hun werk - <i>Basically, unit members perform repetitive activities in doing their jobs</i>
Source: Hartmann (1997)	
8 items, item 2 dropped	
Cronbach's $\alpha = .84$	

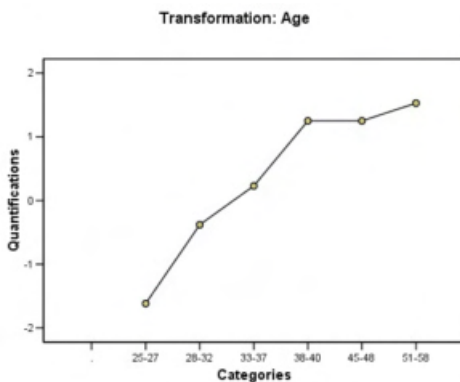
Appendix E. Output of CATPCA analysis reported in § 6.5.2

Figure E-1. Syntax of CATPCA analysis reported in chapter 6

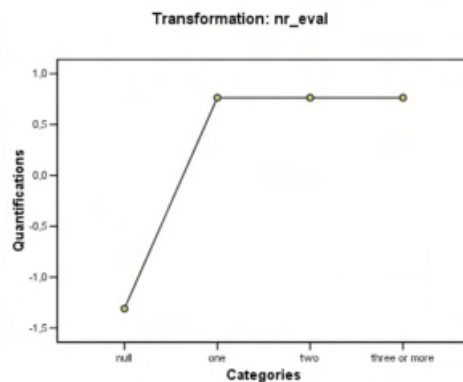
```
CATPCA
VARIABLES=age tenure yrprefu yrpresup tu gender educ leagroup group leader
/ANALYSIS=age(WEIGHT=1,LEVEL=ORDI) tenure(WEIGHT=1,LEVEL=ORDI)
yrprefu(WEIGHT=1,LEVEL=ORDI) yrpresup(WEIGHT=1,LEVEL=ORDI)
tu(WEIGHT=1,LEVEL=ORDI) gender(WEIGHT=1,LEVEL=MNOM) educ(WEIGHT=1,LEVEL=MNOM)
leagroup(LEVEL=MNOM) group(LEVEL=MNOM)
leader(LEVEL=MNOM)
/DISCRETIZATION=age(GROUPING,NCAT=7,DISTR=NORMAL) tu(GROUPING,NCAT=7,DISTR=NORMAL)
/MISSING=age(PASSIVE,MODEIMPU) tenure(PASSIVE,MODEIMPU) yrprefu(PASSIVE,MODEIMPU)
yrpresup(PASSIVE,MODEIMPU) tu(PASSIVE
,MODEIMPU) gender(PASSIVE,MODEIMPU) educ(PASSIVE,MODEIMPU) leagroup(PASSIVE,MODEIMPU)
group(PASSIVE,MODEIMPU) leader(PASSIVE
,MODEIMPU)
/SUPPLEMENTARY=VARIABLE( leagroup group leader )
/DIMENSION=2
/NORMALIZATION=VPRINCIPAL
/MAXITER=100
/CRITITER=.00001
/PRINT=LOADING OBJECT ( leagroup group leader ) QUANT( age tenure yrprefu yrpresup tu gender
educ leagroup group leader )
/PLOT=BIPLOT( LOADING ) ( group leader leagroup ) (20) OBJECT ( group leader leagroup ) (20)
JOINTCAT( educ gender ) (20)
LOADING ( ( CENTR ( gender educ ) ) ) (20) PROJCENTR( group age tenure yrprefu yrpresup tu gender
educ ) (20) TRANS( age
tenure yrprefu yrpresup tu gender(2) educ(2) leagroup(2) group(2) leader(2) ) .
```

Figure E-2. Transformation plots

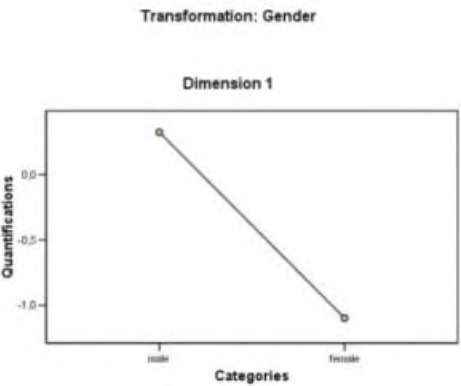
Panel A.



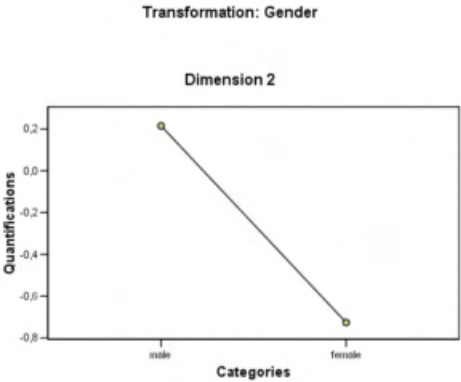
Panel B.



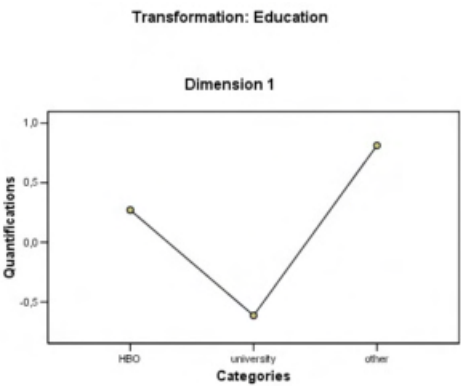
Panel C.



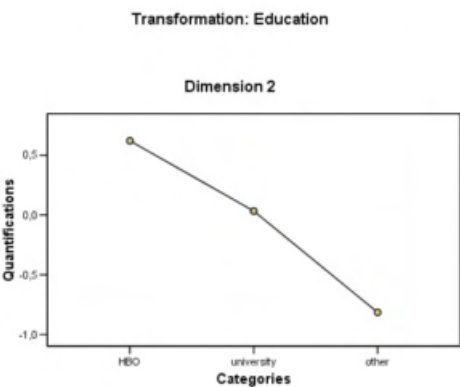
Panel D.



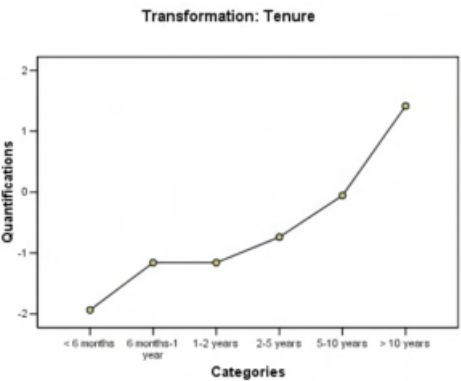
Panel E.



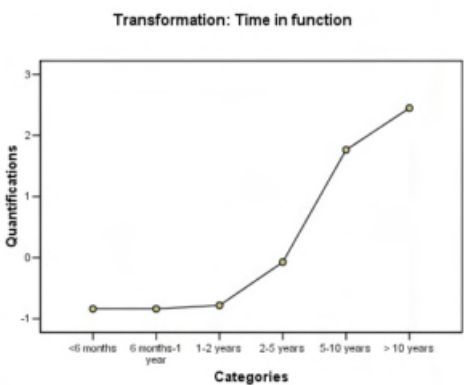
Panel F.



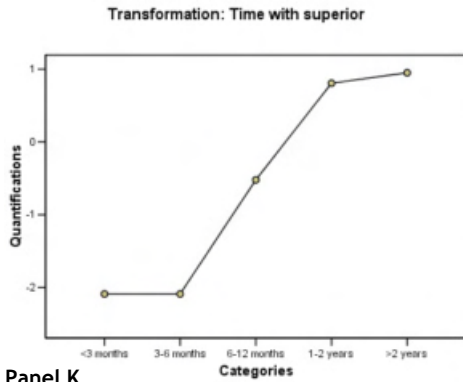
Panel G.



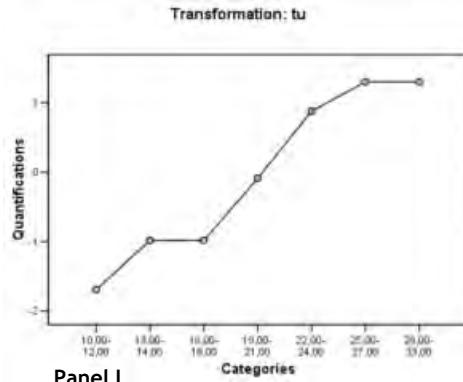
Panel H.



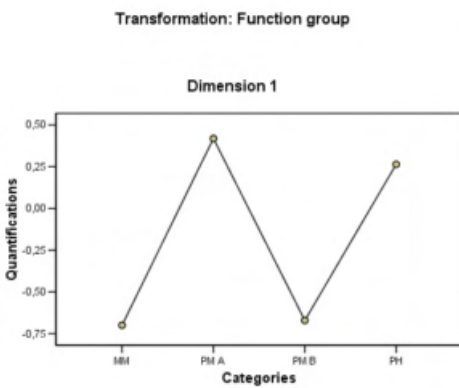
Panel I.



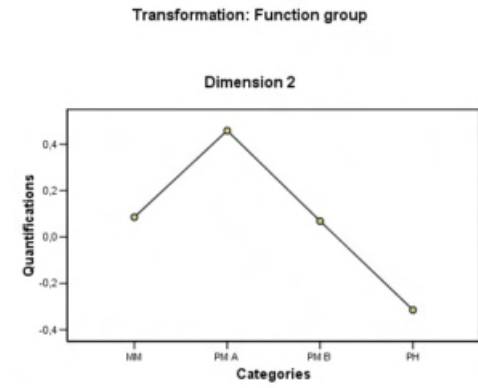
Panel J.



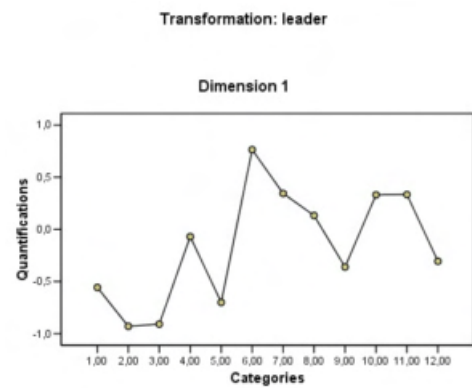
Panel K.



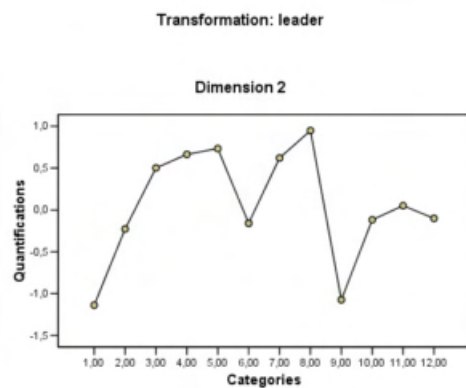
Panel L.



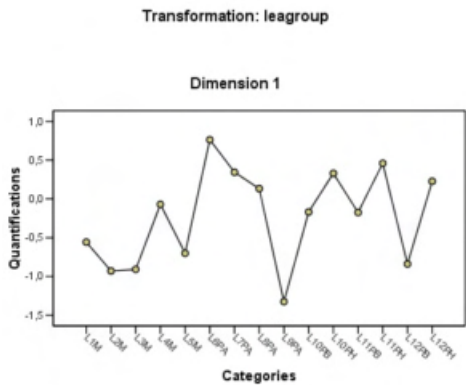
Panel M.



Panel N.



Panel O.



Panel P.

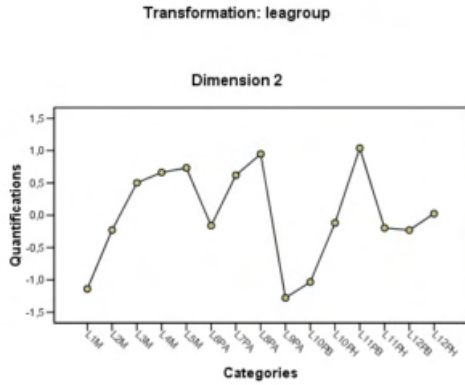
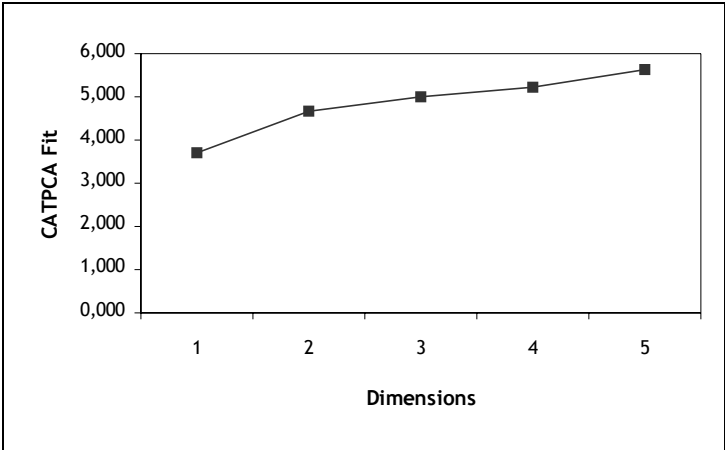


Figure E-3. Total CATPCA fit by number of dimensions



Appendix F. Calculation of within-group agreement coefficients (rwg)

The $r_{wg(1)}$ within-group agreement coefficient for a single item is calculated using the following formula:

$$r_{wg(1)} = 1 - (s_{xj}^2 / \sigma_{EU}^2)$$

where s_{xj}^2 is the observed variance on a single item X_j , and σ_{EU}^2 is the variance on item X_j that would be expected if all ratings were due exclusively to random measurement error (James et al. , 1984). "Random" means

"that each alternative on the measurement scale of X_j has an equal likelihood of response, and, therefore, that the judgements would be distributed uniformly (i.e., a rectangular distribution)." (James et al., 1984, p. 86).

The equation to calculate σ_{EU}^2 is as follows:

$$\sigma_{EU}^2 = (A^2 - 1) / 12 \text{ (Mood, Graybill \& Boes, 1974),}$$

where A corresponds to the number of alternatives in the response scale. In this research project, all items that we used to measure evaluative style were measured on a five-point Likert scale. Thus, A = 5.

For a multi-item scale, consisting of J items, the $R_{wg(J)}$ within-group agreement coefficient can be estimated by applying the Spearman-Brown formula to $r_{wg(1)}$:

$$r_{wg(J)} = \frac{J[1 - (\overline{s_{xj}^2} / \sigma_{EU}^2)]}{J[1 - (\overline{s_{xj}^2} / \sigma_{EU}^2)] + (\overline{s_{xj}^2} / \sigma_{EU}^2)}$$

Where $\overline{s_{xj}^2}$ is the mean of the observed variances on J items.

Appendix G. The calculation and coding of within-group differences

The calculation and coding of within-group differences that are used in § 8.2.2 will be illustrated using the scores of the five respondents of leader 6 on three variables: performance assessment (Assess), performance development (develop), and age.

The original scores were as follows:

Assess	Develop	Age
17	14	54
16	16	52
24	12	29
20	14	58
13	9	55

For each variable, all original scores were raised by 1, before subtracting the lowest score on that variable within the group. So, the lowest score within the group of respondents of leader 6 would be 13 for assess, 9 for develop, and 29 for age.

So, within-group differences were coded as follows:

Dif_Ass	Dif_Dev	Dif_Age
5	6	26
4	8	24
12	4	1
8	6	30
1	1	27

Samenvatting (in Dutch)

In dit proefschrift beschrijf ik een onderzoek naar de relatie tussen beoordelingsstijl van een leidinggevende en het gedrag van ondergeschikte managers. Onder beoordelingsstijl versta ik de manier waarop een leidinggevende manager de prestatie van zijn of haar ondergeschikte managers beoordeelt. Eerdere onderzoeken in de accounting literatuur naar beoordelingstijl en de gedragsmatige effecten daarvan baseren het onderscheid in verschillende beoordelingsstijlen op de mate waarin een leidinggevende manager voor de beoordeling van de prestaties van ondergeschikte managers de nadruk legt op accounting prestatiemaatstaven. De literatuur die uitgaat van deze opvatting van beoordelingsstijlen staat bekend als RAPM (Reliance on Accounting Performance Measures). De resultaten uit de RAPM literatuur rechtvaardigen de conclusie dat de beoordelingsstijl van leidinggevende managers invloed heeft op het gedrag van hun ondergeschikte managers (o.a. Hopwood 1973; Otley, 1978; Brownell & Hirst, 1986; Ross, 1995). Tegelijkertijd laten de resultaten van verschillende onderzoeken zien dat hoe beoordelingsstijl het gedrag beïnvloedt afhangt van situationele factoren, zoals organisatorische factoren (o.a. Otley, 1978; Hirst, 1981; 1983; 1987), omgevingsfactoren (o.a. Brownell, 1985; 1987), en persoonsgebonden factoren (o.a. Hopwood, 1973; Brownell, 1981). Ondanks meer dan 30 jaar onderzoek geven de resultaten uit de RAPM literatuur echter geen duidelijk en eenduidig beeld over de samenhang van beoordelingsstijl van de leidinggevende en het gedrag van ondergeschikte managers in verschillende omstandigheden. Recente overzichten van de RAPM-literatuur (Hartmann, 2000; Fakiolas & Otley, 2000; Otley & Pollanen, 2000; Vagneur & Peiperl, 2000) laten zien dat deze onderzoeksstroom veel problemen en tekortkomingen kent.

Ondanks de problemen en tekortkomingen van bestaande (RAPM-)onderzoeken naar beoordelingsstijlen en de gedragsmatige gevolgen daarvan, blijft onderzoek naar de effecten van beoordelingsstijlen voor bedrijven relevant. Veel grote bedrijven hebben aparte stafeenheden waar mensen betaald worden om allerlei systemen te ontwerpen, implementeren en te onderhouden die bedoeld zijn om medewerkers te beïnvloeden en te motiveren om in het belang van de organisatie te handelen. Grote investeringen worden gedaan in geavanceerde managementinformatiesystemen, prestatiemeting en prestatiebeoordelingssystemen. Maar beïnvloeden deze systemen werkelijk het gedrag van medewerkers op dezelfde wijze als de bedoeling is?

Tegen deze achtergrond onderzoek ik in dit proefschrift het verband tussen beoordelingsstijl van leidinggevende managers en het gedrag van ondergeschikte managers. Met dit onderzoek beoog ik twee doelen. Het eerste doel is om bij te dragen aan een beter inzicht in het begrip beoordelingsstijl, door rekening te houden met tekortkomingen van, kritiekpunten op en

problemen die genoemd zijn in de bestaande RAPM-literatuur. Een goed inzicht in het begrip beoordelingsstijl is noodzakelijk om het tweede doel van dit onderzoek te kunnen bereiken: bij te dragen aan een beter (theoretisch) inzicht in hoe en waarom beoordelingsstijl samenhangt met het gedrag van ondergeschikte managers. Om deze doelen te bereiken bestaat het onderzoek uit twee delen: de hoofdstukken 2, 3 en 4 bevatten een literatuuroverzicht en de hoofdstukken 5, 6, 7 en 8 beschrijven een empirisch onderzoek bij Van den Bergh Nederland (VDBN), een bedrijfsonderdeel van Unilever.

Hoofdstuk 2 geeft een kort overzicht van de literatuur die bekend staat als RAPM-onderzoek. Ook vat ik in dit hoofdstuk de kritiek op bestaande onderzoeken samen. Dit leidt tot de conclusie dat twee onderwerpen voor toekomstig onderzoek van belang zijn:

1. het definiëren en meten van verschillende beoordelingsstijlen die in de te bestuderen context van belang zijn;
2. het ontwikkelen van de theorie/ kennis over de wijze waarop en waarom verschillende beoordelingsstijlen gerelateerd zijn aan verschillen in gedrag van ondergeschikte managers.

Het eerste onderwerp -de definitie en meting van beoordelingsstijl- staat centraal in hoofdstuk 3. Dit hoofdstuk bespreekt eerst de verschillende manieren waarop beoordelingsstijl in bestaande onderzoeken is gedefinieerd en gemeten en geeft op basis daarvan een aantal implicaties aan voor toekomstig onderzoek naar beoordelingsstijlen. Aan de hand van een achttal interviews in een Nederlands dienstverlenend bedrijf is vervolgens de relevantie en waarde van de implicaties 'getoetst'. De belangrijkste conclusie op grond van het literatuuronderzoek is dat de betrouwbaarheid en validiteit van bestaande manieren om beoordelingsstijl te meten gering is. Bovendien blijkt het begrip beoordelingsstijl in de bestaande RAPM-literatuur eng gedefinieerd te zijn in termen van nadruk op financiële, op budgetten gebaseerde accountinginformatie. In het vakgebied van management accounting en management control is de laatste jaren de aandacht voor niet-financiële informatie en prestatimaatstaven als vervanging van of aanvulling op financiële informatie en prestatimaatstaven sterk toegenomen. Met het oog op die ontwikkeling lijken de manieren waarop beoordelingsstijlen in het verleden zijn gedefinieerd en gemeten niet langer relevant voor toekomstig onderzoek. Dit betekent dat beoordelingsstijl in toekomstig onderzoek opnieuw gedefinieerd en gemeten moet worden. Het onderzoek in dit hoofdstuk leidt tot de conclusie dat voor een goede en relevante definitie van beoordelingsstijlen voor een specifieke organisatie rekening gehouden moet worden met:

1. de mogelijkheid dat budgetten een geringe rol in prestatie-evaluatie spelen;
2. het verschil tussen het technisch ontwerp van control systemen (instrumenten) en de wijze waarop deze control systemen worden gebruikt, waarbij beoordelingsstijlen gedefinieerd moeten worden in termen van gebruik van de systemen;

3. de organisatorische context waarbinnen prestatiebeoordeling plaatsvindt, waardoor een onderzoek (field study) binnen één organisatie de voorkeur verdient boven onderzoek onder veel verschillende organisaties op basis van uitsluitend een schriftelijke vragenlijst;
4. de mogelijkheid dat de perceptie van ondergeschikte managers van de beoordelingsstijl van hun leidinggevende is vertekend, of dat verschillen in beoordelingsstijl niet op het niveau van individuen (ondergeschikte managers) maar op het niveau van de leider onderzocht moet worden.

Om recht te doen aan deze vier aandachtspunten is het nodig verschillen in beoordelingsstijlen te definiëren en te meten in de context van een specifieke organisatie. Op die manier kan het begrip beoordelingsstijl worden gedefinieerd in termen van het specifieke gebruik van het prestatiebeoordelingssysteem door een leidinggevende manager bij het beoordelen van de prestaties van een individuele ondergeschikte manager. Daarbij wordt de organisatorische context in het algemeen en het ontwerp van het prestatiebeoordelingssysteem in het bijzonder als gegeven beschouwd. Dit is de eerste stap naar een beter inzicht in hoe en waarom verschillen in beoordelingsstijl in die specifieke organisatie(context) wel of niet tot verschillen in gedrag bij ondergeschikte managers leiden.

Het tweede onderwerp - hoe en waarom beoordelingsstijl samenhangt met het gedrag van ondergeschikte managers - staat centraal in hoofdstuk 4. In dit hoofdstuk en het vervolg van dit proefschrift hanteer ik als uitgangspunt dat de beoordelingsstijl van een leidinggevende wordt gemeten door de perceptie van de ondergeschikte manager van die beoordelingsstijl te meten. De conclusies en de bevindingen zoals beschreven in hoofdstuk 3 (zie vooral aandachtspunt 4 hierboven) roepen echter de vraag op hoe betrouwbaar de perceptie van ondergeschikte managers is en welke factoren op voorhand aannemelijk maken dat ondergeschikte managers verschillen in perceptie van beoordelingsstijlen van hun leidinggevendens zullen rapporteren. Voordat ik aandacht besteed aan de samenhang van beoordelingsstijl en het gedrag van ondergeschikte managers besteed ik daarom in hoofdstuk 4 eerst aandacht aan factoren die een verklaring kunnen geven voor mogelijke verschillen in de perceptie van beoordelingsstijlen tussen ondergeschikte managers. Op grond van een overzicht van de bestaande literatuur kom ik tot de conclusie dat verschillen in de perceptie van beoordelingsstijlen kunnen worden verklaard op verschillende niveaus: op het niveau van organisaties, op het niveau van leiders (of groepen ondergeschikten die rapporteren aan dezelfde leider) en op het niveau van de individuele ondergeschikte. Op het niveau van organisaties gaat het om verschillen in perceptie van beoordelingsstijlen tussen ondergeschikte managers die afkomstig zijn uit verschillende organisaties. Factoren op organisatieniveau, zoals strategie, omgevingsonzekerheid en de financiële situatie van de organisatie, zijn van groot belang om de relevantie van de gekozen definitie en wijze van meten van beoordelingsstijl in die specifieke organisatie vast te kunnen stellen. Verschillen

in percepties van beoordelingsstijlen op dit niveau kunnen ook ontstaan doordat organisaties gebruik maken van verschillende beoordelingssystemen waardoor bijvoorbeeld de nadruk die leidinggevendenden leggen op het behalen van vooraf vastgestelde doelen (targets) tussen organisaties sterk kunnen verschillen. Om mogelijke invloed van factoren op organisatieniveau te voorkomen heb ik in hoofdstuk 3 beargumenteert dat het -op theoretische gronden- de voorkeur verdient onderzoek naar beoordelingsstijlen uit te voeren binnen één organisatie. Een onderzoek dat zich beperkt tot één organisatie maakt het uiteraard onmogelijk de invloed van factoren op organisatieniveau (statistisch) te toetsen. Deze factoren zullen echter niet alleen invloed op de perceptie van beoordelingsstijlen hebben, maar ook op de gedragsmatige gevolgen ervan. Daarom is een beschrijving van de organisatiecontext (factoren op organisatieniveau) ook in onderzoeken die zich beperken tot beoordelingsstijlen binnen één organisatie onmisbaar.

Factoren op het niveau van leiders bieden mogelijk een verklaring voor verschillen in perceptie van beoordelingsstijlen tussen ondergeschikte managers die rapporteren aan verschillende leidinggevendenden. Het niveau van leiders is een belangrijk niveau van onderzoek, omdat onderzoek naar beoordelingsstijlen per definitie veronderstelt dat leiders binnen een en dezelfde organisatie onderling verschillen zullen vertonen in de manier waarop ze hetzelfde beoordelingssysteem gebruiken. Het literatuuronderzoek laat echter zien dat de RAPM-literatuur na de eerste studies van Hopwood (1973) en Otley (1978) dit niveau van analyse volledig heeft genegeerd. Twee mogelijke contextfactoren die verschillen in gepercipieerde beoordelingsstijlen op dit niveau kunnen verklaren zijn de beoogde stijl van de leidinggevende en de mate van taakonzekerheid. Om de invloed van taakonzekerheid op de gepercipieerde beoordelingsstijl op het niveau van de leider te kunnen toetsen ontstaat echter een meetprobleem: om de gepercipieerde beoordelingsstijl op het niveau van de leider vast te stellen moeten de gerapporteerde beoordelingsstijlen van individuele ondergeschikten die aan dezelfde leider rapporteren geaggregeerd worden. Aggregatie van gegevens afkomstig van individuele managers is echter alleen geldig en betrouwbaar als er een aanzienlijke mate van overeenstemming tussen de individuele managers bestaat over de beoordelingsstijl van hun leider. Op grond van het literatuuronderzoek in dit hoofdstuk kom ik tot de conclusie dat enige mate van overeenstemming in gepercipieerde beoordelingsstijl tussen ondergeschikten die rapporteren aan dezelfde leider verwacht mag worden. Tegelijkertijd geeft het literatuuronderzoek ook aanleiding tot de verwachting dat leiders hun leiderschapsstijl en hun beoordelingsstijl zullen aanpassen aan individuele karakteristieken van ondergeschikten zoals leeftijd, werkervaring, opleiding en prestaties. In plaats van uitsluitend de perceptie van individuele ondergeschikte managers te meten verdient het daarom de voorkeur ook te onderzoeken waarom leiders al dan niet ondergeschikten op verschillende manieren beoordelen, door ook gegevens van leiders te verzamelen.

Vervolgens concludeer ik op grond van bestaande literatuur dat (de perceptie van) de beoordelingsstijl van een leider ook af zal hangen van (de perceptie van) de meer algemene leiderschapsstijl van de leider.

Ten slotte kom ik op grond van een literatuuronderzoek tot twee mogelijke verklaringen voor de relatie tussen beoordelingsstijlen en het gedrag van onderschikte managers. Een directe relatie tussen de perceptie van beoordelingsstijl en het gedrag van ondergeschikte managers ligt niet voor de hand. Waarschijnlijk is dat deze relatie afhangt van de mate waarin ondergeschikte managers hun beoordeling als eerlijk ervaren en de mate waarin ze vertrouwen in hun superieur hebben. Ik verwacht daarom een indirecte relatie tussen de perceptie van beoordelingsstijl en gedrag via de mate waarin de beoordeling als eerlijk wordt ervaren en de mate van vertrouwen van de ondergeschikte in de leidinggevende. Om te begrijpen waarom een bepaalde stijl als eerlijker ervaren wordt dan een andere stijl is de context waarbinnen de beoordeling plaatsvindt opnieuw van groot belang.

Na deze theoretische basis volgt in de hoofdstukken 5 t/m 8 het empirische deel van het proefschrift.

Hoofdstuk 5 beschrijft het ontwerp van een empirische studie binnen Van den Bergh Nederland (VDBN). Het onderzoek vindt plaats binnen twaalf zorgvuldige geselecteerde eenheden binnen VDBN. Het onderzoek omvat twaalf leiders en 57 ondergeschikte managers die elk aan een van deze twaalf leiders rapporteren. In het onderzoek zijn verschillende bronnen gebruikt voor de verzameling van onderzoeksgegevens: schriftelijke vragenlijsten die beantwoord zijn door 57 ondergeschikte managers, interviews met de twaalf leiders, interviews met één tot drie ondergeschikte managers voor elke leider (27 ondergeschikten in totaal), en interne documenten van de organisatie (in het bijzonder over het beoordelingssysteem). Het onderzoek is zo ontworpen dat het mij in staat stelt de onderwerpen die in het literatuuronderzoek in de voorgaande hoofdstukken naar voren kwamen te analyseren. Deze onderwerpen zijn als volgt samen te vatten:

1. De conceptualisatie en geldige meting van onderliggende dimensies van gepercipieerde beoordelingsstijl die relevant zijn voor de specifieke organisatie (zie hoofdstuk 3);
2. De mate van overeenstemming in gepercipieerde beoordelingsstijl binnen en tussen groepen van ondergeschikten die beoordeeld worden door dezelfde leidinggevende (§ 4.2.3);
3. Het verband tussen contextuele factoren op verschillende niveaus van analyse en gepercipieerde beoordelingsstijl (figuur 4.1);
4. Het verband tussen gepercipieerde leiderschapsstijl en gepercipieerde beoordelingsstijl (§ 4.3);

5. Het verband tussen gepercipieerde beoordelingsstijl en het gedrag van ondergeschikte managers (§ 4.4);
6. De rol van de organisatorische context, en het ontwerp van het beoordelingssysteem in het bijzonder.

Hoofdstuk 6 geeft een beschrijving van Van den Bergh Nederland (VDBN). Dit hoofdstuk biedt belangrijke informatie over de context waarbinnen de beoordeling plaatsvindt. Daarom is de informatie uit dit hoofdstuk van belang voor een goed begrip van de uiteindelijke resultaten in hoofdstuk 7 en 8. De beschrijving van het beoordelingssysteem van VDBN neemt een belangrijke plaats in dit hoofdstuk in.

Hoofdstuk 7 presenteert de resultaten van het empirische onderzoek naar relevante verschillen in beoordelingsstijlen binnen VDBN. Daarbij besteed ik aandacht aan de eerste twee van de zes bovengenoemde onderwerpen. De resultaten tonen allereerst dat er binnen VDBN geen relevant onderscheid in stijlen gemaakt kan worden op basis van verschillen in het gebruik van budgetten of targets, zoals gebruikelijk is in de RAPM-literatuur. Op grond van een analyse van gegevens uit de schriftelijke vragenlijst die is ingevuld door ondergeschikte managers blijken er drie relevantie dimensies van beoordelingsstijlen te zijn. De eerste dimensie (performance assessment) beschrijft de nadruk die een leidinggevende in de perceptie van ondergeschikten legt op het vormen van een oordeel over de geleverde prestaties. De tweede dimensie (performance development) beschrijft de mate waarin de leidinggevende in de perceptie van ondergeschikten aandacht heeft voor de ontwikkeling van prestaties. De derde dimensie (interpersonal evaluation) heeft betrekking op de mate van communicatie over en weer tussen de leidinggevende en de ondergeschikte wanneer de leidinggevende beslissingen neemt in het kader van prestatiebeoordeling.

Ondanks behoorlijke variantie in stijlen tussen groepen, laten de resultaten vervolgens slechts voor een van de drie dimensies van beoordelingsstijl, namelijk performance development, significante verschillen tussen groepen zien. Daarnaast blijken tussen ondergeschikten die rapporteren aan dezelfde leider behoorlijke verschillen in perceptie van beoordelingsstijl te bestaan. Deze resultaten suggereren dat verschillen in perceptie van beoordelingsstijlen voor een beperkt deel verklaard worden doordat ondergeschikten door een andere leider worden beoordeeld (groepsniveau), maar dat een groot deel van deze verschillen gebaseerd is op verschillen op het individuele niveau van ondergeschikten.

Hoofdstuk 8 presenteert de resultaten van het empirische onderzoek naar het verband van beoordelingsstijl met andere variabelen. Allereerst is in hoofdstuk 8 onderzocht of de verschillen in perceptie van beoordelingsstijlen verklaard kunnen worden door contextuele factoren. Om verschillen in perceptie op groepsniveau te verklaren is onderzocht of deze verschillen samenhangen met verschillen in het niveau van taakonzekerheid. De resultaten laten geen significante samenhang zien. Om verschillen in perceptie van beoordelingsstijlen

tussen individuele ondergeschikten binnen een groep te verklaren is onderzocht of er samenhang is met verschillen in leeftijd, werkervaring en opleiding. De resultaten tonen slechts voor een dimensie van beoordelingsstijl, namelijk performance assessment, een significante samenhang aan met werkervaring gemeten als de tijd dat een ondergeschikte onder zijn huidige leidinggevende werkt. Deze resultaten suggereren dat leiders hun beoordelingsstijl niet aanpassen aan leeftijd, ervaring en opleiding van hun ondergeschikten. Hoewel dat statistisch niet onderzocht is in dit proefschrift, suggereren de gegevens uit de interviews dat leeftijd, ervaring en opleiding slechts belangrijk zijn in relatie tot de werkelijke prestaties. De interviews geven namelijk aanleiding tot de voorzichtige conclusie dat leidinggevendenden hun beoordelingsstijl aanpassen aan de werkelijke prestaties van ondergeschikten gerelateerd aan hun leeftijd, ervaring en opleiding.

Vervolgens heb ik het verband tussen leiderschapsstijl en beoordelingsstijl onderzocht. De resultaten geven aan dat beoordelingsstijl en leiderschapsstijl sterk samenhangen. Wel blijken er, zoals verwacht, verschillen in de precieze samenhang van elk van de drie verschillende dimensies van beoordelingsstijl met verschillende kenmerken van leiderschapsstijl.

Ten slotte beschrijft hoofdstuk 8 de resultaten van het onderzoek naar het verband tussen beoordelingsstijl en het gedrag van ondergeschikte managers. Hoewel de gepercipieerde beoordelingsstijl binnen VDBN een sterke samenhang vertoont met de ervaren eerlijkheid van prestatiebeoordeling en het vertrouwen van ondergeschikten in hun leidinggevende, lijkt er geen verband te zijn tussen beoordelingsstijl en baangerelateerde spanning of tevredenheid. Ook de interviews leiden tot de conclusie dat verschillen in beoordelingsstijl binnen VDBN niet samenhangen met functioneel (lerend) of disfunctioneel gedrag. Evenals bij het verband met leiderschapsstijl, verschilt de precieze samenhang van elk van de drie dimensies van beoordelingsstijl met de ervaren eerlijkheid van de beoordeling en vertrouwen in de leidinggevende. De resultaten geven aan dat performance development een directe significant positieve relatie heeft met vertrouwen in de leidinggevende, maar geen significant verband heeft met de ervaren eerlijkheid van de beoordeling. Performance assessment en interpersonal evaluation hebben beide een significant positieve relatie met de ervaren eerlijkheid van de beoordeling, wat vervolgens leidt tot een indirect effect op vertrouwen in de leidinggevende.

Hoofdstuk 9 vormt de afsluiting van het proefschrift en geeft de belangrijkste conclusies weer. In dit hoofdstuk worden eerst de resultaten uit de hoofdstukken 7 en 8 over de vijf onderwerpen samengevat en vervolgens besproken in het licht van de organisatorische context. De resultaten onderstrepen het grote belang van de organisatorische context voor een juist inzicht in beoordelingsstijlen en het verband met het gedrag van ondergeschikte managers. Vervolgens bespreek ik de betekenis van dit proefschrift voor wetenschap en praktijk, ik ga in op de beperkingen van het beschreven onderzoek en uitgevoerde analyses, en ik sluit af met een aantal suggesties voor verder onderzoek. De belangrijkste conclusie is

dat dit proefschrift inderdaad een belangrijke bijdrage levert aan de twee gestelde doelen: het verbeteren van inzicht in het begrip beoordelingsstijl en het verbeteren van het inzicht in hoe en waarom beoordelingsstijl samenhangt met het gedrag van ondergeschikte managers. Hoewel de resultaten uit het onderzoek context-specifiek zijn en dus op zich niet generaliseerbaar buiten VDBN, blijken de zes onderwerpen die uit het onderzoek naar voren komen en de methoden waarop ik deze onderwerpen binnen een specifieke organisatorische context heb onderzocht wel generaliseerbaar. Context-specifiek onderzoek naar beoordelingsstijlen en de gedragsmatige gevolgen ervan blijft ook in de toekomst relevant, omdat uit het onderzoek in dit proefschrift blijkt dat om het gedrag van medewerkers te beïnvloeden niet alleen het ontwerp van een prestatiebeoordelingssysteem van belang is, maar zeker ook de wijze waarop dat systeem door leidinggevendenden gebruikt wordt.

Biography



Jan Noeverman was born on June 6, 1971 in Haarlemmermeer, the Netherlands. He attended Gymnasium Camphusianum in Gorinchem, from which he obtained his Gymnasium diploma in 1989. From 1989 to 1995 Jan studied Economics at the Erasmus University Rotterdam. He received his Master's degree in Business Economics in 1995 with a thesis on the dysfunctional consequences of management control systems. In 1995 he joined the section Financial and Managerial Accounting of the Erasmus School of Economics, first as a Ph.D. student and since 1999 as Assistant Professor in Management Accounting.

Jan presented his research at several national and international conferences and seminars, including the European Accounting Association (1996; 1997; 2003; 2004), British Accounting Association (1998), and the Management Accounting Section Midyear Conference of the American Accounting Association (2006). His work has been published in MAB and Qualitative Research in Accounting and Management (QRAM).

Jan's teaching experience is broad, ranging from plenary lectures in the first bachelor year to small seminars for Master students at the university, with topics ranging from Cost and Management Accounting and Management Control to Financial Accounting. From 2001 till 2005 he was also involved with NIVE as a lecturer for the Management Control classes that are part of the HOFAM curriculum.

Besides his work at the university, Jan has been involved in many social, religious and political activities. For example, Jan was a member of the municipal council of Prins-Alexander in Rotterdam from 2002-2006, and he has been elder in the Dutch Reformed Church in Rotterdam-Centrum since 2000.

Currently, Jan is Deputy Mayor of Prins-Alexander in Rotterdam and part-time assistant professor in Management Accounting. His research interests mainly focus on behavioural and motivational aspects of design and use of management control systems, with a particular emphasis on performance measurement and the role of leaders in performance evaluation.

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ERIM PH.D. SERIES
RESEARCH IN MANAGEMENT

ERIM Electronic Series Portal: <http://hdl.handle.net/1765/1>

Althuizen, N.A.P., *Analogical Reasoning as a Decision Support Principle for Weakly Structured Marketing Problems*, Promotor: Prof. dr. ir. B. Wierenga, EPS-2006-095-MKT, ISBN 90-5892-129-8, <http://hdl.handle.net/1765/8190>

Alvarez, H.L., *Distributed Collaborative Learning Communities Enabled by Information Communication Technology*, Promotor: Prof. dr. K. Kumar, EPS-2006-080-LIS, ISBN 90-5892-112-3, <http://hdl.handle.net/1765/7830>

Appelman, J.H., *Governance of Global Interorganizational Tourism Networks: Changing Forms of Co-ordination between the Travel Agency and Aviation Sector*, Promotors: Prof. dr. F.M. Go & Prof. dr. B. Nooteboom, EPS-2004-036-MKT, ISBN 90-5892-060-7, <http://hdl.handle.net/1765/1199>

Baquero, G., *On Hedge Fund Performance, Capital Flows and Investor Psychology*, Promotor: Prof. dr. M.J.C.M. Verbeek, EPS-2006-094-F&A, ISBN 90-5892-131-X, <http://hdl.handle.net/1765/8192>

Berens, G., *Corporate Branding: The Development of Corporate Associations and their Influence on Stakeholder Reactions*, Promotor: Prof. dr. C. B. M. van Riel, EPS-2004-039-ORG, ISBN 90-5892-065-8, <http://hdl.handle.net/1765/1273>

Berghe, D.A.F. van den, *Working Across Borders: Multinational Enterprises and the Internationalization of Employment*, Promotors: Prof. dr. R.J.M. van Tulder & Prof. dr. E.J.J. Schenk, EPS-2003-029-ORG, ISBN 90-5892-05-34, <http://hdl.handle.net/1765/1041>

Berghman, L.A., *Strategic Innovation Capacity: A Mixed Method Study on Deliberate Strategic Learning Mechanisms*, Promotor: Prof. dr. P. Mattysens, EPS-2006-087-MKT, ISBN 90-5892-120-4, <http://hdl.handle.net/1765/7991>

Bijman, W.J.J., *Essays on Agricultural Co-operatives: Governance Structure in Fruit and Vegetable Chains*, Promotor: Prof. dr. G.W.J. Hendrikse, EPS-2002-015-ORG, ISBN: 90-5892-024-0, <http://hdl.handle.net/1765/867>

Bispo, A., *Labour Market Segmentation: An investigation into the Dutch hospitality industry*, Promotors: Prof. dr. G.H.M. Evers & Prof. dr. A.R. Thurik, EPS-2007-108-ORG, ISBN: 90-5892-136-9, <http://hdl.handle.net/1765/10283>

Blindenbach-Driessen, F., *Innovation Management in Project-Based Firms*, Promotor: Prof. dr. S.L. van de Velde, EPS-2006-082-LIS, ISBN: 90-5892-110-7, <http://hdl.handle.net/1765/7828>

Boer, C.A., *Distributed Simulation in Industry*, Promotors: Prof. dr. A. de Bruin & Prof. dr. ir. A. Verbraeck, EPS-2005-065-LIS, ISBN: 90-5892-093-3, <http://hdl.handle.net/1765/6925>

Boer, N.I., *Knowledge Sharing within Organizations: A situated and Relational Perspective*, Promotor: Prof. dr. K. Kumar, EPS-2005-060-LIS, ISBN: 90-5892-086-0, <http://hdl.handle.net/1765/6770>

Brito, M.P. de, *Managing Reverse Logistics or Reversing Logistics Management?* Promotors: Prof. dr. ir. R. Dekker & Prof. dr. M. B. M. de Koster, EPS-2004-035-LIS, ISBN: 90-5892-058-5, <http://hdl.handle.net/1765/1132>

Brohm, R., *Polycentric Order in Organizations: A Dialogue between Michael Polanyi and IT-Consultants on Knowledge, Morality, and Organization*, Promotors: Prof. dr. G. W. J. Hendrikse & Prof. dr. H. K. Letiche, EPS-2005-063-ORG, ISBN: 90-5892-095-X, <http://hdl.handle.net/1765/6911>

Campbell, R.A.J., *Rethinking Risk in International Financial Markets*, Promotor: Prof. dr. C.G. Koedijk, EPS-2001-005-F&A, ISBN: 90-5892-008-9, <http://hdl.handle.net/1765/306>

Chen, Y., *Labour Flexibility in China's Companies: An Empirical Study*, Promotors: Prof. dr. A. Buitendam & Prof. dr. B. Krug, EPS-2001-006-ORG, ISBN: 90-5892-012-7, <http://hdl.handle.net/1765/307>

Damen, F.J.A., *Taking the Lead: The Role of Affect in Leadership Effectiveness*, Promotor: Prof. dr. D.L. van Knippenberg, EPS-2007-107-ORG, <http://hdl.handle.net/1765/10282>

Daniševská, P., *Empirical Studies on Financial Intermediation and Corporate Policies*, Promotor: Prof. dr. C.G. Koedijk, EPS-2004-044-F&A, ISBN 90-5892-070-4, <http://hdl.handle.net/1765/1518>

Delporte-Vermeiren, D.J.E., *Improving the Flexibility and Profitability of ICT-enabled Business Networks: An Assessment Method and Tool*, Promotors: Prof. mr. dr. P.H.M. Vervest & Prof. dr. ir. H.W.G.M. van Heck, EPS-2003-020-LIS, ISBN: 90-5892-040-2, <http://hdl.handle.net/1765/359>

Derwall, J.M.M., *The Economic Virtues of SRI and CSR*, Promotor: Prof. dr. C.G. Koedijk, EPS-2007-101-STR, ISBN: 90-5892-132-8, <http://hdl.handle.net/1765/8986>

Dijksterhuis, M., *Organizational Dynamics of Cognition and Action in the Changing Dutch and US Banking Industries*, Promotors: Prof. dr. ir. F.A.J. van den Bosch & Prof. dr. H.W. Volberda, EPS-2003-026-STR, ISBN: 90-5892-048-8, <http://hdl.handle.net/1765/1037>

Fenema, P.C. van, *Coordination and Control of Globally Distributed Software Projects*, Promotor: Prof. dr. K. Kumar, EPS-2002-019-LIS, ISBN: 90-5892-030-5, <http://hdl.handle.net/1765/360>

Fleischmann, M., *Quantitative Models for Reverse Logistics*, Promotors: Prof. dr. ir. J.A.E.E. van Nunen & Prof. dr. ir. R. Dekker, EPS-2000-002-LIS, ISBN: 35-4041-711-7, <http://hdl.handle.net/1765/1044>

Flier, B., *Strategic Renewal of European Financial Incumbents: Coevolution of Environmental Selection, Institutional Effects, and Managerial Intentionality*, Promotors: Prof. dr. ir. F.A.J. van den Bosch & Prof. dr. H.W. Volberda, EPS-2003-033-STR, ISBN: 90-5892-055-0, <http://hdl.handle.net/1765/1071>

Fok, D., *Advanced Econometric Marketing Models*, Promotor: Prof. dr. P.H.B.F. Franses, EPS-2003-027-MKT, ISBN: 90-5892-049-6, <http://hdl.handle.net/1765/1035>

Ganzaroli, A., *Creating Trust between Local and Global Systems*, Promotors: Prof. dr. K. Kumar & Prof. dr. R.M. Lee, EPS-2002-018-LIS, ISBN: 90-5892-031-3, <http://hdl.handle.net/1765/361>

Gilsing, V.A., *Exploration, Exploitation and Co-evolution in Innovation Networks*, Promotors: Prof. dr. B. Nooteboom & Prof. dr. J.P.M. Groenewegen, EPS-2003-032-ORG, ISBN 90-5892-054-2, <http://hdl.handle.net/1765/1040>

Ginkel, W.P. van, *The Use of Distributed Information in Decision Making Groups: The Role of Shared Task Representations*, Promotor: Prof. dr. D. van Knippenberg, EPS-2007-097-ORG, <http://hdl.handle.net/1765/8424>

Govers, R., *Virtual Tourism Destination Image: Glocal Identities Constructed, Perceived and Experienced*, Promotors: Prof. dr. F.M. Go & Prof. dr. K. Kumar, EPS-2005-069-MKT, ISBN 90-5892-107-7, <http://hdl.handle.net/1765/6981>

Graaf, G. de, *Tractable Morality: Customer Discourses of Bankers, Veterinarians and Charity Workers*, Promotors: Prof. dr. F. Leijnse & Prof. dr. T. van Willigenburg, EPS-2003-031-ORG, ISBN: 90-5892-051-8, <http://hdl.handle.net/1765/1038>

Groot, E.A. de, *Essays on Economic Cycles*, Promotors: Prof. dr. Ph.H.B.F. Franses & Prof. Prof. dr. H.R. Commandeur, EPS-2006-091-MKT, ISBN: 90-5892-123-9, <http://hdl.handle.net/1765/8216>

Gutkowska, A.B., *Essays on the Dynamic Portfolio Choice*, Promotor: Prof. dr. A.C.F. Vorst, EPS-2006-085-F&A, ISBN: 90-5892-118-2, <http://hdl.handle.net/1765/7994>

Hagemeyer, R.E., *The Unmasking of the Other*, Promotors: Prof. dr. S.J. Magala & Prof. dr. H.K. Letiche, EPS-2005-068-ORG, ISBN: 90-5892-097-6, <http://hdl.handle.net/1765/6963>

Hartigh, E. den, *Increasing Returns and Firm Performance: An Empirical Study*, Promotor: Prof. dr. H.R. Commandeur, EPS-2005-067-STR, ISBN: 90-5892-098-4, <http://hdl.handle.net/1765/6939>

Hermans, J.M., *ICT in Information Services: Use and Deployment of the Dutch Securities Trade, 1860-1970*, Promotor: Prof. dr. F.H.A. Janszen, EPS-2004-046-ORG, ISBN 90-5892-072-0, <http://hdl.handle.net/1765/1793>

Heugens, P.P.M.A.R., *Strategic Issues Management: Implications for Corporate Performance*, Promotors: Prof. dr. ir. F.A.J. van den Bosch & Prof. dr. C.B.M. van Riel, EPS-2001-007-STR, ISBN: 90-5892-009-7, <http://hdl.handle.net/1765/358>

Heuvel, W. van den, *The Economic Lot-Sizing Problem: New Results and Extensions*, Promotor: Prof. dr. A.P.L. Wagelmans, EPS-2006-093-LIS, ISBN: 90-5892-124-7, <http://hdl.handle.net/1765/1805>

Hooghiemstra, R., *The Construction of Reality: Cultural differences in self-serving behaviour in accounting narratives*, Promotors: Prof. dr. L.G. van der Tas RA & Prof. dr. A.Th.H. Pruyn, EPS-2003-025-F&A, ISBN: 90-5892-047-X, <http://hdl.handle.net/1765/871>

Hu, Y., *Essays on the Governance of Agricultural Products: Cooperatives and Contract Farming*, Promotors: Prof. dr. G.W.J. Hendrkse & Prof. Dr. B. Krug, EPS-2007-113-ORG, ISBN: 90-5892-145-1, <http://hdl.handle.net/1765/1>

Huij, J.J., *New Insights into Mutual Funds: Performance and Family Strategies*, Promotor: Prof. dr. M.C.J.M. Verbeek, EPS-2007-099-F&A, ISBN: 90-5892-134-4, <http://hdl.handle.net/1765/9398>

Huurman, C.I., *Dealing with Electricity Prices*, Promotor: Prof. dr. C.D. Koedijk, EPS-2007-098-F&A, ISBN: 90-5892-130-1, <http://hdl.handle.net/1765/9399>

Iastrebova, K., *Manager's Information Overload: The Impact of Coping Strategies on Decision-Making Performance*, Promotor: Prof. dr. H.G. van Dissel, EPS-2006-077-LIS, ISBN 90-5892-111-5, <http://hdl.handle.net/1765/7329>

Iwaarden, J.D. van, *Changing Quality Controls: The Effects of Increasing Product Variety and Shortening Product Life Cycles*, Promotors: Prof. dr. B.G. Dale & Prof. dr. A.R.T. Williams, EPS-2006-084-ORG, ISBN 90-5892-117-4, <http://hdl.handle.net/1765/7992>

Jansen, J.J.P., *Ambidextrous Organizations*, Promotors: Prof. dr. ir. F.A.J. Van den Bosch & Prof. dr. H.W. Volberda, EPS-2005-055-STR, ISBN 90-5892-081-X, <http://hdl.handle.net/1765/6774>

Jong, C. de, *Dealing with Derivatives: Studies on the Role, Informational Content and Pricing of Financial Derivatives*, Promotor: Prof. dr. C.G. Koedijk, EPS-2003-023-F&A, ISBN: 90-5892-043-7, <http://hdl.handle.net/1765/1043>

Keizer, A.B., *The Changing Logic of Japanese Employment Practices: A Firm-Level Analysis of Four Industries*, Promotors: Prof. dr. J.A. Stam & Prof. dr. J.P.M. Groenewegen, EPS-2005-057-ORG, ISBN: 90-5892-087-9, <http://hdl.handle.net/1765/6667>

Kijkuit, R.C., *Social Networks in the Front End: The Organizational Life of an Idea*, Promotor: Prof. dr. B. Nooteboom, EPS-2007-104-ORG, ISBN 90-5892-137-6, <http://hdl.handle.net/1765/10074>

Kippers, J., *Empirical Studies on Cash Payments*, Promotor: Prof. dr. Ph.H.B.F. Franses, EPS-2004-043-F&A, ISBN 90-5892-069-0, <http://hdl.handle.net/1765/1520>

Kole, E., *On Crises, Crashes and Comovements*, Promotors: Prof. dr. C.G. Koedijk & Prof. dr. M.J.C.M. Verbeek, EPS-2006-083-F&A, ISBN 90-5892-114-X, <http://hdl.handle.net/1765/7829>

Kooij-de Bode, J.M., *Distributed Information and Group Decision-Making: Effects of Diversity and Affect*, Promotor: Prof. dr. D.L. van Knippenberg, EPS-2007-115-ORG, <http://hdl.handle.net/1765/1>

Knapp, S., *The Econometrics of Maritime Safety: Recommendations to Enhance Safety at Sea*, Promotor: Prof. dr. Ph.H.B.F. Franses, EPS-2007-096-ORG, ISBN 90-5892-127-1, <http://hdl.handle.net/1765/7913>

Koppius, O.R., *Information Architecture and Electronic Market Performance*, Promotors: Prof. dr. P.H.M. Vervest & Prof. dr. ir. H.W.G.M. van Heck, EPS-2002-013-LIS, ISBN: 90-5892-023-2, <http://hdl.handle.net/1765/921>

Kotlarsky, J., *Management of Globally Distributed Component-Based Software Development Projects*, Promotor: Prof. dr. K. Kumar, EPS-2005-059-LIS, ISBN: 90-5892-088-7, <http://hdl.handle.net/1765/6772>

Kuilman, J., *The Re-Emergence of Foreign Banks in Shanghai: An Ecological Analysis*, Promotor: Prof. dr. B. Krug, EPS-2005-066-ORG, ISBN: 90-5892-096-8, <http://hdl.handle.net/1765/6926>

Langen, P.W. de, *The Performance of Seaport Clusters: A Framework to Analyze Cluster Performance and an Application to the Seaport Clusters of Durban, Rotterdam and the Lower Mississippi*, Promotors: Prof. dr. B. Nooteboom & Prof. drs. H.W.H. Welters, EPS-2004-034-LIS, ISBN: 90-5892-056-9, <http://hdl.handle.net/1765/1133>

Le Anh, T., *Intelligent Control of Vehicle-Based Internal Transport Systems*, Promotors: Prof. dr. M.B.M. de Koster & Prof. dr. ir. R. Dekker, EPS-2005-051-LIS, ISBN 90-5892-079-8, <http://hdl.handle.net/1765/6554>

Le-Duc, T., *Design and Control of Efficient Order Picking Processes*, Promotor: Prof. dr. M.B.M. de Koster, EPS-2005-064-LIS, ISBN 90-5892-094-1, <http://hdl.handle.net/1765/6910>

Leeuwen, E.P. van, *Recovered-Resource Dependent Industries and the Strategic Renewal of Incumbent Firm: A Multi-Level Study of Recovered Resource Dependence Management and Strategic Renewal in the European Paper and Board Industry*, Promotors: Prof. dr. ir. F.A.J. Van den Bosch & Prof. dr. H.W. Volberda, EPS-2007-109-STR, ISBN 90-5892-140-6, <http://hdl.handle.net/1765/10183>

Lentink, R.M., *Algorithmic Decision Support for Shunt Planning*, Promotors: Prof. dr. L.G. Kroon & Prof. dr. ir. J.A.E.E. van Nunen, EPS-2006-073-LIS, ISBN 90-5892-104-2, <http://hdl.handle.net/1765/7328>

Liang, G., *New Competition: Foreign Direct Investment and Industrial Development in China*, Promotor: Prof. dr. R.J.M. van Tulder, EPS-2004-047-ORG, ISBN 90-5892-073-9, <http://hdl.handle.net/1765/1795>

Liere, D.W. van, *Network Horizon and the Dynamics of Network Positions: A Multi-Method Multi-Level Longitudinal Study of Interfirm Networks*, Promotor: Prof. dr. P.H.M. Vervest, EPS-2007-105-LIS, ISBN 90-5892-139-0, <http://hdl.handle.net/1765/10181>

Loef, J., *Incongruity between Ads and Consumer Expectations of Advertising*, Promotors: Prof. dr. W.F. van Raaij & Prof. dr. G. Antonides, EPS-2002-017-MKT, ISBN: 90-5892-028-3, <http://hdl.handle.net/1765/869>

Londoño, M. del Pilar, *Institutional Arrangements that Affect Free Trade Agreements: Economic Rationality Versus Interest Groups*, Promotors: Prof. dr. H.E. Haralambides & Prof. dr. J.F. Francois, EPS-2006-078-LIS, ISBN: 90-5892-108-5, <http://hdl.handle.net/1765/7578>

Maeseneire, W., de, *Essays on Firm Valuation and Value Appropriation*, Promotor: Prof. dr. J.T.J. Smit, EPS-2005-053-F&A, ISBN 90-5892-082-8, <http://hdl.handle.net/1765/6768>

Mandele, L.M., van der, *Leadership and the Inflection Point: A Longitudinal Perspective*, Promotors: Prof. dr. H.W. Volberda & Prof. dr. H.R. Commandeur, EPS-2004-042-STR, ISBN 90-5892-067-4, <http://hdl.handle.net/1765/1302>

Meer, J.R. van der, *Operational Control of Internal Transport*, Promotors: Prof. dr. M.B.M. de Koster & Prof. dr. ir. R. Dekker, EPS-2000-001-LIS, ISBN: 90-5892-004-6, <http://hdl.handle.net/1765/859>

Mentink, A., *Essays on Corporate Bonds*, Promotor: Prof. dr. A.C.F. Vorst, EPS-2005-070-F&A, ISBN: 90-5892-100-X, <http://hdl.handle.net/1765/7121>

Meyer, R.J.H., *Mapping the Mind of the Strategist: A Quantitative Methodology for Measuring the Strategic Beliefs of Executives*, Promotor: Prof. dr. R.J.M. van Tulder, EPS-2007-106-STR, ISBN: 978-90-5892-141-3, <http://hdl.handle.net/1765/10182>

Miltenburg, P.R., *Effects of Modular Sourcing on Manufacturing Flexibility in the Automotive Industry: A Study among German OEMs*, Promotors: Prof. dr. J. Paauwe & Prof. dr. H.R. Commandeur, EPS-2003-030-ORG, ISBN: 90-5892-052-6, <http://hdl.handle.net/1765/1039>

Moerman, G.A., *Empirical Studies on Asset Pricing and Banking in the Euro Area*, Promotor: Prof. dr. C.G. Koedijk, EPS-2005-058-F&A, ISBN: 90-5892-090-9, <http://hdl.handle.net/1765/6666>

Mol, M.M., *Outsourcing, Supplier-relations and Internationalisation: Global Source Strategy as a Chinese Puzzle*, Promotor: Prof. dr. R.J.M. van Tulder, EPS-2001-010-ORG, ISBN: 90-5892-014-3, <http://hdl.handle.net/1765/355>

Mom, T.J.M., *Managers' Exploration and Exploitation Activities: The Influence of Organizational Factors and Knowledge Inflows*, Promotors: Prof. dr. ir. F.A.J. Van den Bosch & Prof. dr. H.W. Volberda, EPS-2006-079-STR, ISBN: 90-5892-116-6, <http://hdl.handle.net/1765>

Mulder, A., *Government Dilemmas in the Private Provision of Public Goods*, Promotor: Prof. dr. R.J.M. van Tulder, EPS-2004-045-ORG, ISBN: 90-5892-071-2, <http://hdl.handle.net/1765/1790>

Muller, A.R., *The Rise of Regionalism: Core Company Strategies Under The Second Wave of Integration*, Promotor: Prof. dr. R.J.M. van Tulder, EPS-2004-038-ORG, ISBN 90-5892-062-3, <http://hdl.handle.net/1765/1272>

Oosterhout, J., van, *The Quest for Legitimacy: On Authority and Responsibility in Governance*, Promotors: Prof. dr. T. van Willigenburg & Prof. mr. H.R. van Gunsteren, EPS-2002-012-ORG, ISBN: 90-5892-022-4, <http://hdl.handle.net/1765/362>

Paape, L., *Corporate Governance: The Impact on the Role, Position, and Scope of Services of the Internal Audit Function*, Promotors: Prof. dr. G.J. van der Pijl & Prof. dr. H. Commandeur, EPS-2007-111-MKT, ISBN: 90-5892-143-7, <http://hdl.handle.net/1765/10417>

Pak, K., *Revenue Management: New Features and Models*, Promotor: Prof. dr. ir. R. Dekker, EPS-2005-061-LIS, ISBN: 90-5892-092-5, <http://hdl.handle.net/1765/362/6771>

Pattikawa, L.H., *Innovation in the Pharmaceutical Industry: Evidence from Drug Introduction in the U.S.*, Promotors: Prof. dr. H.R. Commandeur, EPS-2007-102-MKT, ISBN: 90-5892-135-2, <http://hdl.handle.net/1765/9626>

Peeters, L.W.P., *Cyclic Railway Timetable Optimization*, Promotors: Prof. dr. L.G. Kroon & Prof. dr. ir. J.A.E.E. van Nunen, EPS-2003-022-LIS, ISBN: 90-5892-042-9, <http://hdl.handle.net/1765/429>

Pietersz, R., *Pricing Models for Bermudan-style Interest Rate Derivatives*, Promotors: Prof. dr. A.A.J. Pelsser & Prof. dr. A.C.F. Vorst, EPS-2005-071-F&A, ISBN 90-5892-099-2, <http://hdl.handle.net/1765/7122>

Popova, V., *Knowledge Discovery and Monotonicity*, Promotor: Prof. dr. A. de Bruin, EPS-2004-037-LIS, ISBN 90-5892-061-5, <http://hdl.handle.net/1765/1201>

Pouchkarev, I., *Performance Evaluation of Constrained Portfolios*, Promotors: Prof. dr. J. Spronk & Dr. W.G.P.M. Hallerbach, EPS-2005-052-F&A, ISBN 90-5892-083-6, <http://hdl.handle.net/1765/6731>

Puvanavari Ratnasingam, P., *Interorganizational Trust in Business to Business E-Commerce*, Promotors: Prof. dr. K. Kumar & Prof. dr. H.G. van Dissel, EPS-2001-009-LIS, ISBN: 90-5892-017-8, <http://hdl.handle.net/1765/356>

- Rinsum, M. van, *Performance Measurement and Managerial Time Orientation*, Promotor: Prof. dr. F.G.H. Hartmann, EPS-2006-088-F&A, ISBN: 90-5892-121-2, <http://hdl.handle.net/1765/7993>
- Romero Morales, D., *Optimization Problems in Supply Chain Management*, Promotors: Prof. dr. ir. J.A.E.E. van Nunen & Dr. H.E. Romeijn, EPS-2000-003-LIS, ISBN: 90-9014078-6, <http://hdl.handle.net/1765/865>
- Roodbergen, K.J., *Layout and Routing Methods for Warehouses*, Promotors: Prof. dr. M.B.M. de Koster & Prof. dr. ir. J.A.E.E. van Nunen, EPS-2001-004-LIS, ISBN: 90-5892-005-4, <http://hdl.handle.net/1765/861>
- Schramade, W.L.J., *Corporate Bonds Issuers*, Promotor: Prof. dr. A. De Jong, EPS-2006-092-F&A, ISBN: 90-5892-125-5, <http://hdl.handle.net/1765/8191>
- Schweizer, T.S., *An Individual Psychology of Novelty-Seeking, Creativity and Innovation*, Promotor: Prof. dr. R.J.M. van Tulder, EPS-2004-048-ORG, ISBN: 90-5892-07-71, <http://hdl.handle.net/1765/1818>
- Six, F.E., *Trust and Trouble: Building Interpersonal Trust Within Organizations*, Promotors: Prof. dr. B. Nooteboom & Prof. dr. A.M. Sorge, EPS-2004-040-ORG, ISBN 90-5892-064-X, <http://hdl.handle.net/1765/1271>
- Slager, A.M.H., *Banking across Borders*, Promotors: Prof. dr. R.J.M. van Tulder & Prof. dr. D.M.N. van Wensveen, EPS-2004-041-ORG, ISBN 90-5892-066-6, <http://hdl.handle.net/1765/1301>
- Sloot, L., *Understanding Consumer Reactions to Assortment Unavailability*, Promotors: Prof. dr. H.R. Commandeur, Prof. dr. E. Peelen & Prof. dr. P.C. Verhoef, EPS-2006-074-MKT, ISBN 90-5892-102-6, <http://hdl.handle.net/1765/7438>
- Smit, W., *Market Information Sharing in Channel Relationships: Its Nature, Antecedents and Consequences*, Promotors: Prof. dr. ir. G.H. van Bruggen & Prof. dr. ir. B. Wierenga, EPS-2006-076-MKT, ISBN 90-5892-106-9, <http://hdl.handle.net/1765/7327>
- Sonnenberg, M., *The Signalling Effect of HRM on Psychological Contracts of Employees: A Multi-level Perspective*, Promotor: Prof. dr. J. Paauwe, EPS-2006-086-ORG, ISBN 90-5892-119-0, <http://hdl.handle.net/1765/7995>
- Speklé, R.F., *Beyond Generics: A closer Look at Hybrid and Hierarchical Governance*, Promotor: Prof. dr. M.A. van Hoepen RA, EPS-2001-008-F&A, ISBN: 90-5892-011-9, <http://hdl.handle.net/1765/357>
- Teunter, L.H., *Analysis of Sales Promotion Effects on Household Purchase Behavior*, Promotors: Prof. dr. ir. B. Wierenga & Prof. dr. T. Kloek, EPS-2002-016-MKT, ISBN: 90-5892-029-1, <http://hdl.handle.net/1765/868>
- Tims, B., *Empirical Studies on Exchange Rate Puzzles and Volatility*, Promotor: Prof. dr. C.G. Koedijk, EPS-2006-089-F&A, ISBN: 90-5892-113-1, <http://hdl.handle.net/1765/8066>
- Valck, K. de, *Virtual Communities of Consumption: Networks of Consumer Knowledge and Companionship*, Promotors: Prof. dr. ir. G.H. van Bruggen & Prof. dr. ir. B. Wierenga, EPS-2005-050-MKT, ISBN 90-5892-078-X, <http://hdl.handle.net/1765/6663>
- Verheul, I., *Is There a (Fe)male Approach? Understanding Gender Differences in Entrepreneurship*, Prof. dr. A.R. Thurik, EPS-2005-054-ORG, ISBN 90-5892-080-1, <http://hdl.handle.net/1765/2005>

Vis, I.F.A., *Planning and Control Concepts for Material Handling Systems*, Promotors: Prof. dr. M.B.M. de Koster & Prof. dr. ir. R. Dekker, EPS-2002-014-LIS, ISBN: 90-5892-021-6, <http://hdl.handle.net/1765/866>

Vlaar, P.W.L., *Making Sense of Formalization in Interorganizational Relationships: Beyond Coordination and Control*, Promotors: Prof. dr. ir. F.A.J. Van den Bosch & Prof. dr. H.W. Volberda, EPS-2006-075-STR, ISBN 90-5892-103-4, <http://hdl.handle.net/1765/7326>

Vliet, P. van, *Downside Risk and Empirical Asset Pricing*, Promotor: Prof. dr. G.T. Post, EPS-2004-049-F&A, ISBN 90-5892-07-55, <http://hdl.handle.net/1765/1819>

Vlist, P. van der, *Synchronizing the Retail Supply Chain*, Promotors: Prof. dr. ir. J.A.E.E. van Nunen & Prof. dr. A.G. de Kok, EPS-2007-110-LIS, ISBN 90-5892-142-0, <http://hdl.handle.net/1765/10418>

Vries-van Ketel E. de, *How Assortment Variety Affects Assortment Attractiveness: A Consumer Perspective*, Promotors: Prof. dr. G.H. van Bruggen & Prof. dr. ir. A. Smids, EPS-2006-072-MKT, ISBN 90-5892-101-8, <http://hdl.handle.net/1765/7193>

Vromans, M.J.C.M., *Reliability of Railway Systems*, Promotors: Prof. dr. L.G. Kroon, Prof. dr. ir. R. Dekker & Prof. dr. ir. J.A.E.E. van Nunen, EPS-2005-062-LIS, ISBN: 90-5892-089-5, <http://hdl.handle.net/1765/6773>

Vroomen, B.L.K., *The Effects of the Internet, Recommendation Quality and Decision Strategies on Consumer Choice*, Promotor: Prof. dr. Ph.H.B.F. Franses, EPS-2006-090-MKT, ISBN: 90-5892-122-0, <http://hdl.handle.net/1765/8067>

Waal, T. de, *Processing of Erroneous and Unsafe Data*, Promotor: Prof. dr. ir. R. Dekker, EPS-2003-024-LIS, ISBN: 90-5892-045-3, <http://hdl.handle.net/1765/870>

Watkins Fassler, K., *Macroeconomic Crisis and Firm Performance*, Promotors: Prof. dr. J. Spronk & Prof. dr. D.J. van Dijk, EPS-2007-103-F&A, ISBN 90-5892-138-3, <http://hdl.handle.net/1765/10065>

Wennekers, A.R.M., *Entrepreneurship at Country Level: Economic and Non-Economic Determinants*, Promotor: Prof. dr. A.R. Thurik, EPS-2006-81-ORG, ISBN 90-5892-115-8, <http://hdl.handle.net/1765/7982>

Wielemaker, M.W., *Managing Initiatives: A Synthesis of the Conditioning and Knowledge-Creating View*, Promotors: Prof. dr. H.W. Volberda & Prof. dr. C.W.F. Baden-Fuller, EPS-2003-28-STR, ISBN 90-5892-050-X, <http://hdl.handle.net/1765/1042>

Wijk, R.A.J.L. van, *Organizing Knowledge in Internal Networks: A Multilevel Study*, Promotor: Prof. dr. ir. F.A.J. van den Bosch, EPS-2003-021-STR, ISBN: 90-5892-039-9, <http://hdl.handle.net/1765/347>

Zhang, X., *Strategizing of Foreign Firms in China: An Institution-based Perspective*, Promotor: Prof. dr. B. Krug, EPS-2007-114-ORG, ISBN: 90-5892-147-5, <http://hdl.handle.net/1765/1>

Zhu, Z., *Essays on China's Tax System*, Promotors: Prof. dr. B. Krug & Prof. dr. G.W.J. Hendrikse, EPS-2007-112-ORG, ISBN: 90-5892-144-4, <http://hdl.handle.net/1765/1>

Management Control Systems, Evaluative Style, and Behaviour

Exploring the Concept and Behavioural Consequences of Evaluative Style

Organisations develop and implement performance measurement and performance evaluation systems to motivate employees to take actions that -in the end- improve organisational (financial) performance. But do these systems really influence employee behaviour as intended? This thesis shows that to answer that question not only the design of the system should be considered, but also the manner in which managers within an organisation use the system. This book describes a study on the influence of evaluative style of leaders on subordinates' behaviour. Evaluative style refers to the manner in which a leader evaluates the performance of subordinates, controlling for the specific design of the performance evaluation system and the broader organisational context. This study improves our understanding of the concept of evaluative style by defining evaluative style, similar to leadership style, as a (behavioural) characteristic of leaders. The study also improves our understanding of the influence of evaluative style on subordinates' behaviour. A literature review first reveals six topics that help improve this understanding. Subsequently, these six topics are investigated in an empirical study at Van den Bergh Netherlands (VDBN) that involves twelve leaders and their subordinates. The findings show that different evaluative styles exist within VDBN. These styles influence subordinate's trust in superior and perceived fairness of evaluation, but no effect is found on functional (learning) or dysfunctional behaviour. These findings are context-specific and cannot be generalised outside the context of VDBN. However, the six topics and the way in which these topics have been investigated within a specific organisational context can be generalised and are relevant for future research.

ERIM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are RSM Erasmus University and the Erasmus School of Economics. ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focussed on the management of the firm in its environment, its intra- and inter-firm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced graduate program in Research in Management. Within ERIM, over two hundred senior researchers and Ph.D. candidates are active in the different research programs. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.