

Rossing, Morten

Doctoral Thesis

Local Adaption and Meaning Creation in Performance Appraisal

PhD Series, No. 29.2013

Provided in Cooperation with:

Copenhagen Business School (CBS)

Suggested Citation: Rossing, Morten (2013) : Local Adaption and Meaning Creation in Performance Appraisal, PhD Series, No. 29.2013, ISBN 9788792977717, Copenhagen Business School (CBS), Frederiksberg, <https://hdl.handle.net/10398/8761>

This Version is available at:

<https://hdl.handle.net/10419/208864>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/3.0/>

COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN
SOLBJERG PLADS 3
DK-2000 FREDERIKSBERG
DANMARK

www.cbs.dk

ISSN 0906-6934

Print ISBN: 978-87-92977-70-0

Online ISBN: 978-87-92977-71-7

Local Adaption and Meaning Creation in Performance Appraisal

PhD Series 29.2013



**Copenhagen
Business School**
HANDELSHØJSKOLEN

Local Adaption and Meaning Creation in Performance Appraisal

Morten Rossing

Doctoral School of Organisation
and Management Studies

PhD Series 29.2013

Local Adaption and Meaning Creation in Performance Appraisal

"Objectives are not fate; they are direction. They are not commands; they are commitments. They do not determine the future; they are means to mobilize the resources and energies of the business for the making of the future."

- Peter Drucker

Morten Rossing
*Local Adaption and Meaning Creation
in Performance Appraisal*

1st edition 2013
PhD Series 29.2013

© The Author

ISSN 0906-6934

Print ISBN: 978-87-92977-70-0

Online ISBN: 978-87-92977-71-7

The Doctoral School of Organisation and Management Studies (OMS) is an interdisciplinary research environment at Copenhagen Business School for PhD students working on theoretical and empirical themes related to the organisation and management of private, public and voluntary organizations.

All rights reserved.

No parts of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without permission in writing from the publisher.

Acknowledgements

There are many people I would like to thank for their help with this piece of work. First, I simply would not have made it without the support of my two supervisors; Associate Professor Christopher John Mathieu, Department of Sociology, Lund University and Associate Professor Magnus Larsson, Department of Organization, Copenhagen Business School. Their persistent support, creative ideas and constructive critique helped me tremendously in all phases of the project.

I am grateful for the support provided by Professor Robert Austin, Dean of the Faculty of Business Administration at University of New Brunswick, whose book "Measuring and Managing Performance in Organizations" has been a great inspiration for me, and who supported me with many valuable comments at my closing seminar.

Similarly, I am grateful for the constructive and supportive comments and challenging questions I received at my project presentation seminars from Professor Jan Mouritsen, Head of Department of Operations Management at Copenhagen Business School and from Professor Paul Du Gay, Department of Organization, Copenhagen Business School.

Also, I would like to thank all faculty and fellow PhD students who have taken time to read and comment on my project presentations during PhD courses and conferences. Many ideas were fostered by these comments.

Indeed, I am very grateful to the managers and employees in PharmaComp who stood up for a series of in depth interviews on top of all their daily tasks and who were so enthusiastic about sharing their views on performance appraisal. Likewise, my gratitude goes to colleagues in PharmaComp's Corporate HR who helped me with providing important documentation and sparring, particularly in the early, forming phase of my project. Thank you also to my own management, who showed me the trust to support the idea of embarking on the PhD journey in the first place.

Last, but certainly not least, I am indebted to my wife, Camilla, who supported and encouraged me all the way. The project was undertaken while I was working full time in PharmaComp. Without Camilla's support, I would not have dared to embark on the project, and I would for sure not have managed to fulfil it either.

Preface

I became interested in the subject of Performance Appraisal (PA) long before deciding to pursue the opportunity to become an Independent PhD Fellow at Copenhagen Business School. For more than a decade, I have worked as a manager in different private organisations that all subscribed to performance appraisal as a systematised process of setting targets for and evaluating the performance of individuals, for the purpose of improving organisational performance.

Both in my role as super-ordinate, as well as in my role as subordinate, I have had mixed experiences with performance appraisal. On the one hand it seems so intuitively logical and right; to agree up front on measurable targets that a later evaluation will be based on, for the benefit of transparency, fairness and well-founded management decisions about reward, promotion, dismissal and more.

On the other hand the tangible experiences are so mixed. Too often I have felt that employees have left appraisal meetings with disappointed expectations because we did not have the same interpretation of how the individual had contributed over the past year. Too often I have felt that employees or managers have been disengaged and unprepared for the process, although the importance of the process in terms of reward allocation and career opportunities should be clear to everyone. And too often I have found that the sum of outcomes on each of an employee's targets did not add up to what my gut feeling told me that the employee had in reality contributed. Sometimes, the employee had done so much more and I felt I had to twist the system to provide the higher rating the employee deserved. And sometimes the opposite was the case.

Of course, these personal experiences could have different causes. One reason could be that I was simply lacking the management capabilities required to handle a simple and straight-forward management tool. However, many dialogues with peers, supervisors and employees who all in different ways shared my mixed experiences told me that the cause likely lies somewhere else. More likely, the truth is that it is not easy to conduct performance appraisal in such a way that targets and evaluations remain controlled and transparent, so that the employees are more motivated than they would have been without it and so that visible organisational performance improvement is the natural outcome of the significant resources put into the process.

Since I joined corporate HR in a large pharmaceutical organisation nearly five years ago I have been working with the further development and support of a centrally defined, globally standardised performance appraisal process. This closer intimacy with the practice of performance appraisal seen from the corporate perspective made me curious to learn more about the experiences of other organisations, and the status of current research into the field. I started reading books and articles on performance appraisal and related theoretical themes like measurement theory, motivation theory and justice theory. From this, it became clear to me that my experiences with the process were not unique. In fact, the subject of performance appraisal is a very well-researched phenomenon, in terms of the number of problems related to the process that have been dealt with in research over the past couple of decades.

However, I also found that for some reason, large parts of the research into the subject of performance appraisal were strikingly uniform. Uniform in terms of methodology, uniform in terms of empirical subject and uniform in terms of narrowly focusing on one or a few single problems in performance appraisal, and analysis of these based on rational economic models without questioning basic assumptions. This left me with a feeling that the ground would be fertile for digging further into the field of performance appraisal to supplement the large volume of narrow focused literature with a more holistic, comparative contribution. So, I decided to do just that, and the present PhD thesis is the result of the following efforts through a three year period from 2010 to 2013.

A Brief Reader's Guide

This thesis is structured into eleven chapters. In the introduction in chapter 1, I will start by presenting the proposition, the phenomenon, the research question and important definitions.

Chapter 2 contains an introduction to the conversation in the literature that I am taking part in. After a high level categorisation of the performance appraisal literature into what I have called a dominating and a radical paradigm, I will present a more detailed analysis of a sample of 125 journal articles from the 20 year period 1990-2010. I will examine the state of the literature in terms of its empirical, methodological and theoretical focus. Further, I will provide an overview of the more than 50 different problems in PA dealt with in the journal articles.

In chapter 3, I will present my methodological approach and my data. This thesis is based on a single case study through a triangulating approach, including participant observation, study of archival material and analysis of a quantitative data set. First and foremost, though, the study is based on 38 qualitative, deep-dive interviews, of which 34 interviews were part of a longitudinal design where dyads of managers and employees in four different workforces were followed during a one year, full performance appraisal cycle. Finally, I will also provide a detailed account of the data coding process through nVivo9, and for documentation and archiving procedures utilised during the study.

Chapter 4 provides an introduction to performance appraisal in the case organisation. The historical background and history of performance appraisal is presented in three stages. The first is the introduction of a standard corporate performance appraisal scheme in 2004. The second is the introduction of a supporting IT system in 2006. The third is the more advanced steps planned and fulfilled in recent years.

In chapter 5 the macro view of performance appraisal in the case organisation is presented. Based mainly on quantitative analysis of a data set covering all electronically captured performance appraisal ratings from the period 2006-2011, supplemented by analysis of archival material from the case organisation, I will draw up the macro situation in the case organisation on a number of dimensions; the distribution of ratings over time and by workforce, the correlation between hierarchy and rating, and the correlation between rating and retention, wage and bonus. I will

also introduce the connection between the design of a corporate performance appraisal system and the corporate culture. Finally, I will introduce the approach chosen in different workforces regarding how to balance business and behavioural performance. The purpose of drawing the macro view in this fairly detailed way is to prepare for the subsequent analysis of local adaption and meaning creation. By having a good understanding of the environments within which actors operate and create meaning, we will be better equipped to analyse the meaning creation of managers and employees in performance appraisal.

Chapter 6 introduces the theoretical toolbox for the meaning creation analysis. First, I give an introduction to Weick's (1995) sensemaking theory, particularly by relating Weick's seven properties of sensemaking to the process of performance appraisal. This will exemplify how Weick's framework will be helpful when we want to understand meaning creation as a process of enactment of environmental differences within which managers and employees operate. In chapter 6 I will also present a categorisation of the task at hand in four workforces included in the study, against Austin's (1996) two management regimes; management by measurement, which is the regime within which performance appraisal is designed, and the alternative; management by delegation. By this, I will have established the framework against which I will investigate patterns of local adaption and meaning creation in performance appraisal across four workforces.

Chapters 7 to 9 contain the detailed analysis of local adaption and meaning creation in performance appraisal in the case organisation across target setting, midyear review and final appraisal. In each of these chapters, a number of challenges in performance appraisal are analysed. In chapter 7 regarding target setting, the challenges of establishing commitment and including behaviour are analysed. In chapter 8 regarding midyear review, the challenges of conveying meaningful feedback, of using rating in feedback, of carelessness and of management disruption are analysed. Chapter 9 regarding final appraisal analyses the challenges of evaluating business performance, evaluating behavioural performance, designing rating distribution policy and of calibration of ratings.

Finally, chapter 10 contains the conclusions and implications of the study, and chapter 11 contains important reservations, qualifications and guidance for future research. Appendices are found in Chapter 12, followed by the list of literature.

Table of Content

ACKNOWLEDGEMENTS.....	III
PREFACE	IV
A BRIEF READER'S GUIDE.....	VI
TABLE OF CONTENT.....	VIII
1 INTRODUCTION	1
1.1 THE PROPOSITION.....	1
1.2 THE PHENOMENON	4
1.3 THE RESEARCH QUESTION AND DEFINITIONS.....	11
2 THE PA CONVERSATION - A LITERATURE REVIEW	14
2.1 THE PERFORMANCE APPRAISAL ATTITUDE DICHOTOMY	14
2.2 INTRODUCING THE LITERATURE REVIEW	30
2.3 THE LITERATURE REVIEW APPROACH.....	31
2.4 ANALYSING THE EMPIRICAL FOCUS OF THE LITERATURE.....	34
2.5 ANALYSING THE METHODOLOGICAL APPROACH OF THE LITERATURE.....	41
2.6 ANALYSING THE THEORETICAL APPROACH OF THE LITERATURE	50
2.7 THE ODD ARTICLES: THE RADICAL PERSPECTIVE.....	65
2.8 PERFORMANCE APPRAISAL - WHAT ARE THE PROBLEMS?.....	69
2.9 SUMMARY OF LITERATURE REVIEW	74
2.10 THE GAP IN THE LITERATURE: DIRECTIONS FOR THE STUDY.....	77
3 METHODOLOGY AND DATA	81
3.1 PARTICIPANT OBSERVATION	82
3.2 ARCHIVAL RECORDS AND DOCUMENTATION	86
3.3 INTERVIEWS	88
3.4 CODING: DIGITAL ANALYSIS OF QUALITATIVE INTERVIEW DATA IN NVivo9.....	98
3.5 PROCEDURES FOR RESEARCH AND DOCUMENTATION	107
3.6 SUMMARY	112
4 CASE INTRODUCTION	115
4.1 PHASE 1: ESTABLISHING A COMMON PA PROCESS (2004)	117
4.2 PHASE 2: IMPLEMENTING THE IT SYSTEM (2006)	127
4.3 PHASE 3: AIMING FOR STRATEGIC PERFORMANCE MANAGEMENT (2007-).....	128

4.4	SUMMARY	132
5	THE MACRO VIEW.....	133
5.1	PERFORMANCE RATINGS AT PHARMACOMP OVER TIME.....	135
5.2	PERFORMANCE RATING DISTRIBUTION BY WORKFORCE.....	136
5.3	PERFORMANCE RATING AND HIERARCHY	141
5.4	PERFORMANCE RATING AND RETENTION	145
5.5	PERFORMANCE RATING AND WAGE INCREASE & SIZE OF BONUS.....	147
5.6	PERFORMANCE RATING AND CORPORATE CULTURE	149
5.7	PERFORMANCE RATING - THE BALANCE OF RESULTS AND BEHAVIOUR	152
5.8	SUMMARY AND CONCLUSION.....	157
6	MEANING CREATION AND THE TASK AT HAND.....	160
6.1	PERFORMANCE APPRAISAL AS IDENTITY CONSTRUCTION	164
6.2	PERFORMANCE APPRAISAL AS RETROSPECTIVE	168
6.3	PERFORMANCE APPRAISAL AS ENACTING ENVIRONMENTS	170
6.4	PERFORMANCE APPRAISAL AS A SOCIAL PHENOMENON.....	172
6.5	PERFORMANCE APPRAISAL AS AN ON-GOING PROCESS	173
6.6	PERFORMANCE APPRAISAL AS EXTRACTION OF CUES.....	175
6.7	PERFORMANCE APPRAISAL AS DRIVEN BY PLAUSIBILITY RATHER THAN ACCURACY	176
6.8	THE TASK AT HAND - PERFORMANCE APPRAISAL ACROSS WORKFORCES	178
7	LOCAL ADAPTION AND MEANING CREATION IN TARGET SETTING.....	190
7.1	INTRODUCTION.....	190
7.2	THE CHALLENGE OF ESTABLISHING COMMITMENT THROUGH TARGET SETTING	191
7.3	THE CHALLENGE OF HANDLING BEHAVIOUR IN TARGET SETTING.....	204
8	LOCAL ADAPTION AND MEANING CREATION IN MIDYEAR REVIEW	219
8.1	INTRODUCTION.....	219
8.2	THE CHALLENGE OF CONVEYING MEANINGFUL FEEDBACK.....	220
8.3	THE CHALLENGE OF USING RATING IN FEEDBACK.....	225
8.4	THE CHALLENGE OF CARELESSNESS	237
8.5	THE CHALLENGE OF DISRUPTION	247

9 LOCAL ADAPTION AND MEANING CREATION IN FINAL APPRAISAL.....	254
9.1 INTRODUCTION.....	254
9.2 THE CHALLENGE OF EVALUATING BUSINESS PERFORMANCE	255
9.3 THE CHALLENGE OF EVALUATING BEHAVIOUR	262
9.4 THE CHALLENGE OF RATING DISTRIBUTION POLICY.....	271
9.5 THE CHALLENGE OF CALIBRATION	283
10 CONCLUSIONS AND IMPLICATIONS.....	303
11 QUALIFICATIONS AND GUIDANCE FOR FUTURE RESEARCH	315
12 APPENDICES.....	321
12.1 APPENDIX 1: ABSTRACT IN ENGLISH	321
12.2 APPENDIX 2: ABSTRACT IN DANISH	322
12.3 APPENDIX 3: LETTER TO SELECTED HR PARTNERS.....	323
12.4 APPENDIX 4: LETTERS TO PARTICIPATING MANAGERS AND EMPLOYEES	325
12.5 APPENDIX 5: INTERVIEW GUIDES	326
12.6 APPENDIX 6: MANAGER’S AND EMPLOYEES’ DIARIES	330
12.7 APPENDIX 7: LIST OF JOURNALS REPRESENTED IN LITERATURE REVIEW	332
13 LITERATURE.....	334

List of Figures

FIGURE 1-1: THE MEANING CREATION FEEDBACK MODEL	4
FIGURE 2-1: DISTRIBUTION OF EMPIRICAL FOCUS BY CONTINENT	37
FIGURE 2-2: DISTRIBUTION OF EMPIRICAL FOCUS BY WORKFORCE	38
FIGURE 2-3: DISTRIBUTION OF METHODOLOGIES UTILISED IN 125 ANALYSED ARTICLES.....	41
FIGURE 2-4: DISTRIBUTION OF ARTICLE OCCURRENCES ON THEORETICAL PERSPECTIVES	52
FIGURE 2-5 OVERVIEW OF PA PROBLEMS IDENTIFIED IN THE LITERATURE REVIEW	72
FIGURE 3-1: DISTRIBUTION OF PARTICIPATION IN THREE INTERVIEW ROUNDS	97
FIGURE 3-2: FULL CODING STRUCTURE FOR QUALITATIVE INTERVIEWS CODING IN nVIVO9 ...	101
FIGURE 4-1: PHARMAComp’s PA JOURNEY AS SET OUT IN 2004.....	117
FIGURE 4-2: THE CORPORATE PA TEMPLATE.....	118
FIGURE 4-3: INITIAL PA PROCESS DEFINITION.....	122
FIGURE 5-1: NUMBER OF EMPLOYEES IN DATASET AND AVERAGE PA RATING, BY YEAR	135
FIGURE 5-2: RECOMMENDED & REALISED RATING DISTRIBUTION IN PRODUCTION, 2008/09 ..	137
FIGURE 5-3: RECOMMENDED AND REALISED RATING DISTRIBUTION IN STAFFS AREA 2008/09	138
FIGURE 5-4: 2010 PERFORMANCE RATING DISTRIBUTION BY WORKFORCE	140
FIGURE 5-5: 2010 PERFORMANCE RATING DISTRIBUTION BY HIERARCHICAL LEVEL.....	142
FIGURE 5-6: ONE YEAR RETENTION RATIOS PER PERFORMANCE RATING, 2008 RATINGS.	145
FIGURE 5-7: RESULTS AND BEHAVIOUR CONVERTED TO PA RATINGS, PRODUCTION.....	153
FIGURE 5-8: PRODUCTION GUIDELINE FOR EVALUATING BUSINESS RESULTS AND BEHAVIOUR	154
FIGURE 5-9: RESULTS AND BEHAVIOUR CONVERTED TO PA RATINGS, GUIDELINE FOR SALES	155
FIGURE 7-1: DISTRIBUTION AND INVOLVEMENT - ALTERNATIVE WAYS TO COMMIT TARGETS.	203
FIGURE 7-2: BEHAVIOURAL TARGETS APPROACH DEPENDENT ON IMPERATIVE	216
FIGURE 7-3: RELOCATION AND ADAPTION - HANDLING BEHAVIOUR IN TARGET SETTING	217
FIGURE 8-1: DISTANCING AND ENLARGEMENT - TWO WAYS TO CONVEY FEEDBACK	224
FIGURE 8-2: MEDIATOR OR ACCOUNTANT - DIFFERENT HANDLING OF DYADIC DISRUPTION....	251
FIGURE 9-1: 'CHANGING THE RULES...' AS MEANING CREATION FEEDBACK PROCESS.....	262
FIGURE 9-2: RELATIVE OR ABSOLUTE - TWO DIFFERENT APPROACHES TO RATING	282
FIGURE 9-3: THE CALIBRATION PROCESS IN R&D	290
FIGURE 9-4: TWO DIFFERENT CALIBRATION PROCESSES ACROSS WORKFORCES	292
FIGURE 9-5: INCLUSION OR NON-INCLUSION OF SELF-EVALUATIONS IN CALIBRATION	301

List of Tables

TABLE 2-1: KEY DISTINCTIONS BETWEEN THE DOMINANT AND THE RADICAL PARADIGMS.....	28
TABLE 6-1: MEASUREMENT COST DRIVERS BY TASK CHARACTERISTICS AND WORKFORCE.....	183
TABLE 6-2: DELEGATION COST DRIVERS BY TASK CHARACTERISTICS AND WORKFORCE.....	187
TABLE 8-1: MAPPING OF SIX STORIES AND THEIR RELATION TO MANAGEMENT REGIME.....	236
TABLE 9-1: SUMMARY OF STORIES ABOUT BEHAVIOUR IN PA, BY WORKFORCE.....	270
TABLE 10-1: TEN CHALLENGES ACROSS THE THREE PHASES OF PA.....	303
TABLE 10-2: SUMMARY OF PATTERNS OF LOCAL ADAPTION AND MEANING CREATION	306
TABLE 10-3: RELATIONS BETWEEN CHALLENGES AND PROBLEMS IN PA IN TARGET SETTING .	312

1 Introduction

In this chapter I will start by presenting the main proposition of the study. Hopefully, this will raise the reader's interest and motivation to dig further into the subject of local adaption and meaning creation¹ in performance appraisal (PA) across workforces. The proposition will be followed by an introduction to the broader phenomenon of PA, and a presentation of the research question and important definitions.

1.1 *The Proposition*

I will argue that it is well established in the literature that PA in modern organisations is infused with lots of problems and that managers and employees have good reason to be frustrated with the process. However, at the same time PA is one of the most institutionalised features of running large corporations everywhere in the Western world. The institution is to a remarkable degree consistently reproduced by business scholars, management consultants, HR professionals and senior managers across sectors, industries and organisations.

Over the past 20 years, more and more subtle problems in PA have been identified, isolated and analysed by scholars without any clear breakthrough in establishing viable solutions to the problems, or in establishing scientifically solid evidence that PA altogether does much to improve organisational performance (Kohn, 1999). Similarly, organisations in practice have continuously strived to optimise their PA process by seeking to eliminate individual problems in PA. But so far the most visible effect of these efforts appears to have been that other problems, old or new, seem to surface as soon as the previous ones have been attended to.

I claim that there is a paradox between the volumes of problems dealt with by academics and practitioners on the one hand and on the other hand the high degree of institutionalisation of PA in modern organisations. Importantly, this paradox is not only apparent in the theoretical field of PA, but also in practice. The problems in PA

¹ Local adaption and meaning creation is a composite concept covering how agents (managers, employees and workforces) create different meaningful stories and local adaption out of a corporate one-size-fits-all PA scheme, in ways that fit their particularities and task at hand.

challenge frames of meaning for managers and employees and provoke important meaning creation processes that affect their perception of the PA process. The paradox is the result of adhering to PA despite the lack of ability to resolve these problems.

Further, I claim that one fruitful way to observe and analyse challenges in PA, local adaption and meaning creation processes is to pay attention to patterns of enactment in differing sub-environments in large organisations. One important place to look for such demarcations of environmental differentiation is between workforces. Through the study of local adaption and meaning creation in PA across four different workforces over a full one-year PA-cycle in a case organisation, I found important differences in environmental dynamics and in manager and employee enactment patterns. I propose that understanding these patterns is not only important for our understanding of how PA schemes affect individuals' and organisations' performance. It is also more fundamentally important to our understanding of the dynamics of the problems in PA which have been uncovered, analysed and attempted to be solved over the last 20 years by scholarly research into PA.

By using Austin's (1996) separation between management by delegation versus management by measurement, a typology of differences in the task at hand in four workforces in a case organisation sets the stage for an in-depth meaning creation analysis across these workforces throughout an entire one-year PA cycle. Although PA as such is clearly designed to operate within the management by measurement paradigm, comparative findings across workforces show that important differences in dynamics are present. Particularly, Sales and R&D form contrasting cases in terms of characteristics of the task at hand, and characteristics of PA adaption. Thus, the present study shows how a standard corporate PA scheme can be twisted in different directions in different workforces to such an extent that in R&D it can even be argued that the scheme in some ways has more similarities with management by delegation than with management by measurement.

This is important, because the clearly dominant approach in the literature on PA is to study large organisations as if they were single entities, while in reality they consist of highly differentiated sub-entities, each adapting standard systems differently in ways that are meaningful in their local contexts.

By reaching a better understanding of this process of local adaption of a standard PA system we also add to the understanding of the paradox in PA. I propose that differences in the task at hand in different workforces form different sub-environments that enable local adaption of the standard PA system. In theoretical terms, local adaption means that particular local practices are established, supported by local meaning creation through stories that make the adapted PA system more meaningful in the local contexts. Thus, while PA schemes are ubiquitous in modern organisations and embedded with many problems, local adaption of PA is an important way for practitioners to establish local meaningful practices and stories about PA that counter the apparent lack of meaning of the standardised PA scheme.

This is also an important proposition in a broader theoretical and practice context because the present study of PA is just one example of how one through the study of meaning creation processes can analyse local adaption of standardised management systems. While the present study illustrates how differences in the task at hand can be an important explanatory factor to understand significant differences in local adaption of a standard corporate PA system, there is reason to believe that similar differences would surface if one studied other standard management systems.

The corporate PA system is not only enacted differently across workforces. The present study also shows how differences in enactment patterns in themselves articulate different problems in PA in a complex feedback process, so that different workforces experience different problems in PA, deal with them differently and thereby experience different feedback mechanisms. These feedback mechanisms again provoke different new problems in PA, or work to reproduce old problems in PA.

To capture these dynamics and interdependencies, I propose the analytical model illustrated in Figure 1-1 below. I call the model the *meaning creation feedback model* because it shows the interaction between problems in PA, workforce specific enactment, meaning scheme challenge to individuals, meaning creation and a feedback loop to new problem articulation in PA.

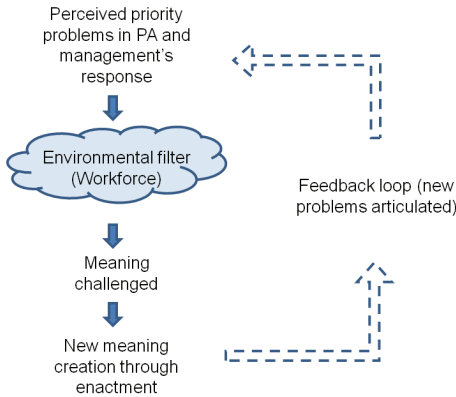


Figure 1-1: The meaning creation feedback model

I propose that what is needed by both practice and theory is to put aside for a while the ever more subtle analysis of isolated problems in PA and turn attention towards dealing with problems in PA at a more dynamic level where local enactment patterns and problem interdependencies through feedback processes come into focus. One fruitful way to do this is to look at how PA affects the meaning creation of managers and employees working within PA in different ways depending on workforce affiliation, and how these enactment patterns again interplay with articulation of other problems in PA.

Thus, although my contribution obviously builds on the existing literature about problems in PA, it attempts to go beyond the more common practice of examining narrow, single problems in PA by introducing a framework for analysing patterns of local adaption and meaning creation through enactment of workforce environmental filters and through feedback processes.

1.2 The Phenomenon

PA can be defined as *the appraisal rating of individuals' work performance and potentially their behaviours by management, covering a specific time period, applied to all employees or specific groups of employees whose participation is typically mandatory or alternatively motivated by access to extrinsic reward, and where results*

*in the form of ratings are stored by the organisation to be used for purposes that require differentiation of employees.*²

PA processes can vary in the details of their design, but as the definition indicates they also have important elements in common. Like e.g. Armstrong and Baron (2005: 23) do, most schemes operate within a rather generic PA process model based on a sequence of sub-processes ranging from "performance planning" and definition of "objectives" over "measuring", "assessing" and "rating" to "documentation".

Although the term "Performance Appraisal" is of newer origin, formal processes for talent differentiation, which is the core content of the concept, goes much further back. It would be far beyond the purpose of this thesis to give anything like a full account of the history of PA, but a few illustrative historical examples may be helpful for understanding how PA has developed into what it is today. This also serves as a short warm up to the literature review of the last 20 years' research into PA that will be presented in the following chapter.

Thus, a very brief genealogy of PA could for example start approximately 1400 years B.C. with the Old Testament, as Wiese and Buckley (1998) do, where Moses evaluates and selects:

...the man who was known to be the most skilled craftsman from the tribes of Israel to build and furnish the tabernacle of the Lord.

Later, in the Chinese Han Dynasty from 200 B.C. merit exams were introduced for selection and promotion of public officials, and "Imperial Raters" were appointed to rate high office holders on a nine point scale (Wren, 1994).

With industrialisation, PA schemes were beginning to take shape in more modern forms from around 1800. So-called "Silent monitors" were introduced in the cotton mills in the shape of colour coded blocks of wood at each work station that made it visible to everyone what level each individual worker was performing at (Wiese and Buckley, 1998).

Wiese and Buckley (1998: 235) specifically identify 1813 as the year when PA formally was introduced in the US. This was done by an army general, who:

² This definition is an elaboration of Coens & Jenkins' (2002: 13-14) five-point "characteristics" of PA.

1. Introduction

Submitted an evaluation of each of his men to the U.S. War Department... The Army General used a global rating, with descriptions of his men such as 'a good-natured man' or 'a knave despised by all'.

In his study of administrative history in the US 1869-1901 called "The Republican Era", White (1958: 353) documents how promotional exams were introduced to avoid unjust promotions, by counterweighing:

importunate solicitations and coercive influence from the outside and prejudice, favouritism, or corruption in the part of the appointing officers.

The documentation of performance was made in the form of "efficiency records" so that:

we can ascertain with almost mathematical certainty the proportion of work done by each clerk in the Pension office in point of quantity as well as quality.³

However, a few years later, the evaluation of the efficiency records initiative by a Department of the Interior Committee was less favourable and it was decided to cancel the efficiency records initiative because of inflationary ratings caused by dynamics not unknown to our days' researchers into PA in modern organisations:

The elaborate system of efficiency ratings and records, based, as it necessarily must be, upon the individual opinion of the officer giving the marks, has become little more than the mathematically expressed opinion of such officer as to the relative standing of the clerks under his supervision... the head of each division in the great majority of cases giving the highest ratings possible in order to put the clerks in his division ahead of or at least on an equal plane with the clerks of other divisions. (White, 1958: 358)

Further, White mentions an illustrative example of how administrative sub-optimisation was caused by different departments in the US Navy being responsible for defining their own deliverables, e.g. for the design of the "hull, the engines, the guns and the sail power" (White, 1958: 164). Although each department's targets and plans may have looked fine on paper and although all targets were met for each department, the combined result was disastrous:

After the ship Omaha had been commissioned and was ready for sea it was discovered that the several bureaus, working independently, had so completely appropriated the available space as to leave coal room for only four days' steaming... The Omaha was at best merely an old repaired wooden vessel, whose rebuilding cost the full price of an up-to-date steel ship, and which could neither fight nor run away from any ship of a foreign nation.

³ The quote is from an internal government correspondence dated 1879, in White (1958: 356).

In her book "The March of Folly" (Tuchman, 1984) gives a number of similar historical examples from Troy to the Vietnam War of wartime mismanagement, among which some have ties to critiques of PA schemes in that they are dealing with situations where targets are set and aimed for although this has an obvious counterproductive effect on organisational performance. In her view, folly is the "child of power" (Tuchman, 1984: 38) in the sense that powerful people define strategies and targets and cling to them long after it has become evident that this is not in the interest of themselves or the organisations they lead.

Turning back to civilian organisations, Chester Barnard in his "The Functions of the Executive" (1938) theorised over how managers can persuade employees to perform to certain standards. In Barnard's analysis, if external incentives are not available in enough quantity to satisfy the need to motivate employees, management must turn to "persuasion". Barnard lists three types of persuasion; creation of coercive conditions, rationalisation of opportunity and inculcation of motives. All three have linkages to PA schemes.

First, coercion can come in the shape of exclusion. Grades of exclusion vary and can e.g. be "homicide, outlawing, ostracism, corporal punishment, incarceration, withholding of specific benefits, discharge, etc." (Barnard, 1938: 149). Although Barnard finds it generally accepted that coercion does not to any great extent support performance in organisations of his age, he nevertheless emphasises how setting examples through coercion can be an effective management tool:

to create fear among those not directly affected so that they will be disposed to render to an organization certain contributions (Barnard, 1938: 149).

This is an argument not so far away from e.g. Kohn's (1999) perception of PA being a punitive system aiming to build motivation from installing fear of low ratings.

Second, the "rationalisation of other incentives" equals what Barnard (1938: 150) terms "propaganda", a term with much relevance in the international political context of his book in the late 1930s. This is a process through which leaders:

Convince individuals or groups that they 'ought', 'it is to their interest' to perform services or conform to requirements of specific organisations.

This second type of persuasion:

consists in emphasizing opportunities for satisfaction that are offered, usually in contrast to those available otherwise (Barnard, 1938: 150-152).

It is not hard to see the analogy to the close relation between PA and monetary incentives like bonuses, wage increases etc. where it is also the "Do this to get that" pedagogy that is used.

The third and most important type of persuasion is the inculcation of motives. Methods for this includes "Precept, example, suggestion, imitation or emulation, habitual attitudes" etc. These are all elements that can be recognised as possible ingredients in modern PA schemes in the shape of business and behavioural standards, targets and evaluation criteria. However, already in 1938, Barnard was aware of some of the possible pitfalls of incentive schemes:

Opportunity for personal prestige as an incentive for one person necessarily involves a relative depression of others (Barnard, 1938: 156).

Nevertheless, Barnard (1938: 160) concludes in what can be seen as a strong defence of the need for external incentives and differentiation of employees:

No enduring or complex formal organization of any kind seems to have existed without differential material payments.... The same principle applies in principle and practice even more to non-material incentives. The hierarchy of positions, with gradation of honors and privileges, which is the universal accompaniment of all complex organization, is essential to the adjustment of non-material incentives to induce the services of the most able individuals or the most valuable potential contributors to organization, and it is likewise necessary to the maintenance of pride in organization.

Lord Wilfred Brown in his classic "Exploration in Management" devotes a chapter to the relationship between "the manager and his subordinates" (Brown, 1960: 72-88). In this chapter, Brown directly relates to issues in employee PA:

It is not difficult for a manager to know, or to find out, whether or not one of his subordinates has carried out the prescribed component of his job... It is much more difficult when we come to the discretionary component... This seems to me to be the really difficult part of a manager's job. He is called upon to judge the executive work of one of his fellow human beings. I think everybody tends to shrink from such a task. The manager is faced with the fact that his decisions in these circumstances may have a considerable bearing on the future career of his subordinate and on the fortunes of his home and his family.

Regarding the basis of assessing subordinates, Brown states that:

The assumption is often made that a subordinate can be assessed on quantitative figures alone which indicate his performance... Unfortunately, the assumption is usually not valid; facts and figures are, of course, an aid to assessment, but they are seldom the whole story.

And Brown even touches specifically upon issues in evaluation of sub-ordinates behaviour:

The right of a manager to criticize work is clear; but the right and duty to make personal criticism of behaviour because it is such as to interfere with work by setting a bad example, disturbing relationships with others, introducing discord into the company, is not so clear.

Modern managers and researchers struggling with theoretical, practical or ethical issues within behavioural target setting and evaluation in PA, can either comfort themselves by, or become worried by, depending on perception, the fact that these issues have been well known for at least 50 years, without much progress being made.

Peter Blau (1955) in a case study in his "The Dynamics of Bureaucracy" examined the use of quantitative measurements for performance evaluation in a state employment agency in the US, and identified what he saw as dysfunctional consequences of the same:

Statistical records facilitated the job of the superior by providing him with information that he would have had to ascertain laboriously otherwise and by improving his relationships with subordinates. It might be expected, therefore, that superiors favoured the exclusive use of quantitative indices for evaluating subordinates, a method of evaluation that would eliminate personal considerations in accordance with bureaucratic principles. This, however, was not the case. Superordinates on all levels explained that it would be impossible to judge all aspects of performance on the basis of these indices alone; 'You can't reduce a man to statistic'. The validity of this opinion can hardly be questioned.

Blau identified a number of dysfunctional effects of the introduction of statistics and quantitative principles in performance evaluation. First, it undermined the managers' authority because his judgment was no longer enough if he could not support it with quantitative evidence. Second, the relation between statistics and an individual's performance was rarely straight-forward, so the translation of statistics became a source of conflict in itself ("figures can't lie, but liars can figure"). Third, documented performance goals were displacing real value-adding goals as the prime focus of managers and employees. Or, in Blau's terms, there was a "displacement of goals whereby an instrumental value becomes a terminal value". Fourth, the thing that was most annoying to one employee who worked as an interviewer in the employment agency, was that the quantification of employee productivity created:

Competition between interviewers to an extent that is - disgusting... they lead to competition and to outright falsification.

Modern PA schemes have close ties to the Management by Objectives (MBO) wave which emerged back in the 1950's (Drucker, 2007). Although MBO later grew out of fashion, the PA process survived and has become so institutionalised that it is today seen as a mere fact of organisational life - something that organisations have to live with and therefore something they may as well learn to live better with. Today, most large corporations have implemented and maintain standardised PA processes.⁴

Based on the feedback from people I interviewed, the institutionalisation seem to best be described in anthropological terms as a set of practices and rituals that define what it means to be part of social entities like modern organisations (Douglas, 1986). Or in broader sociological terms as mimetic isomorphism (Powell & DiMaggio, 1983), i.e. as a process of diffusion and isomorphism through a variety of channels like employee movements, standardised advice from consulting firms, business media discourse, higher education curricula etc. to a degree where there is an almost complete lack of fundamental cognitive reflection on why PA is there in the first place.

This entails that for managers and employees PA has become a basic element in their perception of what is natural in organisational life. It is not the objective of this study to uncover the dynamics of how this diffusion has happened over time, although that would certainly be an interesting study. However, given the lack of reflection among managers and employees on the core question of why PA is there in the first place, it appears reasonable to conclude that mimetic isomorphism has stronger explanatory value than more rational decision making based institutional diffusion theories would claim, e.g. what Powell and DiMaggio (1983) called coercive pressures for isomorphism, where top management seek increased external legitimacy by fulfilling certain standards set by important stakeholders or authorities.

PA is a theme that now plays a role in the work lives of most people in large organisations in the Western world. It has also been the object of broad scholarly interest over time. My database search identified more than a thousand academic journal articles from the twenty year period 1990-2010, where either title, keywords or abstract contained the word combination "Performance Appraisal". Today, a myriad of consulting firm offerings, standard IT solutions, benchmarking institutions, books,

⁴ Surveys indicate that more than three out of four US business organisations have implemented a formal PA process. (Coens & Jenkins, 2002:35).

training courses and "best practice" schemes are offered to organisations to help them implement, re-implement or improve PA schemes. As such, the concept of PA is not just institutionalised in terms of its dissemination into large organisations throughout the Western world; it has also become big business in itself in which organisations invest heavily, both in terms of costs to advisory services and IT but also in terms of spending significant organisational resources on maintaining and fulfilling the schemes.

The spread of PA in modern organisations, the institutionalisation of its design and the significant resources spent on it make it important for both scholars and practitioners to understand what problems there are in PA, if and how they can be solved and more generally how PA affects organisations and individuals within them. The present study focuses on how managers and employees across four workforces create and maintain meaning when they are faced with problems in PA.

1.3 The Research Question and Definitions

In the following chapter I will in detail argue for the existence of an important gap in the literature. It is the aim of the present study to contribute to filling that gap in the literature by exploring local adaption and meaning creation processes in PA. Thus, I will examine patterns of how PA is locally adapted and how managers and employees across workforces establish meaning in situations where they are faced with problems in PA that challenge their existing meaning schemes. My research question is therefore:

How is managers' and employees' meaning creation affected by problems in PA and what patterns of local adaption and meaning creation can be observed across organisational workforces?

By studying local adaption and meaning creation processes I will add to the performance management literature and the on-going conversation about improvement or abandonment of PA. My contribution to the literature will be to extend the understanding of how problems in PA are perceived and dealt with differently across workforces within a corporate one-size-fits-all PA scheme, dependent on the particularities of the task at hand in each workforce. Further, I will add to the literature by showing how these patterns of meaning creation and local

adaption are causing different management approaches to problem resolution, thereby re-enforcing existing problems or articulating new problems in PA. This is what I have called the meaning creation feedback model.

Before proceeding with the presentation of the PA conversation in the literature, a few definitions are appropriate.

By *meaning* and *meaning scheme*, I refer to the establishment of "*plausible images that rationalise what people are doing*" (Sutcliffe and Obstfeld, 2005). Meaning creation is an activity heavily affected by the social and institutional setting in which it takes place. It is important to stress that meaning is not similar to the explication of rational choices and actions. Rather it is the other way around; meaning is constituted by narratives that allow actions and choices to be perceived as rational in retrospect. As Sutcliffe & Weick (2009: 79) put it: "when action is the central focus, interpretation, not choice, is the core phenomenon."

By *problem in PA* I refer to what the academic literature over time has identified as challenges to effectively and efficiently achieving improved individual and organisational performance by utilising PA schemes. In the literature review more than 50 such problems were identified, as will be made clear below.

By *meaning establishment* I refer to the process in which individuals create new interpretations of their relation to the PA process as a consequence of challenges to their current meaning schemes. I deliberately prefer the term "establish" to "re-establish" because although meaning establishment through sensemaking processes are activated by a perceived loss of meaning, it is not so that individuals are necessarily merely reshaping previous meaning interpretations when developing new stories that are meaningful. In fact it became evident through my study that sometimes stories can change quite dramatically when problems in PA challenge individuals' to reformulate what is going on in PA.

By individual *performance* I mean the perceived output or behaviour of an individual relative to certain expectations that can be more or less clear or ambiguous and that can be defined differently by different actors.

By *behaviour* I mean the way a person acts, reacts and interacts with others while performing. Or, in more practical terms, behaviour is the 'how', in the process of performing. Employees may achieve business targets in various ways through

different behaviours, where some are more positively perceived by management than others. In psychology, one often finds a separation between attitude and behaviour, where the former is the inner cognitive pattern while the latter is the outer, observable expression of the same. It is the latter that is of interest in this context, because it is the observable behaviour that is being evaluated and appraised in PA.

By *performance event* I mean any event that serves or potentially could have served as input to performance evaluation (e.g. employees' contribution to a project, employees' contribution to maintaining routine tasks, employee's contribution to teamwork, employees' general behaviour or behaviour in particular situations, etc.). As such, the term is more complex than one may think because it includes the distinction between what becomes "real" performance events (included in evaluation) and what remains as "potential" performance events (not included in evaluation).

By *feedback event* I mean any formal meeting event that is an element in the PA process (i.e. target setting meeting, midyear review meeting, calibration meeting, final appraisal meeting etc.) or other less formal performance feedback meetings between managers and employees.

2 The PA Conversation - a Literature Review

Before digging into my empirical study of PA, a logical initial step is to get acquainted with the positions among scholars to find out how the subject has been handled hitherto and to position the study in its academic conversational context.

I will start my literature review from the helicopter perspective, by demonstrating how the fundamental attitude of scholars towards PA can be used as a way to create some order in a methodologically and theoretically quite heterogeneous field. I will do this by showing how attitudes towards PA are implicitly or explicitly apparent, both in mainstream management monographs and in literature explicitly focused on the subject.

After this introduction, I will present a detailed literature review of 125 journal articles from the 20-year period 1990-2010. First, I will introduce my approach to literature selection and consider some of the advantages and downsides of this approach. Second, I will analyse the empirical focus of the articles by investigating the choice of empirical object, to identify if there are important gaps that I could aim to fill. Third, I will analyse the scientific methodologies utilised in the articles, to understand how the authors have tried to approach the empirical world. Again, the ambition is to identify possible patterns and important gaps that would be beneficial to fill. Fourth, I will analyse the theoretical approach of the articles to understand from what position the authors have attempted to create new generalisable knowledge through their studies. Fifth, I will take a closer look at articles affiliated with the radical perspective on PA. Sixth, under the heading 'what are the problems with PA' I will summarise the problems identified and analysed in the journal articles covered by the literature review. Finally, I will summarise my key findings and formulate the empirical, methodological and theoretical gaps that my study aims to fill.

2.1 The Performance Appraisal Attitude Dichotomy

In the literature, two different normative paradigms can be identified. I will call these the dominant and the radical paradigms. Each paradigm is defined by their fundamental attitude towards and assumptions about PA as a relevant strategic management tool.

The dominant paradigm

The dominant paradigm is characterized by acceptance of PA as a given, institutionalised element in organisational life. At its extreme end this paradigm holds a wide range of authors from consultants with focus on implementation and improvement of processes and systems (Bacal, 1999 & 2004. Grote, 2005) to academic scholars, with focus on improving the process by suggesting alternative tools and concepts (Armstrong & Baron, 2005. Armstrong 2006. Lazear and Gibbs, 2009).

In mainstream economic literature the process is sometimes described as a strategic tool for realising the organisation's strategic ambitions (Kaplan & Norton, 2001). Many mainstream economics textbooks treat PA as a necessary and interrelated phenomenon to incentive management through rewards, most often in financial or career terms. In this context it is routinely taken for granted that

The evidence is clear: Employees tend to respond strongly to incentives. This means that if an incentive plan is designed well, it can be an important source of value creation...pay for performance and other forms of extrinsic rewards are the most important motivational levers that a manager can pull. (Lazear and Gibbs, 2009: 231-32)

As we have already seen in the introduction, in its original form PA has close family ties to the Management by Objectives (MBO) wave which emerged back in the 1950's. Today, most large corporations have implemented and maintain standardised PA processes.⁵

In his analysis of the role of the HR function Ulrich (1997, 2005) contends that PA only plays a subordinate part. Still, it is interesting to see that it is not only taken for granted but also placed in the day-to-day/operational quadrant in Ulrich and Brockbank's (2005: 24) multiple-role model for human resources. This operational placement is interesting since much of the critique against PA is precisely directed against its retrospective, operational focus which is in sharp contrast to the more strategic arguments often put forward for having the process in the first place.

Ulrich and Brockbank (2005) explicitly state the HR delivery from the quadrant to which PA belongs, to be "increased employee commitment and competence". This should be held against their general position that HR has to be maintained in a way

⁵ Surveys indicate that 75-89% of all US business organisations have implemented a formal PA process. Coens & Jenkins (2002: 35)

that balances the benefits of “free agency and control” (Ulrich and Brockbank, 2005: 46). Ulrich and Brockbank refer to McKinsey’s six factors that define how organisations work and place PA as a “consequence” factor, describing how the appraisals should be used as the basis for distributing rewards (Ulrich and Brockbank, 2005: 68). These reflections also are at the core of the academic debate around PA.⁶ Does the process deliver employee commitment and competence or is it ridden with flaws that are counterproductive to this objective? Is appraisal rating a fair foundation for distributing rewards or is it subjective and even arbitrary? Ulrich and Brockbank have another focus and unfortunately do not consider these questions even though they deal with many of the themes that have been criticised by opponents of PA: individual priority setting, control, commitment, collaboration, culture, competence development, etc.

By this, Ulrich and Brockbank take the appraisal process for granted as most other mainstream scholars and HR researchers and also company employees and managers do (Armstrong & Baron 2005. Armstrong 2006. Bacal, 1999). There are huge numbers of both academic and more practical “how-to” publications that deal with PA from this non-critical perspective.

Armstrong (2006) is one of the scholars that have published research into the field of how to implement and optimize the design and maintenance of performance management including appraisal within organisations.

Armstrong gives a thorough presentation of the generic model starting with the PA process and its alleged 12 sub-processes, from “performance planning” through “assessing” and “rating” to “documentation” (Armstrong and Baron, 2005: 23). His research not only documents the current institutionalisation of the process, i.e. the almost universal usage of the model in the US and the uniform vocabulary that has been established around it. He also suggests how the process can be optimized by adding new elements, e.g. 360 degree feedback, team level evaluations etc.

Besides academic contributions, within the dominant paradigm a multitude of non-academic studies are found, much driven by the vast consulting business that has grown around implementation and adjusting the PA process in organisations. PA has

⁶ For an introduction to the academic debate on linking PA to financial reward, see Armstrong & Baron (2005)

become big business due to its universal deployment and the many pitfalls and issues that organisations experience with it.

Bacal (1999) is a good example of this. This author tells much the same story as the state-of-the-art academics. But coming from the consulting business and having a much more practical approach, he presents the challenges that organisations need to overcome and subsequently why they need consulting to help them do this. Thousands of books, manuals, training courses and consulting “offerings” do the same. A search on Google for “Performance appraisal” gives 16.5 million hits.⁷

Although MBO grew out of fashion in the 1990’s, it is not surprising that the consulting business saw an interest in keeping the revenue-generating advising within the field alive. Maybe it is more surprising that it was not fundamentally questioned by the broader academic community. This probably has to be assigned to the fact that the PA process has become so institutionalised that it is seen as a mere fact of organisational life - something that organisations have to live with and therefore something they may as well learn to live better with. HR scholars like Ulrich, PA specialists like Armstrong and consultancy advisors like Bacal may point to weak spots and issues within the practice of PA but they do not fundamentally question the PA paradigm itself. This is what constitutes the dominant paradigm.

The radical paradigm

In opposition to the dominant paradigm an opposing position has emerged since the 1980’s. I call this the radical paradigm to underline the fact that this academic paradigm has emerged from a fundamental rejection of the dominant paradigm. It is radical because it not only rejects the paradigmatic implementation doctrine but also the institution of appraisal itself.

Coens and Jenkins (2002) argue in their book "Abolishing Performance Appraisals" that organisations should abolish PA right away even if they do not have anything to put in its place. The argument is that the costs are so huge and the benefits so few that organisations will improve their performance simply by eliminating the process. The flaws of PA schemes are so deeply connected to the foundation of the process

⁷ Search conducted June 2010

itself that although it is possible to improve and adjust the process, it is not possible to make it become a positive asset to the organisation.

In Coens and Jenkins' analysis, PA schemes generally fail to deliver what they aim for. Coens and Jenkins quote a number of studies to show how organisations working with PA schemes are generally disappointed with the outcome of the efforts. Coens and Jenkins also quote an *Industry Week* survey showing that only 18% of the respondents considered their own appraisal reviews effective while 48% considered their review "second-guessing sessions" (Coens and Jenkins, 2002: 18). More so, Coens and Jenkins claim that it is simply impossible to find any trustworthy academic study that proves a positive outcome of conducting systematic PA in an organisation. Often, consultants and proponents of PA schemes emphasise the fact that most of the very successful companies in the world are using PA schemes. But it would not be difficult to find highly unsuccessful companies also using PA schemes, so this cannot in itself be taken as any kind of proof. For example, huge American companies like Enron and Arthur Andersen were using PA schemes and relied heavily on extrinsic rewards, but both failed to deliver sustainable organisational performance. Neither of those two companies exists today. For Enron, one commentator even directly blamed the PA system for playing a key role in the disaster:

HR ... failed to put "teeth" in its performance management systems that would severely punish (or fire) individuals that kept secrets, took excessive risks or that violated the company's values or ethics. As a result, HR inadvertently sent a message to employees and managers that results, regardless of how they are obtained, are all that matters. (Sullivan, 2002).

Maybe surprisingly to some, it is hard to find scientifically based evidence that supports the common perception that organisations using PA schemes are systematically more successful than organisations that do not. Let alone any evidence that if such relation even could be found, it would not be the cause of other parameters like organisation size, ability to attract talent, organisational culture or other possible explanatory factors.

In fact, DeNisi and Pritchard (2006) found that academic studies into the effectiveness of PA have significant methodological flaws:

... in 68 of the 70 studies there was no control group for comparison... In fact, they reported in their meta-analysis that, when proper controls were considered, feedback actually had the effect of decreasing subsequent performance in one-third of the

studies – and that these results were independent of the sign of the feedback received. Thus, the question of direct evidence for the effectiveness of this performance management intervention for improving performance must still await a more definitive answer.

Coens and Jenkins develop their argument around 10 proposed false assumptions behind modern PA schemes. Since these arguments form the core of the radical critique of modern PA schemes, I will present these at some length in the following.

First, appraisal schemes are assumed to be able to support multiple, often conflicting, objectives at the same time. For example, the objective to provide an honest dialogue about feedback and development on the one hand and the objective of evaluation for bonuses on the other hand. Second, the schemes are built on the assumption that one size fits all - that one uniform setup for dialogue and feedback will serve the needs of all managers and employees, when in fact people have very different needs and strengths that would often be better accommodated in more flexible setups. Third, it is assumed that it is possible to provide employee commitment from a top-down enforced process, when in fact commitment is more often thwarted by such a paternalistic process, with a following decrease in performance. Fourth, performance schemes build on the assumption that it is managers who are responsible for compiling feedback and driving performance, which is detrimental to a more empowering management style where employees themselves take responsibility for compiling feedback and the manager takes a more coaching role. Fifth, it is falsely assumed that it is possible to provide anything near objective, comparable evaluations of employee performance across large complex organisations. In reality, bias is everywhere and ratings will be highly inaccurate and based on incomplete information.⁸ Sixth, it is often assumed that PA schemes will be helpful when dismissals of employees are required. However, PA schemes are in practice detrimental to the organisation's interests in such cases, since most managers are shy of conflicts and thus tend to avoid giving low-end ratings. Such "too friendly" ratings can be ammunition for the employee as proof of unreasonable dismissal, rather than being useful for the organisation as proof of reasonable dismissal. Seventh, it is assumed that it is motivating to compete for high ratings. However, studies show that most humans see themselves as performing above

⁸ Throughout this thesis I will use the term 'bias' understood as a "cause to feel or show inclination or prejudice for or against someone or something" (Oxford Dictionary)

average, so a majority of the employees will in reality be disappointed by the rating they receive, and disappointment is demotivating rather than motivating. Eighth, PA schemes institutionalise in the organisation that feedback on performance and development is a 3-times-a-year standardised and formalised activity, when it would be much better to have a much more frequent, informal feedback culture in the organisation. Ninth, it is assumed that people's motivation is highly dependent on extrinsic factors like performance ratings and bonus, while Coens and Jenkins claim that stimulating intrinsic motivation is a much stronger and lasting way to achieve employee commitment and performance. And lastly, it is assumed that the organisation can be effectively improved by focusing on inspection of the individuals, while it would have much higher impact to look at the systemic level of the organisation and focus on how organisation and processes can be improved.

Kohn (1999) draws similar conclusions in his book "Punished by Rewards - the trouble with Gold Stars, Incentive Plans, A's, Praise and other Bribes". Kohn takes his point of departure in research into children's learning in school and through their upbringing. From this starting point, Kohn also studies the use of rewards in organisations, based on formal appraisal processes. His theoretical perspective is socio-psychological and Kohn sees the appraisal process as based on an old-fashioned, mechanistic view on humans, built on "our reliance on a theory of motivation derived from laboratory animals". This is with reference to the fact that Kohn sees PA to be a legacy from Skinner's (1974) behaviourism. Behaviourism was focusing on how it is possible to affect human action in an intended direction by rewarding wanted behaviour and punishing unwanted behaviour. Kohn simply rejects the idea that it is possible to motivate people through a "do this and you will get that" doctrine (Kohn, 1999: 3). Whereas many modern managers today have less belief in punishment as a tool, it is much more common to see rewards of many sorts, from praise to bonus, as an important way for management to create motivation among employees. In his interesting study Kohn asks two fundamental questions; Is it right to reward? And is it effective to reward? Kohn finds the answer to both questions to be clear nos.

It is quite clear that modern PA schemes are built around a reward and punishment architecture. To put it in Kohn's terms, goal setting is the "do this". The rating and the associated consequences for bonus, promotion, dismissal etc. is the "get that".

Kohn's critique of appraisal systems and other reward systems is a close kin to a much older discussion among motivation theorists and work psychologists, namely the discussion of the effects of intrinsic and extrinsic motivation, and how the latter affects the former. As such, together with Coens and Jenkins' critique of the false assumptions behind PA, Kohn's critique of reward-based systems fulfils the picture of how the radical paradigm builds its argument against modern PA schemes.

So what is the problem with rewards? First, Kohn argues that rewards are in fact just one side of the coin where the other is punishment. Rewards are punishing in the sense that they are controlling by nature. Rewards direct humans to behave in certain ways that may be detrimental to who they feel they really are or how they feel they can best realise their potential, and this control aspect has in itself punitive character. But even more important, rewards will not be obtained by everyone to the degree hoped for and the effect of this on the individual will be similar to the effect of punishment. The individual will experience that something he deserved has not been granted to him. The more desirable the reward is the heavier the employee will feel punished by not getting it, and the more negative impact will it have on the employee's work morale. Kohn finds that unexpected rewards are for this reason in fact much less harmful to employee morale, but

...apart from the practical problems of trying to keep people from expecting another reward tomorrow, it is no coincidence that the great majority of rewards are promised in advance. The whole point is to control people's behavior. (Kohn, 1999: 53)

Second, Kohn claims that reward schemes are destroying relationships, both horizontally between colleagues and vertically between managers and employees. Often rewards are provided based on relative evaluation of peers. As in my case study, many organisations create a ranking scale that introduces an artificial scarcity of high rankings that then can be distributed among peers, with high-end rankings being most rewarding. The scarcity of high-end ratings and the relativity will by nature create competition between peers as they will not only see each other as rivals, in fact they will be installed as rivals. According to Kohn, competition among colleagues creates anxiety that typically interferes with performance. But even worse, the majority of the employees will soon realise that they will realistically not be candidates for high rankings and will thus be discouraged from making an extra effort. Can this be dealt with through team rewards? According to Kohn it definitely cannot. Where most people are against collective punishment, collective reward is

just another variant of the same - exactly as it is at the individual level. Only, at the collective level rewards install even stronger psychological pressures on the individual in the form of group pressure and exposure of weaker individuals as scapegoats for the group's failure to achieve a reward. So, rewards are counterproductive when it comes to the ability to realise a workplace characterised by trust and open communication.

Third, Kohn sees rewards as generally ignoring reasons. It is very easy and convenient for the manager to be able to tick off if a target in an employee's performance plan has been met or not. But in today's complex workplace the realisation of a target very often is a result of multiple factors among which the employee's contribution is only one - and often only a minor one:

Rewards are... gimmicks, shortcuts, quick fixes that mask problems and ignore reasons. They never look below the surface... A sudden deterioration in performance frequently turns out to be due to problems at home. A chronic record of mediocre performance, meanwhile, may indicate, among other possibilities, that there is something wrong with the job itself... Turning the workplace into a game show [] does exactly nothing to solve these underlying problems and bring about meaningful change. (Kohn, 1999: 60-61)

Fourth, Kohn states that rewards generally discourage risk-taking. Rewards narrow people's focus so that anything not immediately instrumental in achieving the reward will be ignored. Rewards also have the effect that people will do exactly what is necessary to get the reward, but nothing more. The purpose of working is not to be creative or innovative. The purpose of working is to "get the goodie". In this way, rewards encourage people to focus on obtaining the reward with as little effort as possible, instead of going into a task solution with the purpose of exploring options of getting the best possible outcome:

If you have been promised a reward, you come to see the task as something that stands between you and it. The easier the job is, the faster you can be done with it and pick up your prize... Do rewards motivate people? Absolutely. They motivate people to get rewards. (Kohn, 1999: 65-66)

Lastly, Kohn finds that extrinsic rewards are not supplements to intrinsic motivational factors; they are directly detrimental to them. Kohn draws on a large number of psychological experimental studies on motivation that show how extrinsic rewards reduce intrinsic motivation. Reward is like a drug that makes employees addictive and destroys their own intrinsically founded commitment to a task. In fact, Kohn believes the effect is so strong that

Rewarding people might even be regarded as a clever strategy for deliberately undermining interest in something. (Kohn, 1999: 72)

Of course, rewards provided as ratings through PA schemes are no better. Kohn finds that they have the same harmful effects on employee motivation as any other reward scheme.⁹ Through his empirical studies, Kohn claims to demonstrate that not only do rewards not work in the workplace, they also do not work in the classroom and neither do they work to help children become good people. The only thing that really works when it comes to creating engagement, commitment and creativity, according to Kohn, is to create an environment that supports the blossoming of intrinsic motivation.

Research from different positions within the paradigms

As my literature review of journal articles on PA will show, only a small minority of today's scholars seem to subscribe to the radical paradigm. However, it would be wrong to assert that scholars uncritically position themselves at the opposite end of the attitude dichotomy. Rather, the vast majority of scholars study aspects of and issues with PA schemes, including some of the themes brought forward by Kohn and Coens & Jenkins. They position themselves within the dominant paradigm because they do not fundamentally challenge the legitimacy of appraisal schemes in the first place. They do not agree or even consider the view of the radicals, that it is plainly impossible to achieve a net benefit from conducting PA schemes, no matter how much effort one puts into designing a system that accounts for all the known biases and flaws associated with such schemes. It is fair to say that the radical scholars I have identified are basically concerned with many of the same issues as the more mainstream scholars. The dividing line is whether a scholar believes or does not believe that it is possible to overcome the issues or whether the issues are symptoms of such fundamental aspects of human nature, that PA schemes in any form are doomed to failure. The issues at stake are the same, but the conclusion is more radical.

Thus, there is a large body of scholars with varying degrees of scepticism regarding the possibility of harvesting significant net benefits from the conduct of PA, but who

⁹ For example, Kohn (1993: 42) refers to behavioural research as far back as from the 1960's which documents that financial reward is not only questionable but in fact directly negative in its impact on human performance.

are not concerned with fundamentally questioning the legitimacy of the process itself. Rather, they are critically investigating challenges with the theoretical concepts or empirical practice of PA. As we shall see below, their critique takes its departure in many different theoretical perspectives, for example justice theory (Narcisse and Harcourt, 2008. Roch et al., 2007), motivation theory (Buchner, 2007. Kuvaas, 2006), organisational politics (Salimäki and Jämsen, 2010), social cognitive theory (Wood and Marshall, 2008), social learning theory (Tziner et al., 2002), Total Quality Management conceptual thinking (Soltani et al., 2004), discourse analysis (St-Onge et al., 2009) and more. For many authors, optimising the value provided to an organisation by conducting PA schemes is very much about handling rating dysfunctions to secure a fair and transparent rating process. Before turning to my systematic literature review of 125 journal articles on PA, let us take a brief look at some of the most important dysfunctions that has been uncovered through research into the rating process.

In his book "Measuring and managing performance in Organizations" Austin (1996) tries to identify which organisational factors allow measurement to work successfully, and which force measurement programs to fail or do damage. Austin dedicates a chapter to what he calls measurement dysfunction, by which he means that measurement in itself becomes the cause of effects that are the opposite of what is valued. According to Austin (1996: 21) a key differentiator is whether measurements are motivational (intended to affect people being measured) or informational (used to build a deeper understanding of what is going on in an organisation). PA is in theory occupied with the former, since target establishment and evaluation is installed to motivate employees to work hard and dedicated towards target realisation. However, a problem is that informational measurement systems are often used to investigate to what degree employees realise the motivational targets (e.g. sales figures, production output, number of customer complaints, etc.). And this gives employees an incentive to try to affect the informational measurement systems so that they operate in favour of proving employees' target fulfilment rather than in favour of showing what "really" goes on. In other words, the instalment of motivational incentives that are to be measured by "objective" informational measurement systems has as a consequence that employees get an incentive to manipulate the measurement systems and thereby undermine the trustworthiness of the output

generated by the measurements. This can be done in many ways, e.g. when a sales person strategically places an order on the right side of a calendar date, to optimise his own reward scheme. Or when customer complaints are not logged or documented, because this would be detrimental to an employees' target realisation.

According to Austin, another key issue is that measurement systems have a built-in tendency to look for measureable indicators. So what is measured is rarely ever the real value of an activity, but an indicator of value, because it is always easier to measure quantity than quality. It is easy to measure if a project delivers on time, if a certain number of cases have been closed or if targets for average turnaround time on case handling have been met. But did the project deliver the value that was anticipated? Or did the agent handle all the easy cases to reach the target, leaving behind the more difficult but maybe more important cases? The solution installed to handle this issue is often even more refined targets that again persistently run into new issues because they again are targets on indicators of value, not of added value in itself. More sophisticated evaluation systems only serve to give more sophisticated dysfunctional reactions:

Employees' true output (such as value to the organization) is often intangible and difficult to measure; in its place, organizations choose to measure inputs (such as the amount of effort devoted to the task as measured by counts of interviews performed). (Austin, 1996: 18)

Austin sees a strong link between McGregor's (1960) theory X style of management and many performance measurement schemes, because they rely so heavily on extrinsic motivation and share the assumption that humans will attempt to avoid work if they are not coerced and controlled.

Austin introduces the concept of measurement cost as something any organisation should consider before installing measurement activities on any working process. Some of the factors affecting measurement costs are the repetitiveness of the task (the more repetitive, the easier to measure), the complexity of the task (the less complex, the easier to measure), newness of the task (the more well-established the easier to measure), the level of specialised knowledge required by the task (the lower level required the easier to measure), interdependence and separability of effort (the less interdependent and the more separability of effort, the easier to measure) and finally also any environmental covariate and other noise (the less of this, the easier to measure). It is not difficult to follow how Austin concludes, that in many modern

workplaces dominated by highly educated knowledge workers, all these factors will turn out to be in disfavour of measurement. However, measurement schemes are everywhere in such organisations.

As the objectivity of employee rating is a cornerstone in almost any PA scheme, other scholars have focused at the subjectivity caused by human biases that are unavoidably present in any human relationship. Some examples of these biases are: managers tend to categorise according to their own prejudice and stereotypes; managers favour people they personally like; managers intentionally or unintentionally discriminate against race, gender or age; managers are shy of conflicts and tend to be too lenient when rating; manager's rating of employees is affected by the rating the manager has received herself; managers tend to cluster their ratings in the middle of the rating scale, not using the extreme ends; managers are affected by the halo effect, causing them to carry perceptions of an employee's performance in one area into perceptions of the employee's performance in other, separate areas; managers (and employees) focus on recent performance, leaving out performance activities not conducted within the last few months up to the PA interview; and finally, managers and employees are prone to attribution errors, so that favourable outcomes are attributed to oneself while unfavourable outcomes are attributed to external factors or someone else (Coens & Jenkins, 2002. Furnham, 2005. Armstrong, 2008. Austin, 1996).

Other examples of authors who are critical towards PA without abandoning it altogether are e.g. Amabile (2011) and Frey (1997). From a motivation theory perspective Amabile focuses in her article "Motivation" on particular issues with extrinsic reward schemes in relation to innovation and creativity. Through a typology of personality traits related to "highly creative people" containing elements like e.g. spontaneity, curiosity, openness to change and experience, flexibility, etc. she concludes that creative people are mostly driven by intrinsic motivation while extrinsic motivational pressure often will have a detrimental effect on creativity.

Frey (1997) has in his book "Not just for the Money" investigated the effect of extrinsic reward on intrinsic motivation and concludes that under certain conditions there is a clear crowding-out effect in the sense that higher extrinsic reward is detrimental to inner motivation:

To offer higher pay then makes people less committed to their work, and may reduce their performance (Frey, 1997: ix)

Here, the conclusion is the same as Amabile's, although for a broader scope, since Frey is not particularly focusing on highly creative people, but more generally on the effect of extrinsic motivation schemes on human behaviour. Frey's argument is particularly interesting because it takes its point of departure in an economic approach like so many PA text books within the pro-PA dominant paradigm. He does not build his argument on qualitative sociological case studies or non-empirical arguments like some proponents of the radical paradigm tend to do (Bowen, 1995. Fisher, 1994. Townley, 1999). Rather, he investigates the crowding-out effect on the home turf of quantitative economics; through controlled experiments, statistical analysis and mathematical theory building. Through his analysis, he identifies a number of important conditions for the crowding-out effect, including intervention in the shape of "monetary reward" or "regulations". In both cases, the condition under which an employee will experience the crowding-out effect is if the monetary reward or the regulation is perceived as controlling. And as we shall see, PA schemes utilise both elements by installing regulations in the form of targets defining the employees focus, deadlines, approach and priorities, while promising reward, if the employee delivers within the frames of these targets. In short, Frey concludes that a rational agent increases his performance until the marginal benefits equal the marginal cost of so doing. The agent's optimal performance is affected when the principal changes the extent of external intervention. So the art of intervention for management is to identify the optimal level of intervention so that the crowding-out effect is not larger than the positive effect of the intervention.

Again, all the above examples of dysfunctions and biases can be investigated with or without accepting or rejecting PA schemes as meaningful tools altogether. As already indicated, most scholars study PA from a critical position to improve or even reform the PA process. But they do this without fundamentally questioning whether it is at all possible to gain a positive net outcome from the process. Table 2-1 below summarises the key differences between the dominant and the radical paradigms.

2. The PA Conversation - a Literature Review

Paradigm characteristics	Dominant paradigm	Radical paradigm
Fundamental attitude	PA will add value to the organisation if used the right way	PA is harmful to the organisation
Fundamental consequence	Work with the process design and practice to improve or even reform it	Abolish PA
Management ideal	Top-down. The PA process is key for the leadership's strategic management and control	Bottom-up. Top-down processes are paternalistic. Managers should see themselves as coaches, not judges
Organisational improvement perspective	Individual. Organisational performance is the sum of the individuals' contribution	Collective. Focus on system and processes is much more effective to improve organisational performance, than focus on the individual
Assumption regarding change	The process can be improved and reformed to increase the value to the organisation	The process can be improved and reformed, but it will still be harmful to the organisation
Assumption regarding competition	Competition is motivating because people want to be winners	Competition destroys motivation because basic psychology makes people overestimate their own contribution and therefore relative ratings will be disappointing to most
Assumption regarding feedback	Responsibility of the manager (push)	Responsibility of the employee (pull)
Assumption regarding motivation	Extrinsic motivation works	Only intrinsic motivation has a lasting positive effect
Assumption regarding relations	PA schemes can be an effective tool for managers to install ownership in others	PA schemes are harmful to the manager-employee and the employee-employee relationships

Table 2-1: Key distinctions between the dominant and the radical paradigms

In Table 2-1 I have summarised the key distinctions between the two paradigms as they have been unfolded above. Although I have presented the distinction as a dichotomy, it should by now be clear that there are of course many scholars who work within the dominant paradigm, who share some of the concerns of the radicals. The deciding separating factor in the dichotomy is therefore what I have summarised in the first two rows in Table 2-1: The fundamental attitude and recommended consequence. There are sceptics, improvers and reformers within the dominant paradigm. The revolutionaries are to be found in the radical paradigm.

The idea about a PA attitude dichotomy grew as I read through a range of key monographs on PA in the early phases of the study. However, I have obviously not read everything ever written on the subject of PA. So some methodological reflexion on the authority and representation (Alvesson and Sköldböck, 2009: 11) of this part of my study may be appropriate, just as I will later argue in detail for the selection of journal articles in my literature review. One may rightfully ask on the basis of what criteria the monographs were selected. The answer to this question is that I started

out in the dominant paradigm, moved on to literature from the radical paradigm and then looked for literature that could add nuances to the dichotomy.

First, I read mainstream management and economic textbooks that dealt with the subject as one among many other subjects, e.g. Lazear and Gibbs (2009), Kaplan and Norton (2001) and Ulrich (1997, 2005). In these I found references to monographs more directly focusing on PA schemes that I then turned to. These included e.g. Kessler (1995), Grote (2005), Bacal (2004, 1999) and Armstrong (2005, 2006). While reading these authors, I became more interested in why books on PA are often so focused on issues and challenges if the concept is as valuable as it is often portrayed. It seemed to me that they all had high hopes for the process but were somehow disappointed with aspects of the process that were apparently difficult to get to work as intended. From the reference lists of these books I started to search for monographs that take a more radical approach to the theme. I started out with Coens and Jenkins (2002) who referenced Kohn (1999). The radicalism of these two monographs was what really formed my idea of a dichotomy.

However, I was still not sure if it was in fact a dichotomy or a continuum. To investigate if there were really fundamentally different assumptions at play, or rather varying degrees of critique, I studied the references of Kohn's and Coens & Jenkins' radical monographs to find the origins of what they were so heavily against. This led me to monographs like Skinner (1976) and Drucker (2007). At this stage I became convinced that there were in fact fundamental differences at play. Therefore, the last step in my monograph search process was to start to look for books that were as close to the border between the two paradigms as possible, because the border often is where the most interesting reflexions are found regarding the relation between two opposing positions. This led me to more narrowly focused monographs covering one or a few aspects of PA in a critical way, e.g. Austin (1996) on the problem of measurement, Vroom (1964, 1989) and Herzberg et al. (1993) regarding motivation and Furnham (2005) on work psychology.

After this helicopter view introduction to the state of the PA research and its affiliation with the proposed attitude dichotomy, let us now turn to the detailed literature review of journal articles on PA over the last 20 years.

2.2 Introducing the Literature Review

I have conducted a meta-study of the past 20 years research literature on PA published in academic journal articles. Considering the widespread proliferation of a seemingly standardised "best practice" approach to PA in the practitioner field, I was somewhat surprised to see how heterogeneous the academic field at first sight seemed to be. But looking closer at the material, I realised that not only could many of the articles be attributed to various established theoretical schools, whether or not this was explicitly or implicitly stated in the articles, they could also be sliced and diced on other interesting parameters that seemed to affect their view and analysis of the subject. Thus, the empirically based articles would have a specific cultural or geographical focus and would normally focus on one or a few specific workforces within either the public or the private sector. Methodologically, the articles differed in terms of empirical approach, varying from literature reviews on PA sub-themes over large scale quantitative survey based analysis through to deep-dive qualitative interview and case studies and even including non-empirical¹⁰ or historical approaches to the field. Lastly, most of the articles were expressing an underlying attitude towards the PA endeavour itself. The articles fit nicely into my proposed dichotomy of attitudes ranging from a position of strong believers who were concerned with how to get the most out of an approach that it was unnecessary to fundamentally question, to a position of outright rejection of PA as being nothing but a source of misery to the organisation and its people.

I approached the literature review in a structured manner and coded the articles according to a range of parameters that I found helpful for my own getting to grips with the subject. This not only helped me to get order out of what at first seemed a rather chaotic state of affairs. It also directed me to other relevant literature, particularly important monographs covering not only specifically PA, but also important sub-elements ranging from themes of work psychology to literature on return on investment and measurability within HR.

Below, I will present the outcome of this literature review. After a short introduction to my overall literature review approach, I will present my findings in five parts. First, I

¹⁰ The term 'non-empirical' is here defined as a study that does not involve generating new empirical data. According to this definition a literature review like the present chapter would be categorised as non-empirical.

will look into the empirical focus of the articles. Second, I will look at the methodological approach used by the authors. Third, I will look closer at the theoretical perspectives of the articles. Fourth, I will look closer at the articles positioned in the radical perspective.

A literature review focusing only at the parameter level of empirical, theoretical or methodological approach or implicit attitude can be interesting. However, of even more interest, of course, is what the authors actually are able to conclude about PA from that position. The author's key findings and contributions will therefore also be presented and unfolded along the way below to illustrate the points I make through the chapter. Before turning to summary and conclusions of the chapter I will fifth, give an overview of the problems identified and analysed in the articles as they have been presented through the chapter.

It will not be possible to give examples from all 125 articles. I have tried to let examples cover as broadly as possible by using many text references and examples all through the chapter. I have also maintained a principle of using only as an exception the same article as an example more than once. This way, I have managed below to reference 82 out of the 125 articles included in the review. One criterion for how I have selected these articles, and thus deselected 43, has been the articles' direct value as an illustration of a given analytical point in my review. I have simply looked for the best examples. When multiple texts have been possible examples of the same point I have either referenced them all, or in other cases simply selected the newest articles. Thus, more than 50% of the non-referenced articles were from 1991-1999, although less than 40% of the articles were from this period. Of course, all 125 articles were coded and included in the different thematic counts and overviews included in the chapter.

2.3 The Literature Review Approach

1.002 journal articles with either title, keywords or abstract containing the word combination "Performance Appraisal" were identified in a database search.¹¹ I surely could have found many more articles if I had enlarged the search to also cover

¹¹ The databases searched were <http://www.emeraldinsight.com/> and <http://www.ingentaconnect.com/>

related concepts like "performance management" or "employee evaluation". However, the purpose was not to find as many articles as possible, but to find a sample of articles of a reasonable and manageable size while still covering the theme of PA broadly. Out of the 1002 articles, 125 articles we selected for coding. The selection was based on the following relevance criteria: The articles should have an adult organisational focus (excluding articles on children upbringing, primary school or high school focus). The articles should have a human PA focus (excluding articles on accounting or financial performance management). The articles should be academic by nature (excluding non-academic articles in professional journals).

The 125 articles were coded on the following dimensions: publication year, source (name and issue of the journal in which the article was published), source type (the audience which the journal primarily addresses, typically either the HRM specific or general business management academic or professional communities, or the psychological academic or professional community). Further the articles were coded for their empirical focus, i.e. which workforce they concerned (managers, blue collar, white collar, R&D, HR), their empirical geographical focus (what country/countries were investigated?) and which sector was in focus (public or private). A detailed coding was made of each article's theoretical approach as well as of the general methodological approach (non-empirical, literature review, quantitative, qualitative, experimental, case and mixed-methods). In addition, each article was coded for its position on the attitude dichotomy (whether the author was positioned within the dominant or the radical paradigm). Finally, for each article selected keywords were also coded to allow for meta-level search and analysis. This coding was all stored in a database that allowed later sorting and analysis of the different parameters in any dimension or combination required.

As structured this approach is, it also has its shortcomings. Firstly, there is no doubt that important articles out there do not meet my search criteria of containing the word combination "Performance Appraisal" in title, summary or key words. Much research on subjects ranging from micro-level studies into employee motivation to macro-level studies of corporate strategy implementation, and a whole plethora of studies in between, could have direct or indirect interest to the study of PA without specifically focusing on this. However, since my aiming point is not to uncover every contribution

that could potentially add to the understanding of PA, but rather to create an overview of how PA is being dealt with in the literature, this limitation is not critical.

Second, a more fundamental critique could be raised against the fact that I have conducted my journal search based on two specific databases. Hence, a few words need to be added to explain this choice and its consequences.

Emerald was founded in 1967 as an offshoot of the University of Bradford. They claim to be "the world's leading scholarly publisher of journals and books in business and management". The database contains peer-reviewed articles from over 200 online academic journals. The database is widely used in academia with over 50.000 article downloads per day globally.

IngentaConnect on the other hand is not a publisher, but a portal that serves to make journals available for searches across a large spectrum of publications. IngentaConnect provides access to search across no less than 13.500 different publications but does not necessarily provide access to the single article itself without prior journal subscription or "pay per view". IngentaConnect provides access not only to academic journals and articles, but also to professional journals.

The validity of the database search was further verified by conducting a similar database search in Business Source Direct finding slightly less articles, 965, with a very significant article overlap between search results at the individual article level.

Thus, the two sources not only provide access to a very comprehensive body of relevant journals, they also supplement each other well in terms of coverage in depth and breadth. However, it is obvious that they are both founded in a Western academic context and they must be expected to have an overweight of Western journals rather than e.g. Asian journals. And although Western journals also take on articles from non-Western authors, as my review also will show, they are in the minority. Hence, when I later will conclude that the body of empirical research made into PA has a significant overweight in focus on white collar, Western populations, this may well be the result of what could be seen as a self-fulfilling prophecy. When one looks in Western journals, one is expected to find a bias towards Western empirical settings. On the other hand, although Asian research institutions are slowly catching up, suffice to say that management studies are still heavily dominated by

Western universities.¹² And since my own empirical object is located in a Western context, it will at least culturally fit to the main body of literature in my review.

Third, one could criticize the choices I have made for slicing and dicing the body of literature. Why focus on methodology? Why be interested in the theoretical perspective of the author? Why not have gender or age dimensions or some other often used independent variable? My aim was to contribute to the field of PA studies both methodologically and theoretically, to be able to maybe shed some new light on the phenomenon. When I started to read through the material and the must-read monographs about the subject, I found that much of the literature was dealing with the same problem of optimising a process that somehow never seemed to work optimally anyway. So, to be able to come up with something new, I thought that by getting a good overview of what had been done already and then do something different, would optimise my chances to succeed. I hope the reader will find that the following chapters will prove me right.

2.4 Analysing the Empirical Focus of the Literature

Before looking at the methodological approaches applied by the authors I will start by turning my attention to what empirical objects the authors are concerned with in the first place.

Public vs. Private sector focus

First, when the articles are coded for their empirical focus on public vs. private sector, we see that 50% of the empirically based articles are investigating private sector organisations, while 36% are investigating public sector organisations.¹³ The remaining 14% are investigating populations across both sectors.

In a global context this distribution seems to overweight public institutions significantly when compared to the public sector's relative share of the job market in most non-Western countries where welfare state institutions are less developed.

¹² E.g. on a list of the world's top 50 business schools compiled yearly by The Economist and Businessweek, all 50 schools are located in Europe and North America (2011).

¹³ In my coding, I have defined higher education institutions as public sector organisations, even though many universities, particularly in US and UK, are often dependent not only on public funds but also on private tuition fees and donations. Even in privately funded universities, for the employees, there are more similarities in culture and roles to other public universities than there are to more classical for-profit private organisations.

But even if the distribution fits better to a Western perspective where public sectors are bigger, public sector studies are still overrepresented in the articles. Public employment generally comprises 10-30% of the entire employment base in Western countries, with Scandinavian countries positioned in the top end of the interval (Handler, 2005). An un-weighted average of public employment share of the entire employment base across EU and US was approximately 16% over the period 1980-2002 (Handler, 2005). Since articles focusing exclusively on public sector make up 36% of all investigated empirical articles, and, as we will see below that European and US populations heavily dominate the empirical focus of the articles, it is reasonable to conclude that public sector studies are over-represented in the PA studies by roughly a factor of two.

One reason for this over representation could be that scholars have a preference for conducting empirical studies using populations from their own world: university students and staffs, as we shall see below is in fact the case for some types of studies. Another reason could be that researchers have better access to public sector organisations, although this is not possible to verify through my review. It is beyond the scope of this study to explore this imbalance in detail. However, this could be an interesting thread to take up in other studies.

In the context of my own study, the perspective of public vs. private sector organisations becomes most interesting if one sees a difference in PA practices across the two sectors. To get a picture of this, a good place to look is in the articles that have a cross-sector comparative design.

Prowse and Prowse (2009) examine research through a literature review of articles that evaluate the aims and methods of appraisal, and the difficulties encountered in the appraisal process. The context is primarily US and UK. It is found that in terms of use of PA as the basis for merit pay, there is little difference between public and private sectors, since 65% of the public vs. 69% of the private organisations having both a merit pay and a PA system had implemented such a link. On the other hand, Prowse and Prowse sees a difference in what the employees in the two sectors see as the top issue with PA schemes: in the public sector it is lack of feedback and awareness, in the private sector it is lack of linkage between effort and performance (Prowse and Prowse, 2009: 74).

In a Jordanian context, Abu-Doleh and Weir's (2007) study reveals that private organizations' PA had significantly greater impact than their counterparts in the public sector on key human resource decisions like promotion, retention, lay-offs and for identifying individual training needs, transfers and assignments.

In an Indian context, Amba-Rao et al. (2000) examine PA differences between public organisations, private organisations and foreign owned/joint venture organisations. The study concludes that foreign owned/joint venture organisations were conducting the most formalised PA schemes. Further, it was found that public sector organisations were the least likely proponents of a link between PA and pay or disciplinary consequences. Lastly, an interesting finding was that while public sector organisations were focusing heavily on external compliance with legal and regulatory guidelines, private sector organisations were more likely to accord ethical standards (behaviour) greater priority beyond economic efficiency and legal compliance.

Unfortunately, only these three of the thirteen articles working with cross-sector populations, makes any conclusions on differences between the sectors. The other ten articles (Millward et al., 2010. Cheng and Cascio, 2009. Sudarsan, 2009. Stark and Poppler, 2009. Wright and Cheung, 2007. Soltani et al., 2006 & 2004. Tziner et al., 2001. Raymark et al., 1999. Snape et al., 1998) have other foci. So, as we have only weak evidence of any differences that would be of relevance also to my own study, I prefer only to conclude two things. First, that there is a significant overweight around a factor of two, of empirical focus in favour of public sector organisations when compared to their relative share of the employment base. My own focus on a private, pharmaceutical sector company will make a little contribution to turn the scale of this imbalance. Second, the scene is still open to researchers who want to make comparative studies of differences in PA schemes and practices across the public and private sectors.

Geographical focus

If the public/private sector study distribution is somewhat skewed in the articles, the picture is not much different when one turns to the geographical point of view, as illustrated in Figure 2-1 below.

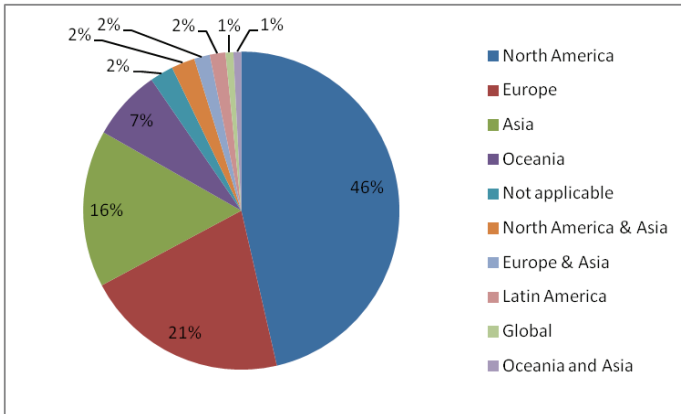


Figure 2-1: Distribution of empirical focus by continent

Figure 2-1 shows how North America with 46%, or almost half of all journal articles selected, dominated the empirical focus of the articles. As runner up Europe reached 21%, Asia 16%, Oceania 7% and various combinations of countries across continents covered the rest. Looking at a country by country level, the top 5 countries in terms of share of the 125 analysed articles were: US 40%, UK 17%, Canada 6%, Australia 6% and China 4%.

However, even this strong bias towards North America and Europe at a closer look could be argued to over-estimate the empirical focus of academic articles on non-Western populations. Since Asia also includes countries we would normally consider western in terms of economic and political system, i.e. here Japan and Israel, and Oceania covers Australia and New Zealand which we would also consider Western, one could argue to see the distribution as skewed with up to 82% Western studies. This needs to be held up against the fact that Western populations only account for approximately 15% of the total world population. So, no matter if one argues that the 3 articles focusing on Japan and the 4 articles focusing on Israel should belong to one or the other category, it is clear that there is a strong and systematic bias in the empirical focus in favour of Western populations.

Since my own study places itself in this mainstream tendency by focusing on a Scandinavian context, I can hardly claim to be setting out to do anything really original in terms of geographical focus. The broader question if and how non-Western

cultural particularities may or may not have an impact on practice of PA in organisations and the way scholars analyse it, I will leave for others to investigate.

Workforce focus

Turning the focus from the geographical and public/private sector dimensions to the workforce dimension, other interesting insights become available. Each of the 125 articles was coded for their focus on specific or mixed workforces. My taxonomy included five main categories: White collar, blue collar, R&D specific, Management only and HR professionals only. The outcome of the coding is illustrated in Figure 2-2 below.

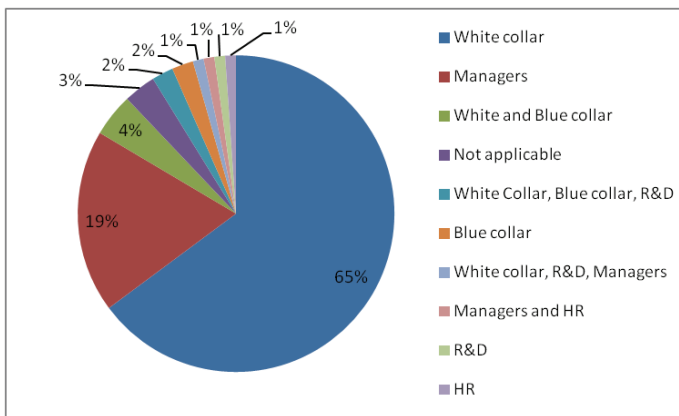


Figure 2-2: Distribution of empirical focus by workforce¹⁴

As is evident from Figure 2-2, 65% of the empirical articles were focusing on white collar employees and 19% on managers only. Very few articles focused on blue collar, HR and R&D or on combinations of the categories.

Further, one could reasonably argue that the management workforce is most often also white collar work even in e.g. a production or retail company setting with blue collar workers. And HR definitely is a white collar category. Thus, it would be reasonable to conclude that no less than 85% of all empirical articles have a white collar focus. It is obvious that the above documented geographical focus plays a role in this, since the Western economies over the past decades have turned more and

¹⁴ In percentage of 91 empirically based articles out of 125 analysed journals

more into service oriented economies with increasingly smaller proportions of classical blue-collar jobs. However, there is also another bias at stake. Probably for convenience reasons scholars often use higher education students or university staffs as their empirical basis when doing social research, particularly when using experimental or quasi-experimental methods. Thus, three out of the four articles coded as using experimental or quasi-experimental methods in my review, were using university staffs and students as their empirical object (Payne et al., 2009. Cardy et al., 1998. Wagner and Goffin, 1997). Peterson (2001) conducted a large meta-analysis of the implications of the use of students in empirical studies as representatives for other groups. He concluded that not only did students show a slight but statistically significant higher homogeneity than non-students; he also concluded that "effect sizes derived from college student subjects frequently differed from those derived from nonstudent subjects both directionally and in magnitude" (Peterson, 2001: 1).

If the heavy focus on white collar workers and managers was not surprising, I still was surprised to see that no attention has been given to cross-workforce comparative studies. In fact, only 8% of the articles were looking at populations covering more than one workforce when studying PA (Brown et al., 2010. Soltani et al., 2006. Brown and Heywood, 2005. Woods, 2003. Shibata, 2002. Tziner et al., 2002. Shibata, 2000. Shaikh, 1995). But of these, none actually performs a comparative study to examine workforce differences. Woods (2003) conducts a cultural comparative study including multiple workforces, but his comparison parameter is national culture rather than workforce, as he sets out to investigate differences in Australian and Singaporean PA practices. Shibata (2002) is working from the same approach, but compares US and Japanese PA practices and underlying cultural causalities. Tziner et al. (2002) use both white collar and blue collar workforces in US and Israel, but his comparison is not between workforces, but between two different approaches to rating scale and performance feedback systems. Shibata (2000) conducts a diachronic comparative analysis of the interrelation between performance rating and wage before and after implementing a new PA system in a Japanese organisation. The other four articles with cross-workforce empirical data are not comparative studies. So, while these comparative

studies based on cross-workforce populations are interesting, none of them are actually dealing with workforce as the comparative dimension.

Summary of empirical review

By looking at the empirical focus of my sample of the last twenty years' research into PA a number of findings of interest to my own study appeared. First, there is a significant overweight of public sector studies as opposed to private sector studies, when compared to the share of the full employment base of the two sectors. Since some studies indicate that there are differences in the experiences with PA between the two sectors, this indicates a need for further focus on private sector studies.

Second, there is a huge overweight of studies within a Western context. Almost two thirds of all the empirical studies in my sample were made on populations from three countries: US, UK and Canada. Only one of the articles focused on a Scandinavian empirical context, where one could expect to find deviations from the broader Anglo-American findings, since Scandinavia is often presented as having a more democratic leadership tradition and a less significant power distance than almost any other region in the world. This could leave room for my own study to add to a more balanced picture of PA, although I will not be able to change the overall imbalance between Western and non-Western studies.

Third, there is a significant overweight of empirical studies into one single workforce. 85% of all studies were focusing on a white collar empirical population. Only 1% of the articles were focusing on R&D specifically. Even more interestingly, only 8% of the articles were studying more than one workforce, but none of these were in fact comparing experiences between workforces.

Thus, my literature study clearly indicates that there is a lack of comparative research into PA practices across workforces, in a non-Anglo-American context and within the private sector. Is this an important gap? In my opinion it is. Multiple authors have investigated particular groups of employees to identify the specific challenges they meet in a PA context. For example Ubeda and Santos (2007) focused on R&D, Abu-Doleh and Weir (2007) focused on HR, Yariv (2006) focused on managers and many focused on white collar workers. But comparative studies were surprisingly absent. To improve our understanding of possible differences between workforces it is necessary to take one more step from all these single workforce analyses, and

conduct a truly comparative study of workforce particularities. Having good access to a large organisation that has implemented a uniform PA system across multiple workforces puts me in a good position to start filling that gap.

2.5 Analysing the Methodological Approach of the Literature

The 125 articles were coded for their methodological approach, and the distribution can be seen in Figure 2-3 below.

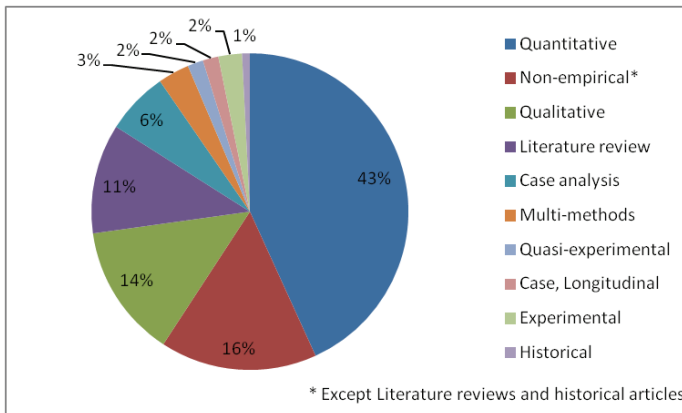


Figure 2-3: Distribution of methodologies utilised in 125 analysed articles

Figure 2-3 shows the distribution of methodologies utilised in the 125 articles. The largest category is by far the quantitative studies comprising 43% of all 125 articles. This is not surprising, since it is often referred to as a fact that many top tier journals have a preference for quantitatively based analysis (Marschan-Piekkari & Welch, 2004: Chapter 27). But here we are more concerned with what it means for the type of findings the authors are able to produce.

Quantitative methods

Quantitative methods rely on numbers and statistical analysis of numbers. In this context the quantification is most often based on survey questionnaires, and sometimes correlated to other quantifiable data, e.g. size of bonus, salary increase or other rewards. The quantitative approach is well suited if one is out to produce statistically reliable tests of hypothesis about already defined problems where a few

independent variables can be defined and isolated. Often, this approach is chosen based on a desire to test hypothesis about respondents' opinions or attitudes towards certain elements in the PA system under which they operate.

For example, Salimäki and Jämsen (2010), Thurston and McNall (2010), Roch et al. (2007), Vest et al. (2000), Waite and Stites-Doe (2000) and Holbrook (1999) all set out to test hypothesis regarding employees' perception of the degree of fairness in the PA scheme under which they work. Salimäki and Jämsen (2010) analyse employee survey responses from three governmental organizations (N=367). The purpose is to see if perceptions of politics and fairness are predictors of whether the pay system was perceived effective in achieving its objectives. The study concludes that there is such a correlation.

Thurston and McNall (2010) conduct a nested confirmatory factor analysis of US employee responses (N=188) and investigates if four different categories of perceived justice in the organisation can be correlated to satisfaction with the organisation's PA system, appraisal ratings and employee satisfaction with their managers. The study concludes that these relations can be verified.

Roch et al. (2007) conduct two US studies, (N = 208) and (N = 393) to investigate employees' preference for absolute vs. relative rating scales. The studies find that absolute formats are perceived as significantly fairer than relative formats. However, it is also found that significant differences in fairness perceptions also occur within relative and absolute formats.

Other examples are Cook and Crossman (2004) and Snape et al. (1998) who both investigate the more general employee attitudes towards PA schemes. Cook and Crossman (2004) conduct a UK survey (N = 382) to test a hypothesis that a person who's role is appraisee will be less satisfied with a PA scheme than a person who is both appraisee and appraiser (typically managers). In the public sector sample examined, there was no significant correlation so the hypothesis could not be verified, although other studies have indicated so (Cook and Crossman, 2004: 533).

Snape et al. (1998) examine attitudes on PA in the UK and Hong Kong. Respondents (N = 276) were mature part-time MBA students and graduates; (N = 141) were from Hong Kong and (N = 135) from UK. The study finds that British appraisals tend to be more participative and to place greater emphasis on discussing objectives,

development and career plans. Hong Kong appraisals on the other hand appear to be more directive and Hong Kong respondents perceive a higher level of 'negative' appraiser behaviour. Nevertheless, Hong Kong respondents show stronger support for appraisal's use for reward and punishment and less support for the objectives-setting and training and development uses than do the British sample.

Tziner et al. (2002, 1998), Langan-Fox et al. (1998), Shore et al. (1998) and Greller (1998) all examine the effect of bias in PAs. Tziner et al. (2002) examined a hypothesis that managers' conscientiousness moderates the impact of their attitudes toward PA on their rating behaviour. The hypothesis was verified, as correlations between attitudes and rating behaviour were significantly weaker for managers with high conscientiousness.

Langan-Fox et al. (1998) conducted a survey among shop floor employees in an Australian government enterprise (N = 476) where participation in the PA scheme was voluntary for the employees. Here, the bias the researchers were looking for was the prejudiced attitude of the employees when deciding whether to participate in the scheme. It was found that attitudes towards PA in general, perception of consequences of PA, satisfaction with the supervisor, anticipated feedback and awareness of the system were correlated to the decision to adopt or reject the scheme.

Other hypothesis handled by utilisation of quantitative methods include employee coping strategies, general process satisfaction, rating effects for employees, relation between PA and stress, rating reliability for employees working in teams etc.

As should be evident from these examples, authors utilising quantitative methods often have well defined, narrow and unambiguous questions and/or hypothesis that can be investigated by asking closed questions in a survey format.

Qualitative methods

Quantitative studies are in many ways contrary to the kind of studies conducted by the authors utilising qualitative methods, which account for 14% of the articles. Qualitative research is research that produce findings without statistical procedures or quantification (Strauss & Corbijn, 1990: 17). These authors are most often embarking on more open questions. They are not interested in setting up hypotheses to be tested through closed question surveys. Rather, their questions and

approaches are of a more exploratory character, often based on few but lengthy interviews. Some examples below will illustrate this.

Millward et al. (2010) give a psychological analysis of managers' feedback seeking. Over two in depth individual interviews with ten UK managers, the managers were reflecting systematically on their own feedback seeking in the organisation. Their approach was found to be very instrumental, in the sense that "managers sought their feedback remotely, and from largely external sources, to reconcile development needs with self-protective considerations (i.e. image and ego-costs) in relation to subordinates and peers" (Millward et al., 2010: 384).

St-Onge et al. (2009) adopt in-depth interviews with 15 MBA students at a North American university. The interviewees were selected because they formed the extreme ends of respondents to a survey that had been administered to all 85 enrolled MBA students, surveying their level of motivation to conduct employee PAs at their workplace. The study revealed that researchers must integrate various theoretical perspectives (rational, affective, political, strategic, cultural, justice, and symbolic), to understand managers' motivation to conduct PA. Also, the study revealed that managers' motivation to conduct PAs are not so much about technique or system design, but more about the degree of leadership support and overall performance culture in the organisation.

Yariv (2006) examines what he calls the "mum effect", i.e. managers' reluctance to give negative performance feedback. Forty Israeli elementary school principals were interviewed about cases where they had to give feedback on poor performance. Four stages with growing degrees of directness and criticism were identified as a possible typology; each stage depended on the outcome of the previous one. Half of the principals initially simply ignored the poor performance while at a later stage most principals were conducting a "soft negative feedback" interview with the poor performing employee. In the cases where this did not result in improvement, most principals criticized the employee, but when even this did not result in improved performance, only one third proceeded to the last identified step: critique in writing. The process was on the principal's side accompanied by strong emotions of both anger and compassion, illustrating how interpersonal relations were influencing next steps for the principal and how manager's behaviour fruitfully can be observed through an "emotional lens".

Spicer and Ahmad (2006) conducted 22 in-depth cognitive mapping interviews with appraisers in the Malaysian education system, to map the cognitive processing models underlying appraisers' decision making. Through the interviews, the appraisers' cognitive processing was mapped against a six-step conceptual model to examine its applicability and to learn if appraiser experience was influencing the use of the model. The six-step conceptual model was not revealed to the interviewees and the interviews were undertaken with departure in the interviewees' own wording, which the interviewer then later translated into theoretical categories. The interviews revealed considerable variability between individual maps, but still the maps were found to be aligned with the six-step conceptual map, with the more experienced appraisers being most aligned to the conceptual model.

In a last example of utilisation of qualitative methods, Narcisse and Harcourt (2008) interviewed 20 employees in a West Indian public service organisation. The study found that distributive, procedural, and interactional justice factors were influencing employee perceptions of fairness in their appraisals. The interviews uncovered four concrete justice factors affecting perceptions of justice: one distributive - the consistency in reward distribution; and three procedural - appraisal frequency, job relevant criteria, and appraiser and appraisee training.

In the above and other qualitative studies, the authors are searching for patterns from a more or less structured approach but in all cases based on in-depth interviews with relatively few interviewees and with the purpose of theorising through concept building, uncovering complex interdependencies and combining multiple factors rather than isolating independent variables as would be required in a quantitative context. From this work the authors do not reach the hard-data reliability that users of quantitative methods are looking for. Rather, instead of generalising to populations, they generalise to theory building.

Non-empirical methodologies

Having dealt with the two most common methodological approaches for analysing empirical data above, let us for a moment turn to the non-empirical studies which naturally do not utilise any empirical methodological approach. There are three categories of non-empirical studies: Literature reviews, historical, and a mixed group of other non-empirical studies.

Literature reviews are studies that scan the current body of literature as a means to investigate a research question. For example, Claus and Briscoe (2009) examine 64 articles identified via database search, from the period between 1985 and 2005 dealing with international PA. The study concludes that the academic literature on cross-border PA is relatively non-theoretical and exploratory in nature and that the design and substance of the research studies are weak.

Rasch (2004) reviews literature regarding the process of PA in higher education, focusing on articles supportive of the use of PA, and those that are not. It is concluded that the advocates of PA cite implementation flaws as the cause for dissatisfaction among users, while critics of PAs question the validity of the process in general.

Flint (1999) examines the literature on organizational justice and proposes a model to determine how employees will react to PA ratings. Based on this, the author suggests that when PA ratings are low, employees will use procedural justice evaluations to determine the fairness of those ratings. If an employee perceives a low rating as procedurally fair, he may act to improve performance. However, the employee's performance may decrease if the rating is perceived to be unfair.

Other examples on themes analysed through literature studies are: The impact of computer technology monitoring (Bates and Holton, 1995), connections between rating and like/dislike personal relations between manager and employee (Lefkowitz, 2000), investigation of a literature focus movement over time from bias research to motivational effects (Fletcher, 2001) and identification of factors that affect the success of team based PA (Mendibil and MacBryde, 2006).

Among the 125 articles there is only one that utilises a historical approach (Wiese and Buckley, 1998). The study goes back to the earliest days of PA which according to the study can be traced back to the Chinese Han dynasty 206 BC, where evaluations played a part in building a merit based administration. The study continues with historical examples and identifies the early 1800s as the time when formal PA started to become integrated in the UK cotton industry and in the US military. The study examines how the use of goals and tools has shifted over time, as has the view on rater errors. In more modern times, the study examines various shifts

in views on PA: management by objectives, focus on behaviour, focus on legal issues and to today's attempts to create PA schemes that are valid and reliable.

The "Other non-empirical studies" category is a mixed group. Three articles are of purely theoretical academic nature, e.g. DeNisi and Pritchard (2006), who introduce expectancy theory and suggest its possible use to investigate PA. The rest are more practically oriented "how-to" articles from professional journals like *Work Study*, *Managerial Auditing Journal*, *Personnel Review*, *Journal of Managerial Psychology* and *Career Development International*.

Case studies and mixed methods studies

The term 'case study' is not a trivial one. Ragin and Becker, in their discussion about what a case study is, emphasise that there are at least four different basic conceptions of this. Some studies treat cases as if they are "out there" waiting to be discovered and analysed. Other studies use cases as objects without challenging their boundaries, when analysing for example nation states. A third category of case studies see cases as made by the researcher. Here the researchers do not see cases as empirical as such or as given, but rather as a theoretical construct made by the researcher to define and analyse a subject. The fourth category of case studies see cases as conventions based on collective scholarly work, like for example "industrial society". In their book Ragin and Becker conclude that the discussion of "what is a case?" is a discussion without any real beginning or end. And that it may sometimes be more fruitful to ask oneself the question "what is it a case of?" instead (Ragin and Becker, 2009: 7-11).

Methodologically, Yin defines a case study as one that,

investigates a contemporary phenomenon in depth and within its real life context... which relies on multiple sources of evidence, with data needing to converge in a triangulating fashion. (Yin, 2009: 18)

In my coding of case studies among the 125 journal articles I used the principle of recognition of the authors' own labelling - i.e. when an author labelled her study as a case study, I coded it as a case study. By this procedure, ten studies were case studies and four studies were mixed methods studies that did not use the label case study.

I have here grouped case studies and mixed methods studies together because they are often overlapping categories, in terms of Yin's definition. However, although case studies are often using a mixed methods approach, and Yin suggests all should, not all cases studies live up to this. And it is also seen that some studies use a mixed methods approach without labelling it a case study. Four examples of studies coded as case studies or mixed methods studies among the 125 journal articles will help to illustrate this complexity.

In a case study, Wilson and Western (2001) examine a UK hospital to identify potential inhibitors which can reduce the effectiveness of a hospital PA system in relation to training and development plans. They do this by using a mixed methods approach containing participant observation, a survey questionnaire, semi-structured interviews and review of training records in the case organisation.

Orpen (1995) on the other hand conducts a single method case study in an Australian private company to identify the extent to which employee job performance and employee relations with their superior moderate the impact of goal setting. He does this by leveraging a purely quantitative study design in four phases where supervisors and employees filled out different schemas regarding evaluation, self-evaluation, process evaluation and motivational effect assessment.

For Latham et al. (2008) a case study is just one among four different methods utilised in a study. Latham et al. investigate the influence of manager's own performance rating on their evaluation of others. In a case study, the ratings received by two area coordinators in a university academic department seemed to affect their subsequent ratings of faculty. Here the case is narrowly defined as limited to the study of the effect of own rating for only two area coordinators in a university department. Separately, in a simulation, 30 managers received hypothetical feedback regarding their own job performance. The managers subsequently evaluated an employee on videotape. Managers who received positive feedback about their performance subsequently rated the employee significantly higher than managers who received negative feedback regarding their own performance. This occurred despite the fact that the managers knew the evaluation of them was bogus. The results of two additional follow-up field studies involving 74 manager–employee dyads in a manufacturing company in Canada and 39 manager–subordinate dyads in a retail organization in Turkey were consistent with the view that one's own PA is

related to the subsequent appraisal of one's subordinates. So in this design Latham et al. use the narrow case study and the other three studies as supplementary to each other, rather than seeing all four of them as being one integrated case study.

McAllister et al. (2003) do not label his study of appraiser behaviour and appraisal smoothing a case study, but still use a mixed methods approach. A qualitative interview survey of fund managers in UK and their appraisers is combined with an empirical study of the number of appraisals which change each month within a recognised market Index.

Of the ten studies that the authors label as case studies, only two are utilising a longitudinal approach. Shibata (2000) examines how a Japanese firm has seen that wage and performance-appraisal systems have been shifting to a performance basis, and away from the previous heavy reliance on age and seniority. Lee and Son (1998) investigate a Korean company to understand whether appraisal reviews actually change employees' subsequent performance.

Experimental, quasi-experimental studies

Four articles use experimental or quasi-experimental approaches. This type of approach is often chosen when the purpose is to test very concrete hypothesis in a controlled setting. For example, Payne et al. (2009) utilise a quasi-experimental design to investigate participation level and degree of participant satisfaction by comparing the reactions of a group of 83 employees evaluated with a traditional paper based PA instrument to the reactions of a group of 152 employees evaluated with an online version of the same assessment tool.

Cardy et al. (1998) use an experimental design to investigate the effect on PA of system vs. employee responsibility for good or bad unit performance. System produced errors and person produced errors were manipulated in two studies and the outcome was mapped to performance ratings given by management. The study showed that ratings were lower when the ratee had greater responsibility for production than did the system.

Summary of methodological review

By looking at the methodological approach utilised in my sample of the last twenty years' research into PA a number of findings of interest to my own study appeared.

First, quantitative studies were significantly more frequent than any other methodology. There were three times as many quantitative studies than qualitative studies in my sample.

Second, the group of quantitative studies, experimental studies and quasi-experimental studies were generally characterised by narrow and unambiguous research questions, often formulated as one or more hypothesis testable against closed question survey responses from a representative group of employees. On the other hand, articles concerned with more open ended, exploratory questions were typically utilising a qualitative approach based on deep dive interviews or a case analysis approach.

Third, except for only two studies, all empirical articles are concerned with data collected at one given point in time. Only two studies have a longitudinal design. When presenting my study protocol in a later chapter, I will elaborate on why I find this to be very strange since PA in my opinion is an obvious process to study by longitudinal designs.

Fourth, only four of the articles are using a mixed-methods approach, although it is generally agreed that methodological triangulation is a way to increase validity in most studies (Yin, 2009. Alvesson and Sköldbberg, 2009).

Thus, my literature review clearly indicates that there is a lack of longitudinal, qualitatively oriented studies which handle the more complex aspects of PA. Further, if such a design could be combined with a methodological triangulation through the use of supplementary data establishment methods, there seems to be a good chance to create new knowledge about PA.

2.6 Analysing the Theoretical Approach of the Literature

Coding for theoretical approach was the most difficult part of the literature review. Not all authors are so nice as to start by stating under which single theoretical approach they intend to conduct their study. In fact, surprisingly many authors do not reflect on what existing theoretical perspectives their contribution relate or contribute to. Others are multi-theoretical in their approach - some by explicit, up-front choice, others de facto by the way of arguing along the way. So the first challenge was to define a distinction between when to code an article as belonging to a theoretical category at

all. In my opinion, it would be possible to identify elements of theory in any of the 125 articles, even if the author did not explicitly do this. In fact, it would probably be possible even where the author never gave it a thought. However, I find it valuable to distinguish between two types of articles: A) articles that are explicitly relating to theory, either by drawing on it in the analysis or by contributing to theory, or by doing both. And, B) articles that are non-explicit about their relation to existing theory.

Although I am aware the term is somewhat misleading, I have, in lack of a better term, used 'non-theoretical' to describe the latter category. Thus, non-theoretical means in the following that no explicit overall theoretical approach at the framework or meta-level is used or referred to in the article. That is, an article may look for general concepts like "best practice", it may investigate "issues", it may examine "motivation effects" or "participant satisfaction" or similar, and it may also propose generalised statements about the possible use of the article findings beyond its own study object, a manoeuvre one could easily justify as being theoretical by nature. However, the author does this without explicitly attaching these reflections to any broader theoretical perspective.

Going through the 125 journal articles, two initial key findings became apparent. First, a majority of 64% of the articles were non-theoretical according to the above definition. Second, there was not a handful of dominating theoretical approaches across the articles. Rather, there was a plethora of different theoretical approaches so that I ended up identifying 27 different theoretical approaches across the 44 articles that were coded as belonging to one or more theoretical category. Let us start by looking at the articles coded as theoretical. The distribution of these is illustrated in Figure 2-4 below.

2. The PA Conversation - a Literature Review

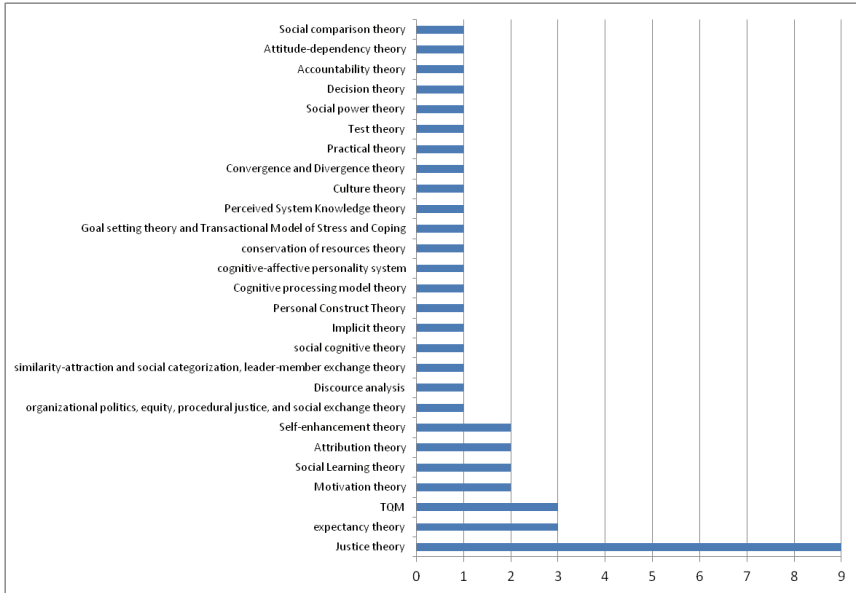


Figure 2-4: Distribution of article occurrences on theoretical perspectives

Figure 2-4 shows how many articles that claim to contribute to each of the 27 identified theoretical perspectives. The figure seems to illustrate a very heterogeneous theoretical composition of the articles. However, at closer look some initial considerations appear to be appropriate.

First, there is the question of what defines theory, as opposed to e.g. methodology. Some articles claimed to use grounded theory, which I, despite the reference to theory, coded as a method used in qualitative studies. But what about discourse analysis, then? Although normally seen as a method, I have here taken the liberty to code it as theoretical, because the St-Onge et al. (2009) article not only uses discourse analysis in a qualitative study as a means to work with the material, but also theorises about how the manager discourse can be actively changed to improve the effectiveness of the PA process.

Second, the question of what defines a theory is important. For example, is justice theory one theory, or are there in fact two fundamentally differing justice theories; one based within psychology with departure in the justice perception of the individual, and another based in sociology which is relatively and collectively based? I have

here decided to use the authors' own claim at face value without proposing even more fine grained definitions than the above 27 categories. This is not to ignore the fact that such an analysis would bring improved understanding of the theories in use, but it lies outside the objective of this study.

Third, readers acquainted with more than a handful of the theories listed in Figure 2-4 above will probably already have seen that some of the theories are closely related to each other. Before turning towards a more detailed discussion of the most used theories and the findings they produced, it could be beneficial for the purpose of maintaining an overview, to consider how one could group the 27 theories at a meta-theory level. This would be beneficial in terms of adding to the understanding of the overall theoretical perspectives into which the 27 theories fit. One way to make such a grouping would be to categorise the theories into three broad perspectives ranging from individual psychology over social psychology and to sociologically oriented theories.

In such a construction, the first meta-category would be the broad category of theories based on individual psychology. Roughly split, this category would include expectancy theory, attribution theory, personal construct theory, cognitive processing model theory, cognitive-affective personality system, test theory, self-enhancement theory, conservation of resources theory and attitude dependency theory.

The second meta-category would be theories oriented towards social psychological explanatory factors. This category would include implicit theory, accountability theory and social comparison theory.

The third meta-category would be theories oriented towards sociological explanatory factors. This category would include TQM, social power theory and possibly culture theory.

However, some theories would be harder to put into one of these three boxes. For example, as already mentioned, justice theory can operate at least at the first two levels. Practice theory is also somewhat cutting across the three levels. However, as a rough rule of thumb I believe it can be beneficial to see theoretical approaches to analysing PA to focus at one or sometimes more of the three categories: individual psychology, social psychology and the sociological system/process level. Or in other words; at the level of the individual, the group or the system.

For the purpose of the present analysis it will not be beneficial to take this categorisation discussion any further. Below, the most used theories and the findings they produced will instead be analysed as they have been named and put forward by the authors. The one theory that sticks out as being the most used is Justice Theory, used in nine articles. Let us start by looking at these articles.

Justice Theory

Justice theory, or organisational justice theory, is used by Thurston and McNall (2010), Narcisse and Harcourt (2008), Roch et al. (2007), Cook and Crossman (2004), Poon (2004), Holbrook (1999, 2002), Flint (1999) and Korsgaard and Roberson (1995). According to Thurston and McNall (2010) Organizational justice theory is deeply rooted in social exchange theory which takes departure in the assumption that

social relationships are viewed as exchange processes in which people make contributions for which they expect certain outcomes; and, individuals evaluate the fairness of these exchanges using information gained through social interactions. (Thurston and McNall, 2010: 204)

As a consequence of this assumption and in the context of PA, Justice Theory would in its simplest form claim that there has to be a balance between people's contribution and their share of reward, for people to perceive the exchange as fair. However, later developments of the theory separate between three different forms of justice; distributive, procedural and interactional justice.

Distributive justice concerns the distribution of resources among a group of people. In a PA context this could be the performance ratings but it could also be other rewards like bonus, promotion, better assignments or praise. This concept deals with the justice perception of outcomes.

Procedural justice concerns the perception of process fairness. The claim here is that people will tend to accept even disappointing ratings if they perceive that the rating process is transparent, consistent and fair, i.e. without bias.

Interactional justice concerns the perception of the personal treatment someone receives while being rated. This concerns the quality of the dialogue including the ability of the manager to deliver the rating message in a way that is perceived as respectful to the employee. This type of justice also includes what some authors

choose to see as a separate form, informational justice, which concerns the procedural explanations for why an outcome happened.

As Thurston and McNall (2010: 204) put it:

Perceptions of inaccuracy and injustice as well as feelings of dissatisfaction have long plagued PAs and the organizational processes that generate them.

After this short introduction to Justice Theory, let us turn to look at some of the key findings that have been generated from this perspective in the articles belonging to this category.

Thurston and McNall's (2010) study claims that the different theoretical types of justice perception are distinct but also so highly correlated that it can be discussed how much practical importance the separation has. In the study, justice perceptions were investigated at the general employee-organisation relation level rather than specifically for the PA process. However, the study did show a significant correlation so that employees who rated their perception of justice within the organisation as high, were also the more satisfied with the PA scheme in the organisation.

Poon (2004) states that there is evidence that performance ratings are often manipulated for political purposes which in turn are perceived as unfair by the employees. With departure in justice theory she examines the effects of employees' perceptions of political motives in PA by relating this to their job satisfaction and intention to quit. Her analysis shows that when employees perceived performance ratings to be unjustly manipulated because of raters' personal bias and intent to punish subordinates, they expressed reduced job satisfaction that, in turn, led to greater intentions to quit their jobs.

Holbrook (2002) looks into how the use of justice mechanisms can be conceptualised into the PA interview situation. The study shows how the PA process contains a number of activities that each affects the justice perception of the employee. Specifically, he mentions the evaluation process, the assignment of ratings and the interview between manager and employee. Although not mentioned by Holbrook, one could easily add other activities like goal setting, prioritisation, assignment workload etc. as events that can affect the overall PA process justice perception of the employee. According to Holbrook it is rare to find organisational events that contain all three theoretical forms of justice. But since these are all present in the PA

process, Holbrook sees the process as an excellent study object for scholars working within Justice Theory.

Generally, the relatively high number of articles concerned with justice theory can be seen as an indication of the fact that perceptions of injustice or unfairness appears to be a central issue to many PA scheme implementations. It is not hard to see how a scheme characterised by low justice perception will have a hard time being motivating for the employees. What Justice Theory brings to the table is a systematic approach on how to investigate the elements constituting justice perceptions.

Expectancy Theory

Expectancy theory is a variant of motivation theory proposed by Vroom (1964) which focuses on how employees' motivation is a function of the perceived reward of conducting a specific task and the expectancy, or confidence, the employee has in his own ability to accomplish the task. In a PA context this means that if conducting the task to a certain quality level is perceived by the employee to be rewarding to himself, then the employee will be motivated to conduct the task. However, the job of the manager is to find out what the employee perceives as "rewarding" enough to find it worthwhile to conduct the task. The employee could be valuing a financial bonus but he could also be more interested in recognition or other less tangible rewards. Second, the employees' expectancy of own capability to conduct the task satisfactorily directly affects motivation. Even if the reward is perceived as attractive, the total motivational effect will be negative if the expectancy factor is negative, i.e. if the employee perceives it as unlikely that he will be able to meet the desired standards. In fact, if expectancy is negative so that the employee finds the task unobtainable, a higher perceived reward will only add to the level of negative motivational effect (Vroom, 1964).

With departure in this framework, DeNisi and Pritchard (2006) in a non-empirical contribution define the PA process as a sequence where actions lead to results that are then evaluated and rewarded to a certain level of need satisfaction. It is then proposed that the stronger the perceived linkages between each step in this sequence, the greater will the employee's motivation be. This may sound trivial, but considering the amount of literature on rating and measurement bias, interdependency in outcomes and differences in employee reward preferences, this

has significant impact on the way one deals with PA in practice. Based on this assumption, DeNisi and Pritchard raise a number of claims. First, if the perceived connection between an individual's contribution and the desired outcome is weak, which would often be the case when more strategic goals are cascaded down through an organisation, the motivation effect will be negative. Second, the perceived ability of the individual to fulfil the required task; be it due to capabilities or resource availability, will directly affect the employee's perception of the link between action and outcome and thus the employee's motivation and performance. Hence, PA systems should include a systematic aid to the employee's perceived ability to fulfil the tasks, to strengthen the perceived link between action and outcome. Third, PA systems should aim to strengthen the perceived link between outcome and evaluation. If the employee evaluation is expected by the employee to have only a weak connection to the outcome obtained, this will have a negative impact on motivation and performance. In other words, the perceived fairness and transparency of the rating system has a direct impact on employee motivation and performance. Fourth, PA systems that are expected to deliver a weak link between evaluation and the reward preference of the individual will have a negative impact on motivation and performance.

Smith and Rupp (2004) examine knowledge worker's perception of performance ratings with departure in the expectancy theory claim that for a PA system to have a positive effect on motivation and performance, high performance ratings should entail high merit increases while low performance ratings result in low merit increases. Through a survey to knowledge workers in 5 medium to large US companies, Smith and Rupp found a weak correlation between appraisal rating and merit increases. 58% of the knowledge workers perceived the link between performance rating and merit increase to be weak. In fact, "it appeared that decoupling performance ratings and merit increases is a common practice" (Smith and Rupp, 2004). Smith and Rupp present a range of explanations to this paradox, proposed by the respondents to the survey. The most common explanations were budget constraints, subjective factors, manager bias, external conditions and a general perceived decoupling of employee contribution and outcome. Using expectancy theory, Smith and Rupp claim that this decoupling of performance and reward may in fact be the key to understanding the widespread failure of PA schemes.

Lee and Son (1998) conducted a longitudinal study in a Korean petrochemical company to investigate employee perception of review quality and utility, and impact on later employee performance. Three independent variables were tested: degree of employee perceived opportunity to participate in a discussion with the manager during the review, the degree of perceived goal clarity, and the degree of perceived inclusion of career themes in the review interview. Somewhat surprisingly, the study concluded that although all three factors were correlated to the employees' later perception of review quality and utility, none of the factors had any significant impact on later performance. Lee and Son suggest that this has to do with a more general fact of the PA practice itself, i.e. that the performance-outcome relationship is weak since performance evaluations for a certain employee have a tendency to be stable over time no matter the fluctuations in actual performance.

What expectancy brings to the table is insight into how important it is that performance goals are set in a way so that the employees see them as not only relevant and rewarding, but also realistic to achieve. If individual targets are unrealistic or if the number of targets is so huge that fulfilling them all becomes unrealistic, then the process will have a negative impact on employee motivation. As we shall later see, this is something managers and employees in my own study are in strong agreement on.

Motivation Theory

Motivation theory is of course a much too broad categorisation of a group of theories that concern human work motivation. Most writings on PA somehow relate to the subject of motivation. However, some relate more specifically to motivational theories. Buchner (2007) investigates in a non-empirical article how applied models of PA do not take full advantage of three key theories of work motivation: Goal-setting Theory, Control Theory and Social Cognitive Theory. He claims that the general predominance of top-down approaches to PA has a negative impact on motivation in all these three perspectives and that it is thus constraining the positive impact of PA schemes.

Kuvaas (2006) explores alternative relationships between PA satisfaction and employees' perceptions of own work performance, affective organizational commitment and turnover intention. Through a survey of employees in 64 Norwegian

savings banks he shows that PA satisfaction is directly related to affective commitment and turnover intention. He also shows how the relationship between PA satisfaction and work performance was moderated by employees' intrinsic work motivation. The moderation revealed a negative relationship for employees with low intrinsic motivation and a positive relationship for those with high intrinsic motivation.

Generally, motivation is a crucial concept in studies of PA because most opponents and proponents of PA schemes claim that it is the effect on motivation that is one of the key arguments for embarking on, or rejecting, the process in the first place. Motivation theory contributes to this discussion by offering a framework for analysing different sources of motivation and how appraisal schemes affect them.

Total Quality Management (TQM)

A separate small group of three articles specifically deal with the links between TQM and PA. This may at first sight seem to be a rather narrow subject. However, for any sector with keen interests in quality issues - and the pharmaceutical sector certainly is one - it may be of interest to consider how PA schemes combine with a quality focus.

Soltani et al. (2004) examine the consistency and congruity of PA systems with quality management precepts in UK-based quality-focused organisations. Soltani et al. claim that a fundamental reason for a purported incompatibility between TQM and PA is the assertion that in the TQM framework most variance in performance is caused by system factors, but that PA schemes focus on person factors rather than on system factors. Soltani et al. found what he calls a "remarkable inconsistency" between the currently dominating PA schemes and TQM demands. Most PA systems fail to meet quality management expectations and contradict TQM assumptions about individual and system performance. In fact, Soltani et al. claim that their study supports Deming's theory that PA schemes are a root cause of quality management problems, because they direct focus so strongly towards the individual, and away from the team and system level. Soltani et al. conclude that,

unpalatable though it may be, is that HR performance evaluation in the majority of the TQM-based organisations surveyed is locked into a vicious circle of individual performance, control approach, HR dissatisfaction, and a low degree of success for TQM programmes. (Soltani et al., 2004: 415)

Along the same lines, Waite and Stites-Doe (2000) examine employees' reactions to an abrupt abolition of a PA and merit pay scheme in the name of a TQM implementation. Through a survey of non-managerial bank employees, it is found that replacing a traditional PA and merit pay system with a non-appraisal feedback and coaching scheme combined with a blanket pay increase enhanced the employees' perceptions of pay fairness, pay satisfaction and job satisfaction. It is concluded that in a quality focused organisation, the managers

...must recognize that their role has changed from that of judge to that of coach.
(Waite and Stites-Doe, 2000: 202)

Through a literature review, Deadrick and Gardner (1999) is aiming to examine each TQM concern with PA, to identify what a PA system that addresses the TQM concerns could look like. It is suggested that it is possible to design an appraisal system that is congruent with TQM, but the system would need to hold the following features: explicit recognition and evaluation of system constraints on performance, both individual and group rating is required, strong strategic links to the individual's performance targets and on-going updates to safeguard the link, and finally focus on the overall utility of the performance rather than a more formal target realisation focus.

The Total Quality Management movement brought an interesting and allegedly critically destructive relation between quality focus on the one side and the typically more individual, quantity oriented focus of performance measurement tools on the other. As we have already seen, measurement researchers like Austin (1996) also found this dilemma intrinsic to PA in that measurement systems would always look for something to count, hence favouring quantitative targets rather than the much more complex quality parameters. But even more important is the point made by TQM proponents, that PA is misleading managers and employees to overestimate the importance of individual's impact on output, and thus take focus from the more impactful system factors.

Attribution Theory

Attribution theory is an example of a psychologically oriented theory that is used to examine bias in the PA process. Attribution Theory is concerned with the mechanisms of how people, through their observation of others' and own behaviour,

try to establish whether something is caused by internal or external forces. The theory holds a concept called the fundamental attribution error, which prescribes a general tendency to take greater account of situational factors' impact on own behaviour while underestimating the same where other people's behaviour is concerned (Furnham, 2005: 264).

Within this framework, Cardy et al. (1998), as we have already seen, established an experiment that showed how ratings were lower when the ratee had greater responsibility for production than did the system.

Colella et al. (1997) review the literature on persons with disabilities to examine the attribution effects of disabilities on performance ratings. To the extent that raters hold certain expectations and stereotypes about people with disabilities, this can influence performance ratings. It is found that some articles did suggest a positive attribution effect in the sense that poor performance of people with disabilities were expected to be attributed to the disability, thus causing a positive bias in the performance evaluation. However, only two empirical studies were identified that directly examined the impact of disability on performance attributions and none of them found any significant disability effect for attributions regarding performance evaluations.

Generally, even the strongest proponents of PA will agree that bias is apparent in a process where one human being is evaluating a full year's performance of another human being. But there are of course huge differences in how much emphasis is put on the effect of the biases, how realistic it is seen to overcome the biases and what effect this has for the overall conclusions about PA. Attribution Theory is just one example of many different ways to deal with the multiple sources of bias inherent in PA schemes.

Other Theoretical Approaches

As illustrated in Figure 2-4 there are a number of other theoretical perspectives apparent in the literature. It would be beyond the purpose of this literature review to introduce them all in a way that paid them the credit they deserve. However, some examples from articles that have not already been dealt with above, can illustrate the variety of questions posed and theoretical approaches utilised in the field.

Stark and Poppler (2009) investigate from a Similarity-Attraction Theory perspective if any differences can be found in performance ratings across gender and ethnicity.

Wood and Marshall (2008) use Social Cognitive Theory to investigate how personal factors and organisational context affects rating accuracy and perceived effectiveness. Heslin and VandeWalle (2008) use Implicit Theory to show how managers who assume that personal attributes are stable inadequately recognise performance changes over time. Spicer and Ahmad (2006) use Cognitive Processing Model Theory to investigate the effectiveness of rating decision making in PA. Nikolau (2003) uses a Cognitive-Affective Personality System approach to show how there is a relationship between personality and job satisfaction but not between personality and job performance-related variables. In their article "Rated to exhaustion?" Brown and Benson (2003) use Conservation of Resources Theory to investigate the level of emotional exhaustion reported by employees due to bad PA process experiences. And Townley (1999) uses Practice Theory to examine responses of chairs of university departments to the news that they would be required to conduct PA.

Non-Theoretical Articles

According to my categorisation, no less than 74 out of the 125 articles were non-theoretical. This is in itself an interesting observation although not a new one. Also Prowse and Prowse (2009) identified the lack of theoretical development in appraisal research in their literature review. Claus and Briscoe (2009) reached a similar conclusion from a literature review on cross-border PA. What are then the questions the authors aim to answer without referencing an established theoretical framework?

One large group of articles focuses on improving the current use of PA by showing how the process will work and deliver good results, if it only is used in a more intelligent way than is mostly the case today. Samarakone (2010) is looking into how the PA process can be improved by using a real-time talent management IT system. Caruth and Humphreys (2008) investigate how to make a PA scheme best support the strategy execution of the organisation. Marques (2008) looks at how upward evaluation is a necessary supplement to traditional PA schemes which will provide a path to greater effectiveness. Catano et al. (2007) investigate how behaviour based competencies can be appraised in a way that is supported by both managers and employees. Mendibil and MacBryde (2006) look at the factors that determine whether

team based PA schemes are successful. Picket (2003: 237) somewhat frustrated with the state of affairs, concludes that,

The annual fiasco of the performance review continues in many organisations despite evidence that it has the potential to be one of the most effective management tools in the entire executive toolkit.

And Harrington (2000) asks rhetorically in his article "Was W. Edwards Deming Wrong?" and provides himself the answer that the problem is not the concept of performance evaluation but the way it has been implemented.

Another large group of articles focus on specific issues or bias in the current practice of PA. Like the category above, the authors in this group are often looking for opportunities to improve the practice of PA, but they do it with outset in specific challenges they see as imminent in the process. For example, Varma et al. (2005) examine the role of interpersonal affect in PA. Molleman and Timmerman (2003) look at the specific challenges of conducting PA in R&D environments with less routine work and more interdependency among employees. Simmons (2002) investigates how a new philosophy of PA is required for it to be successful among knowledge workers in academic institutions. Wilson and Western (2001) look at how PA schemes can have negative impact on employee training and development if not carefully addressed. Dewberry (2001) looks at ethnicity and race bias in PA. Vest et al. (2000) examine the factors that determine employees' beliefs that merit pay increase is a biased decision or is related to real performance. Raymark et al. (1999) investigate the information sources raters utilise and the pitfalls these hold. Fink and Longenecker (1998) attempt to identify the top ten rater skills necessary to effectively conduct formal PAs, and why organizations fail to conduct effective rater training with their managers. Finally, Cook (1995) conducts a broader investigation into rater bias, and finds that problems among others include politicking, impression management and undeserved reputation.

The rest of the non-theoretical articles include a variety of themes, from how to automate the process and get rid of human bias by introducing Computer Performance Monitoring (Bates and Holton 1995), to issues with keeping a link between strategic aims of the organisation and the targets of the individual (Cameron, 1993), and to tips on how to prepare for meeting the boss (Knippen and Green, 1995).

Besides the focus on optimisation, issue handling and bias, the most prominent denominator of the non-theoretical articles is that they are so many. One could argue that this is because the field of PA studies is very practice oriented in that so many studies are focusing on investigating specific features and issues of the process, with the purpose to generate ideas for process improvement in practice. However, the lack of theory generation can be seen as a problem because it leaves a fragmented view on PA rather than adding to a deeper and more reflected understanding of the process and its effects on the people affected by it.

Summary of theoretical review

By looking at the theoretical approach utilised in my sample of the last twenty years' research into PA a number of findings of interest to my own study appeared.

The non-theoretical category constituted no less than 64% of the articles. These articles were primarily focused on investigating specific issues and bias and to propose suggestions for improvement activities. The significant lack of theory building within PA studies is confirming the findings from previous literature review studies within the field.

For the 44 articles categorised as theoretical, the picture was very heterogeneous as no less than 27 different theoretical perspectives had been utilised. I argued that these could be ordered into a three-layer, meta-level categorisation according to whether they were affiliated with either psychological, social psychological or sociological explanatory factors. However, for the purpose of introducing the theoretical conversations going on in the literature, I kept loyal to the labelling made by the authors themselves.

The most popular was Justice Theory which was used in 9 out of the 44 articles, indicating the significance of the fairness and justice perspective which has been central to the critique of PA in practice.

Another key theoretical concept utilised is motivation, which under various headlines like expectancy theory and motivation theory plays a more or less direct role in many, if not most, articles. Often the focus is on the positive and negative effects on employees' intrinsic motivation, of extrinsic reward paradigms so often linked to PA schemes. Based on the research investigated here, the relation can be concluded to

be at least problematic, and it is no surprise that this is a key argument of proponents of the radical paradigm, as will be clear from the discussion below.

Another interesting theoretical approach is the quality movement around the concept of TQM. The TQM standpoint generally distance itself from the tendency of PA schemes to focus on individuals rather than systems and on quantity rather than quality. Again, some of the hardest critique against PA schemes comes from proponents of TQM.

Twenty-one out of the 27 theories utilised in the 44 articles coded as theoretical, are only used in one article. Many of them provide interesting new views on detailed aspects of PA from very different angles. But together they also add to a picture of a very heterogeneous field which could, as it seems, benefit from more fundamental theoretical clarity around the core question of what the PA process means to the people who work every day within it.

2.7 The Odd Articles: The Radical Perspective

Out of the 125 articles in my literature review, 119 fall into the dominant paradigm in the sense that except a few all-embracing articles, they are concerned with critically analysing aspects and issues of PA schemes, but without rejecting the idea of PA as an acceptable approach altogether. Above I have given numerous examples of this and will not dwell more on it. Instead, I will below take a closer look at the articles where the authors insist on entirely distancing themselves from the dominant paradigm. These are the articles belonging to the radical perspective.

Out of the 125 articles, six are radical in their approach as they argue for abolishing PAs altogether. One of the six articles contrasts the two perspectives in an interesting literature review. I have already above shown how Waite and Stites-Doe (2000) empirically examined employees' reactions to an abrupt abolition of a PA and merit pay scheme in favour of a TQM implementation. In the study it was found that replacing a traditional PA and merit pay system with a non-appraisal feedback and coaching scheme combined with a blanket pay increase, enhanced the employees' perceptions of pay fairness, pay satisfaction and job satisfaction. The authors drew a very strong conclusion from the study, stating that in a quality focused organisation,

the managers "...must recognize that their role has changed from that of judge to that of coach" (Waite and Stites-Doe, 2000: 202).

In her study of the introduction of a PA scheme to chairs of university departments, Townley (1999) examines why chairs were so strongly against the scheme. The author found that there was a complete mismatch between what people found was the management rationale behind the scheme as opposed to the management rationale they were operating under in the university world. A somewhat lengthy quote is required at this stage because it so nicely describes the perception of PA schemes seen through the lenses of the radical perspective:

Although PA may operate in a range of ways, the underlying model is technocratic. It offers a disengaged, formal rationality involving the identification, specification and measurement of 'work', a unit of work, and the components of work - contract, workloads, output, performance, etc. Management is presented with a knowable, standardized, empirically verifiable reality, ordered through universal and abstract regulations that attempt to clarify the relationship between inputs and outputs. Its operation requires no particular type of social setting. It portrays management as a task that can use the same standardized, formalized procedures in an inherently substitutable manner. (Townley, 1999: 298)

The university chairs were remarkably aligned in their resistance to the scheme. They could not see how such a scheme could relate to their daily work where it was hard to specify what even the term "work" meant, as responsibility was socially defined rather than given by management. Neither could they define concrete, desired outputs in the form of performance targets that in any meaningful way captured the very independent mode of working characteristic of the employees. And they found it impossible to set up effective and fair measuring mechanisms.

Despite the initial resistance in the case organisation, PA has today become a widespread management tool also in universities. But according to Townley, this has happened at a high cost:

They [PA schemes] have become normal. But with them comes subtle shifts in the ability to argue about the goods of the tradition. The debate becomes circumscribed, individualized and privatized. Individuals are accountable because they complete annual appraisal forms; an organization is accountable because it has such procedures in place. Through such mechanisms the opportunity for collective engagement and debate about the goods of the tradition and their sustainability is inhibited and diminished. There is a loss of vitality. (Townley, 1999: 302)

So, according to Townley, the appraisal schemes bring with them a bureaucratized and individualised, compliance fulfilling management process that has negative

impact on the more practically grounded and socially oriented coordination and exchange paradigm it has replaced.

Bowen (1995) raises the question why society relies on ranking and grading from the first day of school, through the educational system, into the working environment. He argues that PA schemes are creating winners and losers in an organisation to the detriment of overall organisational performance and to general human wellbeing. According to Bowen, ranking schemes rely heavily on a strong belief that extrinsic motivation is what drives human beings, while so much research shows that it is in fact intrinsic motivation that holds a lasting and good effect on performance. But not only are ranking schemes relying on extrinsic motivation - according to Bowen they are in fact destroying intrinsic motivation. Hence, Bowen suggests that ranking should be replaced by a much stronger focus on strengthening co-operation and shared responsibility.

Piercy (1994) raises a more instrumental critique against PAs. He accuses management of repeatedly substituting bureaucracy and complex reporting and control systems for the things that really matter to competitive success. According to Piercy, PA schemes are excellent examples of this, since they are so bureaucratised and focused on what he terms "administrative neatness" that they completely lose any connection to the real strategic issues of the organisation.

Fisher (1994) comes from a consultant background with years of experience in assisting with implementing appraisal schemes in different organisations. In his article, he is presenting some thoughts on how different schemes can be categorised and what the advantages and pitfalls are for each category. The author is not explicitly departing from PA as a useful management tool as such. But since some of his categories are in practice quite radically departing from the mainstream PA concept of manager-rating-employee, I have included it as an example of a radical approach to PA. Fisher classifies performance schemes on two dimensions. First, he distinguishes between whether they are focused on accountability or development. Second, whether the appraisal is to be carried out by someone hierarchically superior to the appraisee or by a peer. Since the mainstream PA scheme would clearly fall into the accountability/hierarchical combination, and the purely developmental interview would normally be seen as an activity that can be maintained in parallel

with the PA process, let us take a look at what Fisher has to say about schemes where it is up to peers to conduct the review:

The purpose of this type of appraisal is to provide feedback to the appraisee and to help her or him to plan their future self-development. The focus is primarily on the needs of the individual with only a broad and generalized regard being given to the needs of the organization. (Fischer, 1994: 37)

Fisher did not find any really good empirical examples of organisations with performance evaluation by peers. However, he mentions academic peer review as something close to this, even in the form of a yearly review by the head of department, if these were able to create an atmosphere where they,

...were acting in the role of academic peer rather than as line manager... and if the interview was not seen as a management exercise but as a personal opportunity to reflect and plan. (Fischer, 1994: 37)

Clearly, this is something completely different from the mainstream performance review where there are clear roles between rater and ratee, and where there is a singular, objectified outcome in the form of a ranking on a numeric scale.

To round up the presentation of the articles coded as belonging to the radical perspective, let us now turn to Rasch (2004) who also places himself in the radical perspective and who through a literature review gives an account of how the arguments of proponents and opponents of PA schemes contrast. Generally, Rasch finds that proponents of PAs blame implementation flaws as the reason for the widespread dissatisfaction of expectations to PA implementations. Proponents of PA would claim that the process would work if only it was properly used. However, opposing this view are the critics of PA who,

...view the PA process as divisive within the organization, and thus counterproductive to organizational development. They state that the use of ratings in an employee PA establishes arbitrary divisions among employees. Furthermore, they contend that intrinsic rewards are far more effective as a motivator of employees. (Rasch, 2004: 407)

Rasch quotes Deming's (1986) proposal that 95% of the employees perform well but that PA schemes are basically designed to address issues that are only relevant for the remaining 5% of the employees. As a logical consequence of this, Rasch suggests that the current practice of PA should be substituted with another scheme consisting of three elements. First, the scheme should build on a focus on individual growth and development through dialogue and without ranking. Second, the scheme

should include a PA system that deals with the 5% special cases, so that it allows managers to dedicate their evaluation and close target follow up activities on those employees who actually need that. And third, the scheme should focus on developing a continuous improvement mindset in management so that managers can become good coaches to the employees, rather than performance raters.

Summary of the radical perspective review

Although authors within the radical paradigm all reject the idea that PA schemes will stand a chance to add net value to organisations, they do it by emphasizing different aspects as critical. As we have seen, Coens and Jenkins (2002) emphasize false assumptions while Kohn (1999) emphasizes the criticality of the more narrow perspective of dependence on extrinsic rewards. The authors of the six radical articles in my literature review reached the same conclusion but from other angles. From Waite and Stites-Doe's (2000) TQM position to Townley's (1999) practice argument, through Bowen's (1995) critique of society's dependency on ranking and Piercy's (1994) critique of the technocratic and bureaucratic focus of PA schemes and finally to Fisher's vision of a more development and cooperation oriented manager role, rather than the currently dominating evaluator role.

2.8 Performance Appraisal - What are the Problems?

In my literature review I have so far analysed a number of monographs and 125 journal articles to create an overview of the state of research into PA. While the monographs were used to create an overview of the subject and identify a dichotomy of attitudes, the analysis of the 125 articles was based on three different dimensions of the approach of the articles: the empirical, methodological and theoretical approach. I supplemented this by taking a separate look at the six articles belonging to the radical perspective.

Through the chapter, I have given numerous examples of the kinds of questions the authors have dealt with and the findings they have presented. It may at this stage be beneficial to take a step back from methodology and attitude, and try to summarize the content of the author's findings as they have been presented above. I will do this under the headline question 'What are the problems with PA?'

The literature review meta-study of the past 20 years research published in academic journals has shown that only a small minority of today's scholars subscribe to the radical paradigm. The majority of journal articles on PA are of a more pragmatic nature focusing on particular practical and theoretical problems in PA and possible improvements. Thus, my review uncovered more than 50 different practice problems that have been analysed in the literature over the past 20 years.

In the literature, some problems in PA are described as specific to the target setting sub-process, e.g. that appraisal schemes remove focus from adding real value to reaching targets (Kohn, 1999), targets are almost always only indirect represents for real organisational value (Austin, 1996), targets become input (effort) rather than output (result) oriented (Soltani et al., 2004), targets are either too ambitious or too easy and in both cases de-motivating (DeNisi and Pritchard, 2006), targets do not incorporate external factors (Furnham, 2005) and targets are quantity rather than quality focused (Deadrick and Gardner, 1999).

Other problems are closer connected to the on-going feedback and midyear appraisal sub-process of PA. For example, the informal feedback dialogue between manager and employee is thwarted by appraisal schemes (Waite and Stites-Doe, 2000), feedback is manager centric (Coens and Jenkins, 2002), there is a weak link between effort and feedback (Langan-Fox et al., 1998), there is a general lack of feedback outside the formally required meetings (Prowse and Prowse, 2009), managers are reluctant to give negative feedback (Yariv, 2006) and appraisal schemes Install managers as judges rather than coaches (Waite and Stites-Doe, 2000).

Another category of problems in PA concerns the sub-process of finally appraising and rating employees. Judged on the number of contributions, this sub-process of PA seems to have seen particular attention from scholars, who have worked with issues like e.g. that rating processes are subjective (Armstrong, 2008), ratings lack comparability (Coens and Jenkins, 2002), appraisal schemes are in practice counterproductive in dismissals (Coens and Jenkins, 2002), PA schemes build on an extrinsic reward philosophy which is undermining intrinsic motivation (Kohn, 1999), measurement is often more difficult than one would expect (Austin, 1996), there is an inconsequent relation between performance, appraisal and reward (DeNisi and Pritchard, 2006), there are serious issues with perceptions of justice and fairness

(Thurston and McNall, 2010), relative rating schemes are generally perceived as unfair (Roch et al., 2007), there are differences in employee engagement in the PA process which affects rating but is unrelated to performance (Knippen and Green, 1995), disappointing rating decreases motivation and increases employees' intention to quit (Poon, 2004), focus of PA schemes is on formality rather than utility (Deadrick and Gardner, 1999), and PA schemes (and research) ignores or downplays within-person variability (Murphy 2008). Further, many scholars have investigated particular biases in the rating process, e.g. that manager's own appraisal affects employee appraisals (Latham et al., 2008), attribution bias (Cardy et al., 1998), gender, race and age bias (Stark and Poppler, 2009), how managers falsely tend to assume performance is stable over time (Lee and Son, 1998), interpersonal affect bias (Varma et al., 2005), manipulation, impression management and politics (Cook, 1995), undeserved reputation bias (Cook, 1995), memory bias (Coens and Jenkins, 2002), central tendency of ratings bias (Bacal, 1999), and bias due to managers' provision of self-serving evaluations (Coens and Jenkins, 2002).

Finally, there are scholars who have investigated problems in PA that are of a more cross-process nature, e.g. that appraisal schemes serve conflicting objectives and are built on a one-size-fits-all perception of the organisation (Coens and Jenkins, 2002), PA processes are essentially top-down processes (Coens and Jenkins, 2002), the entire PA process is founded on incomplete information (Raymark et al., 1999), PA undermines intrinsic motivation (Kohn, 1999), focus in PA is on individuals rather than systems and processes (Soltani et al., 2004), PA favours quantity over quality (Deadrick and Gardner, 1999), PA installs competition at risk of deteriorating cooperation (Kohn, 1999), PA discourages innovation and risk taking (Kohn, 1999), PA requires significant resources, particularly from management (Coens and Jenkins, 2002), PA gets heavily politicised (Poon, 2004), cultural differences makes comparison and global use questionable (Woods, 2003), there are typically multiple implementation flaws (Rasch, 2004), PA is based on extrinsic motivation philosophy and has questionable outcome (Bowen, 1995), PA is bureaucratic and control focused (Piercy, 1994), PA is stressful and mental resource requiring for managers and employees (Brown and Benson, 2003), managers are too lenient and focus on avoiding conflicts and therefore undermine the trustworthiness of targets and ratings (Coens and Jenkins, 2002), PA is characterised by a lack of transparency (DeNisi

and Pritchard, 2006) and last but not least, it is still unproven that PA schemes ultimately or at all increase organisational performance (Kohn, 1999).

I have illustrated this process dimension in Figure 2-5 below, where I have listed all the problems referenced through the analysis conducted in this chapter. For the purpose of simplicity, I have ordered the problems by three main process stages that are characteristic for almost any PA scheme: Target setting, feedback & control, and calibration, appraisal & reward.

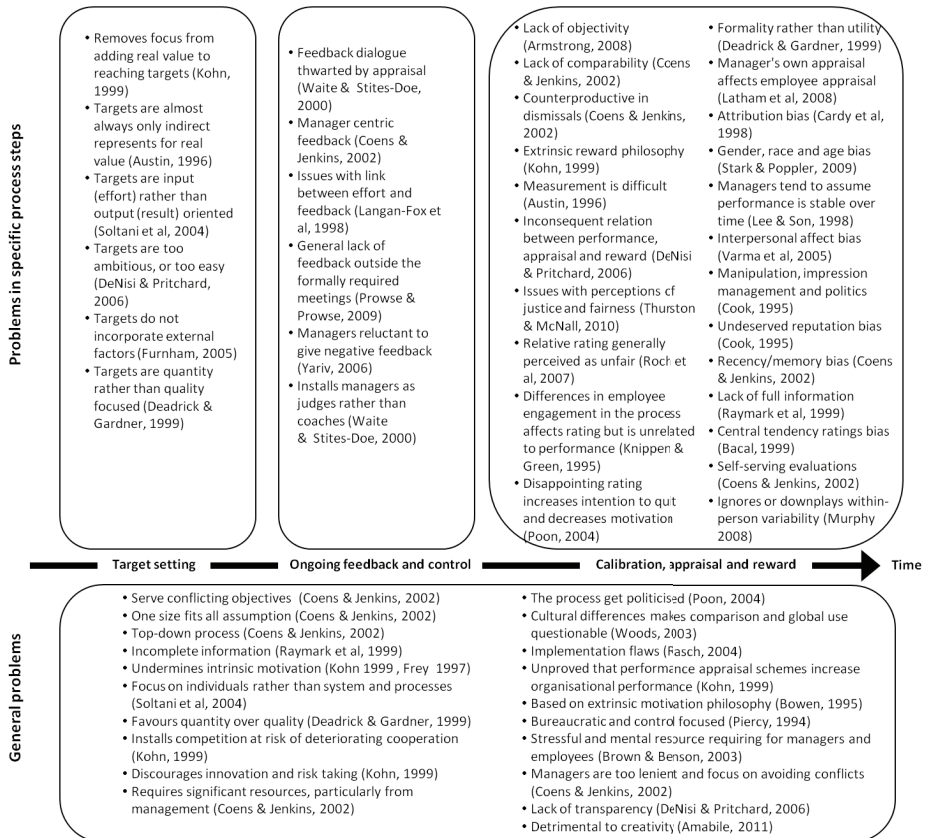


Figure 2-5 Overview of PA problems identified in the literature review

The top three boxes in Figure 2-5 summarise the problems primarily connected to one stage of the PA process, as they have been identified and analysed by the authors included in my literature review. The lower box summarises the more

general, cross-stage problems handled by the authors. To avoid making the illustration too complex, I have only included one reference into the figure for most problems. However, in reality almost all the problems are dealt with by more authors in scope for my literature review.

Of course, not all authors would agree about the seriousness or possibly even the relevance of all problems presented here. Some problems have been presented by the authors as contextual - for example Austin (1996) made a point on the particular issues with measuring performance in a knowledge worker context. On the other hand, it is highly unlikely that the list of problems presented here is anything near exhaustive. If one used broader search criteria than I have done in this study, one would likely find more problems described out there. The purpose of this study was not to draw 'the full picture' but rather to provide a comprehensive taste of the challenges and complexities one faces when embarking on studying or implementing PA schemes in organisations. Looking at Figure 2-5, it should be quite clear that installing a successful PA scheme is not a trivial task. It is a resource intensive endeavour with many real and potential problems and risks. Realising this up front is probably a good place to start when embarking on an implementation of a PA scheme in an organisation.

One could ask why I have been so focused on the problem side of PA. Would it not be fairer to also dig into all the good things that come out of the process? Partly, this prioritisation is a logical consequence of the nature of my analysis. It is hard to find any academic journal articles that are not problem focused, so a literature review will naturally cover a range of problems. Also, the problem side is an interesting contrast to the obvious fact that my case organisation, together with most other large Western private organisations, at some stage has decided that implementing a PA scheme was a worthwhile investment. I will soon give a more thorough introduction to the rationale behind the implementation and the details of the PA process in the case organisation. At this stage it will suffice to reiterate that there is an interesting tension, a paradox, between the prevalence of performance management schemes and the huge number of problems, risks and costs identified by the last 20 year's research into the subject. This is the tension that leads an author like Pickett (2003), who is by the way himself in favour of conducting PA, to conclude that in too many organisations the process is best characterised as 'the annual fiasco'. It is also this

tension that made me so interested in the question of local adaption and meaning creation. How do managers and employees create meaning from being part in a process so prevalent in business and so strongly enforced by their organisation, while at the same time being faced with some or all the problems listed in Figure 2-5?

2.9 Summary of Literature Review

Based on a literature review of 125 journal articles from 1990-2010 on PA I have documented tendencies and identified possible research gaps within the field from four angles.

First, I examined the empirical focus of my sample of articles, I found that there is an overweight of public sector studies as opposed to private sector studies, when compared to the share of the full employment base of the two sectors. Similarly, I found that there is a huge overweight of studies within a Western context with almost two thirds of all the empirical studies in my sample being conducted on populations from three countries: US, UK and Canada. Also, I found a significant overweight of 85% of all empirical studies being conducted on white collar populations. Only 8% of the articles were studying more than one workforce, but none of these compared experiences between workforces. I concluded that this lack of comparative analysis across workforces is a significant gap if we want to improve our understanding of how experiences with PA can differ intra-organisationally.

Second, I examined the methodological approaches utilised in my sample of articles. Quantitative studies were significantly more frequent than any other methodologies and there were three times as many quantitative studies than qualitative studies in my article sample. Also, I found that the group of quantitative studies were generally characterised by narrow and unambiguous research questions, often formulated as one or more hypothesis testable against closed question survey responses from a representative group of employees. On the other hand, far fewer articles dealt with more open ended, exploratory questions, and these typically utilised a qualitative approach based on deep dive interviews or a case analysis approach. Only two studies had a longitudinal design and only four of the articles were using a mixed-methods approach.

Third, I examined the theoretical approach utilised in my sample. Interestingly, the non-theoretical category constituted no less than 64% of the articles, which were generally concerned with investigating specific issues and bias within PA and proposed suggestions for improvement activities. For the articles categorised as theoretical, the picture was very heterogeneous as no less than 27 different theoretical perspectives had been utilised. The most popular was Justice Theory which was used in 9 out of the 44 articles, indicating the significance of the fairness and justice perspective which has been central to the critique of PA in practice. Also, the concept of motivation played a central role in many articles, often with focus on effects on employees' intrinsic motivation of the extrinsic reward paradigms so often linked to PA schemes. Another theoretical category was TQM which was generally seen to distance itself from the tendency of PA schemes to focus on individuals rather than systems and on quantity rather than quality. Finally, it was found that 21 out of the 27 theories utilised in the 44 articles coded as theoretical, were only used in one article. However, we also saw how theories only used once were in some cases closely related to each other.

Fourth, I found that out of the 125 articles only six belonged to what I have called the Radical Paradigm. Of course, many authors are critical towards many or a few elements of PA schemes, but only six came to the conclusion that abolishing the scheme rather than improving it would be the recommendable way forward.

Fifth, I summarised all the problems identified and handled by the authors included in the literature review and ordered these in accordance with a simple three-stage model of the PA process. This overview illustrated how the past 20 years' research into PA schemes has identified many costs, problems and risks connected to PA. In my mind it is justified to conclude that according to the state of academic studies into the field, there is a clear tension between the prevalence of such schemes and the reality managers and employees meet when fulfilling the process in practice. This seems to be a good place to look for processes of meaning creation.

It is to this large and on-going conversation about problems in PA I aim to contribute by shedding light on meaning scheme challenges, local adaption and meaning creation around problems in PA. One important observation when considering the above mentioned paradigmatic dichotomy between the dominant and the radical perspectives is that it is a scholarly rather than a practice dichotomy. This may seem

surprising when so many problems have been identified in PA over the past many years and so few have seen significant improvements. In practice, numerous PA process improvement and optimisation initiatives are continuously being kicked off in public and private organisations but cases of organisations abandoning PA schemes are rare.¹⁵ The lack of a practice dichotomy is linked to the degree of institutionalisation in the field. But it may also have to do with another factor; so far it has been very limited what scholars and practitioners have been able to offer as an alternative to PA. In his book "The Strange non-death of Neoliberalism" Crouch (2011) argues that the viability of neo-liberalism is to some extent caused, not by the virtues of neo-liberalism itself, but by the lack of a better alternative. In parallel we may say that the lack of tangible alternatives to PA schemes to support organisations' requirements for differentiation of talent makes it difficult to break the iron cage of the PA institution. Scholarly proponents of the radical paradigm may argue that people's engagement and motivation as well as organisational performance will increase by simply abandoning PA schemes even without putting anything else in its place. But that suggestion does little to help organisations make systematic differentiation of employees needed for reward, talent identification, promotion and demotion processes, etc. Instead, organisations continue to struggle with seemingly endless improvement initiatives to cushion the negative side effects of PA. These initiatives generally seem to bring very little tangible improvement. Measured on the development in the literature the problems in PA are persistent over time although different problems dominate in different settings.

However, this does not necessarily mean that the PA improvement initiatives are meaningless to the organisations that invest in them. On the contrary, even though the initiatives may not bring much tangible improvement in PA schemes' contribution to organisational performance, they still may bring improvements in peoples' belief in the schemes. Such improvement initiatives may be fully meaningful because they contribute to the rationalisation of the system. We may say that if it is perceived as hopeless to fundamentally change or even dismiss PA schemes because: a) PA has become so heavily institutionalised; and b) we have nothing to offer in its place, it is in fact quite meaningful for individuals and organisations to avoid a fundamental discussion of their PA schemes.

¹⁵ My literature review uncovered only one such example (Waite and Stites-Doe, 2000).

2.10 The Gap in the Literature: Directions for the Study

Besides the empirical, methodological and theoretical gaps described above, the current literature can for the purpose of simplicity be described as two-dimensional. One dimension is the paradigmatic question of improvement vs. rejection, represented by the dominant and the radical paradigms. The other dimension is the problem focus dimension concerning which problems scholars investigate and whether they are focusing on particular sub-processes, i.e. target setting, feedback or rating and calibration, or on more general cross-process problems.

The current literature largely focuses on questions around "what is wrong with it?" and "how could it be done better?" These kinds of questions are usually framed within the dominant paradigm. A few contributions from the radical perspective are concerned with "why we should abandon it!" Very few contributions are concerned with micro level type of questions like "how do people make meaning of it?" Of course, these different types of questions are not unrelated – they all can contribute to shed light on PA although from different angles. Therefore, neither are some of the questions necessarily more relevant to ask than others. But there is a clear imbalance so that what could be called the "improvement" questions dominate in the literature, while questions regarding meaning creation by and large are missing.

Among the 125 articles I took under closer examination in my literature review only Wright and Cheung (2007) dealt specifically with meaning creation at the micro level although not from a sensemaking perspective but from a Personal Construct Theory position. Wright and Cheung concluded that,

Not much is known about how organizational actors (both raters and ratees), party to the appraisal process, construe and make sense of their appraisal experiences within the bigger framework of the performance management system. Research of this kind in the past, have purely focused on the rater (per se) in their decision making role when making judgements about employee performance.

Wright and Cheung statistically investigated how managers see, interpret and make sense of their PA experiences. The authors did this through the use of a cognitive mapping methodology inspired by clinical psychology. However, their study focused on a limited set of general and high level, pre-defined categories of issues in PA and only dealt with the view of managers. The study did not elaborate on how managers' and employees' meaning creation in PA interrelated. Neither did the study deal with

the question of how a better understanding of this on-going complex of interrelated meaning creation, meaning scheme challenge and meaning establishment in turn could potentially contribute to our understanding of the problems in PA. In fact, I have searched but not been able to find any academic articles except Wright and Cheung's that specifically deal with PA from a meaning creation or sensemaking perspective. This indeed is an under-researched phenomenon.¹⁶ Thus, at the theoretical level my study will contribute to our understanding of the way individual meaning creation and PA practices mutually interrelate, and the environmental factors framing this interrelation. But at the broader level my study will also contribute to the literature on PA by drawing on a non-rational perspective on the understanding of well-known problems in PA, thereby deepening our theoretical conceptualisation of the same problems.

Outside the scope of my journal articles search, I have also come across a couple of doctoral dissertations that deal with PA and meaning creation, although from different angles. Mjoen (2011) in his PhD thesis "The HRM-Performance Trap" is investigating the corporate purpose and approach to PA, to analyse how humans are conceptualised as resources in what he calls "the HRM ideology". In this, Mjoen is more interested in how corporate stakeholders are affected by trends in HRM ideology, whereas I focus specifically on the dynamics at the level of managers and employees across workforces. As such, it is beyond the purpose of my study to investigate whether it was HR that invented PA as an institution in corporate life or whether HR was just riding on a wave set in motion by other forces.

Kane (2010) in his thesis "Performance Talk" examines "measurement as one representational form out of many others, for example text." The author empirically does this by conducting thirty interviews with senior managers to uncover various representational forms of measurements and to uncover their links to perceived measurement dysfunction. Also this study is quite far from my own investigation by still perceiving PA schemes as integrated and aligned across organisations, where as

¹⁶ A full text search on the academic database *Business Source Direct* on the word combination "Performance Appraisal" and "sensemaking" gave no hits. A search on the word combination "Performance Appraisal" and "meaning" gave one hit; Wright and Cheung (2007). If "Performance Appraisal" in the two searches is substituted by the related term "performance management", two more articles can be added; One about performance management implementation in primary schools in UK and one about institutional path dependency in Swedish governmental performance management schemes. None of them focus on meaning creation processes at the individual level.

I focus on identifying patterns of differences across workforces. And Kane's study focuses only on the view of senior managers, while I focus on the perceptions of employees and first line managers.

I have already introduced what I call the paradox in PA; that there is an important tension between the institutionalisation and prevalence of PA schemes in large Western organisations today and the number and magnitude of problems faced by managers and employees performing the process, as indicated by the literature review above. The issues and problems in PA processes are numerous and they appear to be persistently present despite decades of practical experience with PA processes and hundreds of research publications into the subject of what is wrong with it and how we should improve it. I propose that a first step to move forward from this standstill would be to deepen our understanding of meaning processes within PA. Before we have a thorough understanding of patterns of meaning creation in PA, we will have difficulties with understanding the nature of the problems in PA and we will therefore also struggle to identify the best solution alternatives to problems in PA.

In short, practitioners face a multitude of problems in PA which seem to be persistent over time despite years of attempts to solve them, both in practice and in scholarly research. This standstill is caused not by lack of technical knowledge about how to design a PA process, but by two core and related issues: a) Lack of attention to and understanding of organizational practice as it is affected by patterns of actors' practical meaning creation; and b) A common misconception of PA as being generally rational and de-contextualized and thus a candidate for standardised process design (Armstrong & Baron, 2005. Armstrong, 2006. Bacal, 1999 & 2004. Grote, 2005. Lazear and Gibbs, 2009). PA may in fact be better understood as highly situated and dependent on individual meaning creation within frames set by actors' personal, non-rational interpretations and experiences as well as macro level factors like organisational culture and subcultures, and institutions of management practices.

Meaning scheme challenges in PA can come in many forms. They can be challenges to peoples' meaning creation regarding themselves (identity related challenges), their work (task related challenges), their intra-organisational relations (challenges to their relations to managers and colleagues), their organisational perception (challenges to perceived culture and affiliation) and to a range of other meaning creation processes. The important object for this study is that there is a perceived meaning scheme

2. The PA Conversation - a Literature Review

challenge related to problems in PA and that this challenge is countered by actions and sensemaking processes at the individual level through enactment of local workforce environments.

3 Methodology and Data

In this chapter I will present the strategy for establishing empirical data as it was laid out in my research protocol, as well as reflections and learning made during the data establishment process.

The term "establish data" is purposefully constructionist. In a qualitative data analysis context it can be somewhat misleading to use the more common label "data collection", as if the data is out there in an unambiguous format, just waiting to be collected. Participant observation and qualitative interviews both include intensive interaction between the researcher and the people forming the empirical objects. They are both methods where the researcher cannot avoid having an effect on the object. Rather, they are methods where bias avoidance is not fully realisable and in fact often not even desirable, since the common interaction also invokes reflections on the object's side thereby unveiling insights previously unavailable. Thus, the researcher must reflect on the impact he has on the object and rigorously consider how this affects the findings.

My study took place with departure in a detailed research protocol. It is important to emphasize that all through the data establishment and analysis phases the protocol was a "living document" to reflect that a general research design principle has been flexibility. The protocol was updated in correspondence with changes in propositions, methodology, procedures etc. that - as would be expected in a qualitative, exploratory design - occurred during the research process.

Below, I will first introduce the two data establishment methods that played the least significant role in the study; participant observation and archival records and documentation. Then I will progress to presentation of my primary data establishment source: Qualitative, semi-structured interviews with managers and employees in the case organisation. From this, I will turn to presenting the details of the coding process including how I have taken advantage of digital coding software. Lastly, I will present and discuss the practical procedures used for conducting research and handling documentation.

The fact that qualitative methods are always to some degree impacted by the researcher does not contradict the aspiration to aim for the highest possible level of validity when doing research. In a single case study one way of increasing construct

validity is to utilise a triangulating strategy by analysing multiple sources of evidence and leveraging multiple data establishment methodologies (Yin, 2009).

3.1 Participant Observation

By being a manager within corporate HR I am close to the continuing development of the firm-wide conceptual PA solution. I am involved in the strategic development of the underlying IT tools and reporting capabilities that allow correlation analysis to other HR practices, e.g. bonus, salary increases and more. This all gave me a unique opportunity to undertake participant observations of the corporate dialogue and activities. In short, my privileged position as a manager in corporate HR gave me direct access to the PA conceptual machine room. It also gave me the ability to “perceive reality from the viewpoint of someone ‘inside’ the case study rather than external to it” (Yin, 2009: 112). Understanding the corporate politics and lingo improves the ability to conduct relevant analysis of data established.

The cost of being a participant observer obviously is the risk of introducing personal bias. This concern was from the start inherent in my work and only became more so along the way, as my understanding of the phenomenon grew and as impressions unavoidably transformed into opinions. As a manager within corporate HR directly involved in the PA process development I of course had my own professional interests and agenda and it was a clear risk that this would affect my ability to perform a balanced analysis based on participant observation alone. Thus, participant observation was from the beginning only thought of to be one among other data establishment methods and it was always planned to be inferior in importance to the much more extended qualitative interviewing approach described below. As it turned out, participant observation became in fact even more limited in its use than originally planned, due to ethical considerations but also because the research question through its refinement process developed in a way where participant observation was a less valuable data establishment source. Below, I will elaborate on this development while introducing the participant observation approach in more detail as it was described in the protocol.

3. Methodology and Data

Drawing on Emerson et al. my aim was to produce field notes in an on-going, structured manner and on the basis of these notes, “transform[ing] observation and experience into inspectable texts” (Emerson et al., 1995: xii).

I did this by producing a written record of relevant observations made on an on-going basis. By being in a position of what Emerson et al. call immersion (Emerson et al., 1995: 4), I had already walked a significant part of the path that participant observing researchers normally need to go since they often approach a setting not already familiar to them. My position also deviates from the classical participant observer role in another way: Emerson et al. describe how the participant observer never becomes a member in the same sense that those ‘naturally’ in the setting”. (Emerson et al., 1995: 4)

Being a member was exactly the position I was in. How did this position affect my field notes? The first thing to realise is that field notes are never just a list of objective facts. They constitute narratives that include not only the event in itself but also the interpretation of the event by the observer. Self-reflection is a requirement to be able to understand how one’s own interpretation affects the accounts and how different or even contrasting interpretations could be possible. This was the approach I decided to follow and it was my aim to reflect on my own role as a participant and observer in all three steps to handle bias rather than aiming for the impossible mission of avoiding bias.

Scholars agree that writing field notes is an activity that must be conducted immediately after or as close in time as possible to the event covered (Yin, 2009 and Emerson et al., 1995). Further, I had an advantage over other types of ethnographic observers who often find it awkward to do jotting during the actual observed interaction. This was a perfectly normal activity for me, as I participated in meetings and workshops regarding the PA process. Thus, I developed a simple template to hold my field notes as they unfolded from my jottings. The template was a two-column table with my field notes in the first column. Each event constituted a row in the table. Methodologically, I took Emerson et al.’s advice and aimed to keep the field notes in the first column as concrete as possible without “making statements characterizing what people do that rely on generalisations” (Emerson et al., 1995: 32).

3. Methodology and Data

In the second column I stated my initial analysis of the notes. This could be concerns, ideas suited for generalisation, methodological reflections, etc. I found such reflections valuable as they are not only a way to bring forward further analytical thoughts early in the process of the study but also because they constitute the first steps towards theorising or relating to theoretical frameworks. As such, they are elements in the analytical coding process which is a necessary step similar to when data is established through qualitative interviews as described below.

So, what was I looking for while conducting participant observation? Emerson et al. (1995: 146) among others suggests focusing on "What are people doing?... how exactly do they do this?... what assumptions are they making?... why did I include them?"

These questions was a good starting point but as already indicated I added more analytical secondary order notes to capture my own reflections. These notes were not as such conclusions but rather ideas to be utilised in the later analysis.

From the start I did not have a very precise idea about what events I would cover with participant observation. Since I work full-time in the organisation and in periods spend significant parts of my time on work related to the PA process, particularly the development and further technological support of the process, it was clear that I could not cover everything in which I was involved. Instead, I initially had the idea that I would cover three main categories of activities by taking field notes.

First, I would cover my own experiences from being part of the PA process by being both a department manager for a group of employees, and an employee myself in relation to my manager. By being a middle manager I have of course been a practitioner of the PA process. As a manager I have had years of experience with setting targets, conducting midyear evaluations, performing yearly appraisals, participating in management team calibrations and having tough or happy appraisal conversations with employees. I have also gone through the same as an employee with the managers to whom I have reported over the years in the organisation. Second, I would take notes from work experiences related to the strategic development of the PA process, particularly the part I was myself heavily involved in. Third, I would generally take notes on an *ad hoc* basis whenever I heard or experienced things that I found important in relation to the study.

3. Methodology and Data

Since the participant observation method part of the study methodology was the one that changed the most during the course of the study in terms of purpose, extent and outcome, I will try to explain these changes below.

As it turned out, I chose to refrain from most of the systematic participant observation activities due to ethical and practical considerations. Evaluating the PA process conduct of my direct reports or my manager would in the last instance involve stepping out of the confidential dialogue and take the position as an outside observer “judging” the behaviour of not only myself but also my employees or manager, through my analysis of appraisal events of which I was myself part. Although this may be possible to do, I believe it would be very challenging to do so without compromising the trust and confidence required. A founding principle of my design was that I wanted to be open about what I was doing and thoroughly inform involved individuals up front. Doing participant observation without informing my employees or my manager was a no-go, both held against my design principles and because I found it unethical. But taking notes and conducting analysis with their consent would for both them and me be so awkward that the value of the notes would be at risk, and even worse that the relationships would be negatively affected.

Taking notes from the general development of the process and the above mentioned *ad hoc* notes, was less ethically challenging. Of course, getting participant consent would still be required had the note-taking involved quoting other people. But as this would generally not be required as the notes were merely my own, on-going notes on thoughts and reflections along the way, it was sufficient to inform people at a more general level, that I was conducting a study of PA with the organisation as the case object. Since I had access to so much background information, the outcome of my notes turned out to be more supplementary to the two other data establishment methods utilised through the study, rather than a means to break new ground. It became a structured way of reflecting on the things I was myself involved in at the process development macro level, rather than a source to sensemaking at the individual micro level.

Besides these ethical and practical considerations I also found that most of the knowledge that this part of the study design would bring could equally as well be derived from qualitative interviews with other pairs of managers and employees, where the outside observer role would not contrast with my own process participant

role. Particularly as the research question was refined and narrowed over the first year of the study, it became more and more clear that the critical data establishment process would be qualitative interviews, and that participant observation would play a less significant role.

Nevertheless, understanding the business background of implementing a global PA process in the first place, and the later and on-going rollout and development of the process, has been a major advantage during the study. I already had a good understanding of this before embarking on the study. After all, it was my own experiences with the process at the macro and micro level that made me interested in the subject in the first place. In that sense, my participant observation notes made me a more observant observer, and in this way added to my ability to see and reflect on events with relevance for my study.

Thus, although participant observation was from the start of my study a weighty methodological part of the protocol, over time it developed to be less prominent. However, it still was a fruitful way to constantly mirror my findings against my own experiences as they were uncovered through personal experience and observation.

3.2 Archival Records and Documentation

Personal experience from a six year long employment history within the case organisation at the start of the study together with participant observations gave me a solid foundation for asking questions. But being an insider in corporate HR also gave me another thing which was invaluable to my study - privileged access to internal documentation.

I initially had focus on using this access primarily for two purposes. I wanted to analyse archived material produced in connection with the corporate performance management project that ran from 2004 to 2006. This was the project that first implemented the new, global PA process in the organisation and later rolled out an IT solution to support it. This documentation included presentations, training material, communication material, steering committee meeting minutes, status reports, benefit realisation plans etc. The archival access also covered documentation of later initiatives to further rollout or improve the IT solution. The thorough study of all this material made it possible to draw a picture of what was intended to happen and why.

Second, archival sources were identified that helped me establish the frames within which sensemaking takes place. Within Corporate HR statistical experts over the years had gathered data through their analysis of correlations between appraisal ratings and rewards, promotions and more. It was on the basis of this data combined with my own participant observation that I became able to conclude that the case organisation's experience with PA was in so many ways congruent with established research on PA: the same macro level purpose, the same bias, the same micro level frustration. Thus, studying archival material and matching the findings to the literature review and the participant observation findings were necessary activities, not only to understand the generalised character of so many features and challenges of the PA scheme in the case organisation, but also to get to the final formulation of a research question with potential to generate new knowledge.

I analysed five main categories of archival material. First, documentation from the corporate project which in 2004-06 implemented a global, uniform PA process, named *wave I*, and subsequently implemented a global PA IT system, named *wave II*. This material included documentation of the original purpose and business case for the initiatives, steering committee material, project presentations, analysis documentation, communications etc. Second, documentation from the two later initiatives; *wave 3* that further rolled out the process and the IT solution globally, and *wave IV* that added process support in matrix organised parts of the organisation and further rolled out the IT solution to such units globally. Third, material from the 2010-11 initiative aiming to establish Global Minimum Benchmarks, i.e. a further standardisation of the practical conduct of PA across the global organisation. Fourth, documentation of statistical analyses of rating distributions and correlations to variables such as rewards, position in the organisation etc. These analyses were conducted by corporate HR in the period 2010-2011. Fifth, a data set covering all electronically captured performance appraisal ratings from the period 2006-2011, allowing linear regression analysis of correlations between ratings and elements like hierarchical position and retention ratio.

When analysing archival material it is important to observe a consistent source critique. The various PA projects were all managed by corporate HR with certain intentions and success criteria. Corporate agendas are always to some degree political in the sense that there are vested interests and power at play. Hence, I saw

it as important to understand the agenda of senior management and corporate HR when translating the material into statements about what went on and why.

3.3 Interviews

In all, I conducted 38 qualitative, longitudinal interviews. By longitudinal I mean a study design of multiple engagements over an extended period of time. I.e. a study design built on following fixed manager-employee dyads over a full year (corresponding to a full PA cycle) and interviewing them multiple times in connection to recurring PA events (target setting, midyear review and final appraisal). Of course, I am aware that even a one year performance cycle is still only a partial image, since employees through their work life in an organisation will go through multiple PA cycles that mutually affect each other, such that performance delivered in one year may affect targets set for the coming year and so forth. However, observing a full one-year PA cycle through multiple interviews with the same people is broadening the analysis significantly compared to most other studies dominated by the tendency to conduct analysis gathered at a single point in time, e.g. at the target setting or at the final appraisal only.

Below I will describe in some detail the approach of the interviews and the methodological lessons learned from them. I will start by explaining why I find a longitudinal interview approach such a fruitful way to investigate the PA process. Second, I will discuss how I approached a tricky question that challenges most researchers working in qualitative designs: Who to interview? This will include reflexions on the consequences to the study object and the study results, of my own intervention as researcher. Third, I will discuss the practical planning and conducting of the interviews. Fourth and finally, I will discuss the robustness of the study design.

Why a longitudinal design?

The literature review of 125 scientific journal articles regarding PA published over the last 20 years (1990-2010) showed an empirical-methodological emphasis on quantitative approaches. Only ten articles used a methodologically triangulating approach through case studies. But what was really striking was that only two out of the 125 studies used a longitudinal setup, although many authors suggested the likely value of conducting longitudinal research on PA in their 'direction for further

research' section (Korsgaard and Roberson, 1995. Salimäki and Jämsén. 2010. Lefkowitz, 2000). Thus, the lack of longitudinal research into PA seems more a consequence of lack of resources in terms of time and capacity, than a consequence of lack of longitudinal design benefits. In fact, the benefits of a qualitative, longitudinal research design are in my view significant. Besides the fact that longitudinal design in PA research is rare though recommended, I see at least three important reasons to aim for a longitudinal design when investigating meaning creation in a PA context.

First, the appraisal process itself runs over a full year with at least three formal events and many other opportunities for informal feedback events. Thus, to conduct research at any given single point in time will be at a high risk of missing both any overall pattern, but also of missing important fluctuations in attitude developments over time. Second, in a cross-sectional analysis it will be highly likely that one will either entirely miss out on, or get a distorted impression of, important events simply because of time lag between investigation and event. Third, if my proposition is correct that the field of PA is heavily institutionalised so that managers and employees think of the process as a given, this does not encourage reflexion on the process. The degree of institutionalisation must be expected to be a significant filter between what is at face value interpreted from the appraisal process and what would be interpreted through deeper reflexion. To raise the level of reflection would be almost impossible in the course of a single interview or survey research design. It takes time to build confidence between the interviewees and the researcher and to provoke reflexivity around the PA process.

Identification of interviewees

The interviewees should ideally be identified on the basis of their likely contribution with new insights or reflections around PA. Within such a design, the principles and process for strategic sampling was critical for the success of the study.

To identify appropriate candidates, a dialogue with HR Partners in Research & Development, Staffs, Production and Sales was conducted during November and December 2010.

I did not look for participants who were particularly high or low performers. Neither did I look for participants that were either strongly for or against the appraisal process design. Rather, I looked for participants who were first of all interested in participating

in my study and spending some time and effort on this and who were reflective and willing to talk openly about their personal experiences around the PA process.¹⁷ It was my hope that the participants would also themselves benefit from participation, through personal reflection and discussion with the researcher.

It was the initial intention to identify two pairs of managers and employees within each of the four workforces. The reason for looking for pairs of managers and employees is directly derived from my research question about the meaning creation of managers and employees. Differences between any random manager and employee without a direct reference line could be derived from the fact that their experienced performance events differed by nature. The opposite is true for pairs of managers and employees with direct reference lines. Here the manager and employee would at least theoretically observe and evaluate the same performance events and surely participate in the same performance feedback events with each other. Interviewing pairs of managers and employees may uncover varying interpretations and meaning creation processes around identical events.

The reason I was looking for two pairs of managers and employees from each workforce was a question of design robustness. If one pair fell out during the study, I would still have another pair left. However, as it turned out, it was only in the Staff area that I interviewed two pairs of managers and employees. In Sales, I interviewed one manager and two of her employees. In R&D and Production, I interviewed one paired manager and employee. The reason for this was partly practical, partly due to more principal issues with the study design, as will be evident from below.

In all four workforces, I approached the managers first. I assumed that the managers would be the ones with most time constraints and possible reservations about participating in a study of this duration, so I wanted to let the invitation to employees await the invite acceptance from the managers. In Sales, I was asked from the start if I could accept to interview only one manager since local sales managers were very busy and the sales HR partner wanted to disturb the business as little as possible. I

¹⁷ This is in line with Van de Ven's (2007) notion of Engaged Scholarship, i.e. that informants are chosen because they want to engage.

did not have anything against this and accepted the increased risk to my study design that this entailed.

In Production I got a yes from one manager but was held waiting by another. She was principally positive about participating, but was very busy in the start of the year and requested to wait to start until March. I accepted this, but as it later turned out she did not have time to participate even from March. At that stage it had become too late to look for a whole new pair of manager and employee in the production organisation, and I thus had to accept to go with only one pair.

In R&D, one of two selected managers interestingly rejected participation from a principal standpoint. I will quote the rejection mail at full length below, because it illustrates some important points.

R&D Manager: "As our business in our research department is very scientifically and project focused, it may be misleading to follow our PA interviews. Further I think we are far from representative to the degree that I think you could very easily misunderstand our PA (too many scientific issues and too many project/people related matters that it will take hours to get you updated on) and further this could be a too strong influence on our PA interviews. So I am sorry, I cannot prioritize this for now."

There are no less than three different arguments against participation hidden in this quote. The first argument is that it would be misleading to follow the PA process in R&D because the business is very scientifically and project focused. This argument is being reinforced by the argument that R&D is not representative and that I would easily misunderstand the process as it is conducted in R&D. It is hard to see why this should be an issue to my design. On the contrary, it was one of the key points in identifying interviewees from four workforces to see how meaning creation within a seemingly standardised process was differing across workforces.

The second argument against participation concerns the more common matter of time constraints. "...it will take hours to get you updated on" refers to the fact that the manager would not be able to justify the amount of time she thinks would be required to help me to the right level of knowledge. In all modesty, and based on the interviews I finally did conduct in R&D and elsewhere, I believe this objection does not hold. Of course, there would be targets of such R&D-specific scientific complexity that I would not understand the detailed content of it. But that would not do any harm to my study since I was interested in the meaning creation of the participants from

being part of the PA process. I was not particularly interested in the detailed content of each single target.

The third argument against participation is the most interesting and caused a lot of reflection on my side. The argument was, that "...this could be a too strong influence on our PA interviews." The manager was simply worried that by participating in such a study design, I would not be able to avoid affecting the object of study.¹⁸ As I see it there can be two concerns in this. One concern is that my intervention could somehow be harmful to the relation between the manager and the employee and the way they conducted the process activities. The other concern is the more scientific concern that the object of study would change from my intervention so that my results would not be representative for the "real world" where intervention had not taken place.

The other pair of R&D manager and employee did not have any reservations, so I ended up with just one pair in R&D. However, the rejecting R&D manager's arguments made me think a lot about what I was actually setting out to do. I asked myself the question what the most likely effect of my intervention would be on the participants. As I see it, the most likely effect would be that managers and employees would put more effort into the process than they would have done without my intervention. The rationale behind this is that they do not want to be caught being careless about their responsibilities in the process when someone from outside is watching them. They will most likely be better prepared, provide more examples in their feedback, be more reflected and well founded in their arguments for a given rating and so forth. So, if my intervention would have an effect on the manager-employee relation and the quality of their PA dialogue, I see it as more likely that this would be a positive effect. Hence, I did not see any ethical issue inherent in this. On the contrary, one could argue that the more I interviewed the better for the organisation.

The other concern about the scientific issue of object intervention is not so surprising when provided by a representative from R&D. R&D in the case organisation is strongly committed to conducting world-class statistically based scientific studies,

¹⁸ This critique is in the extreme case, where all observed behaviour change can be traced back to the researcher intervention, also known as the "Hawthorne effect" (French, 1950).

including double blinded, placebo based designs. In such a context, intervention effects are to be minimised and at best completely avoided. My study, on the other hand, is a qualitative study with the purpose to uncover - and provoke - reflections that reveals meaning creation processes of participants in the PA process. I do not claim to reach any statistical representation and my intervention can therefore by definition not be jeopardising such representation. Nevertheless, the feedback from the rejecting R&D manager made me very aware that I needed to capture reflections of my own role and influence through the study, and document these as part of the analysis.

Planning and conducting interviews

Planning for interviews is in part about handling all the practicalities, from preparing the necessary letters and documents to inviting interviewees well in time and handle all the unavoidable changes that happen through a study that runs over a full year. I will go into these elements in more detail when presenting the procedures I set up for research and documentation in a separate section below.

One thing that I found crucial for building the necessary trust between the interviewees and myself was that I followed what I would call a strategy of shared control. By this I mean that I always made it clear from the start of the interview that a transcript of the interview would be forwarded to the interviewee for review. I also made it clear that this review had two purposes. One purpose was to let the interviewee correct any errors from my side in transcribing the interview. There were examples where I had spelled a product name or a person's name incorrectly, and the interviewee could use the review opportunity to correct this. The other purpose was to let the interviewee think again about the discussion at the interview and if required correct any statements they made during the interview that they now felt they regretted.

The premise for this was that during an hour-long interview one can sometimes be carried away by the dialogue and maybe say something that one on closer reflection did not mean. I wanted to assure the interviewees that the study was not about searching for sensational expressions and then nail the interviewees to what they said. Rather, I wanted to understand their honest reflections and if they said something they did not mean it was ok to pull it back later. So I allowed the

interviewees to make material changes to the interview transcripts. Although I sensed that this was very comforting at least to some interviewees, interestingly there was not a single example of an interviewee suggesting any such material changes to the transcript through their review. Based on the input I got this was surely not because nothing controversial was said during the interviews. Rather, I found it was because the participating managers and employees had strong opinions that they felt were important to put forward, and maybe also because of lack of time or prioritisation of reading carefully through the transcripts and initiating a process of changing previous statements.

The interviews took place in meeting rooms at the premises of the organisation, at a location as convenient as possible for the interviewee in terms of transportation. There was not a lot of small talking around the interviews because most managers and employees are generally busy working and I did not want to take more than necessary of their time. The interviews lasted everywhere between half an hour and one and a half hour. Typically, the target setting and the final appraisal interviews were longer than the midyear review interviews. At target setting there was often a broader dialogue about the PA scheme in the organisation. At midyear, the interviews typically were more narrowly focused on the experience at the midyear review meeting. At final appraisal the broader concepts of calibration, fairness and impact made some of the interviews quite long. For all interviews I made sure that I did not have to close the interview due to time constraints on my side. However, in respect of the time of the interviewees, I made sure to cover the most critical parts of my interview template within the set time. As it turned out, this worked very well and I did not in any single case experience that an interview had to be prematurely closed due to time constraints.

Generally, interviews were held in a relaxed atmosphere. One significant benefit of the longitudinal design which I experienced was that the dialogue became more open and direct as the study progressed. This was particularly the case with some of the manager interviews where I at the first interview experienced a tendency of managers being politically correct in some of their expressions. In a highly successful organisation with a PA scheme strongly supported by top management, it can be rather anxiety-provoking to be open about concerns, particularly if they are of a more fundamental character. Managers are naturally most aware of the lingo around

3. Methodology and Data

corporate politics since they are much more involved in the tactics of making the PA scheme work in practice than are employees. Managers are also the ones that are expected to strictly represent corporate policies towards employees. Therefore it is only to be expected that managers are the most cautious when it comes to putting words on any critique against corporate schemes. However, as the interviews went on and we revisited the same subjects in the second and third interviews, some of the managers seemed to soften up and became more direct in their expressions.

One aspect that I experienced worked particularly well being a middle manager myself, was the relation this created to the managers I interviewed. We were in many ways in the same situation. I could easily relate to their experiences - good as bad - with PA. One could then think that my position as manager would on the other hand scare some of the employees from being open. However, this was not what I experienced. Though being a manager I am of course also, in relation to my own manager, an employee. So I found it as easy to relate to the employees' experiences as I did to the managers' experiences, and that created a foundation for a good dialogue and for building trust.

Even more fundamentally than being a manager, I was an insider to the organisation as such and I did nothing to hide this. On the contrary I decided from start that I would be very explicit about it. For example, I decided to use my company mail address for all correspondence with the interviewees so that they could easily identify who I was and where in the organisation I belonged. When my own position in relation to the PA process came up during interviews, I took time to explain my own role in the organisation and how this was in fact the original cause of my interest into the subject.

Being an insider was in many ways a huge advantage. I had a very good understanding of the PA process as it was intended from the corporate perspective. I also had a very good understanding of the corporate lingo including the ocean of abbreviations, nicknames, organisational unit names etc. that interviewees mentioned through the interviews. That saved us a lot of time that would have been required for explanation to any outside interviewer. That time could then be used for discussions at a more analytical and value-adding level.

On the other hand, being an insider who is also an active participant in the object of study at the same time contains some risks while conducting interviews. I experienced a couple of times that interviewees turned my questions around and wanted me to answer them myself, since "you are coming from HR, so you should know". In one case I also experienced an employee pulling me in as an explanatory co-factor. She mentioned that her manager was not well prepared for a performance review meeting, but at the same time excused this by referring to "you managers are of course also so very busy". In such cases I always tried to rephrase and reiterate my question in another way and to explain that it was not my perspective on things, but the interviewee's own perspective, that interested me. With very few exceptions I experienced that this worked well and that the advantages of being an insider far outweighed the disadvantages.

Robustness of the design

When doing longitudinal qualitative interviews one should always consider the design's robustness against organisational or other changes. It is likely that one or more of the interviewees leave their current position during a one-year period, e.g. due to promotion, dismissal, expatriation or other reasons. It was central to my design that the interviews concerned managers and employees who shared the same appraisal events. Thus, I defined a set of clear research design procedures to follow in case any changes should happen that affected my object of study.

The principles were as follows. First, if a manager left her position I would try to get the substituting manager involved in lieu. The disadvantage of this of course would be a break in continuity and rapport in the dialogue between the manager and her employee. On the other hand, from experience I knew that in many areas the average time in position for a manager was around three years, so there would be a good chance that such a swap would unveil insights that were actually very relevant and important in the real life and meaning creation processes of the new manager and the employee who gets a new manager. How would they cope with this lack of continuity and rapport?

Second, if it was an employee that left and this was to another internal position, I would try to follow the employee through the finalisation of the yearly appraisal cycle that the employee was in when leaving. From a research point of view this could

potentially provide interesting insights into what happens in the appraisal process for an employee that transfers internally.

Third, if an employee left the company as such, it would be interesting to see how the appraisal process was closed down during termination. Hence I would continue to follow the manager-employee relation until the employee was fully terminated from the position and the PA process finalised and closed.

Fourth and finally, if an employee left early in the one-year research period I would try to get a substituting employee into the research design.

Thus, the research outcome would not be unaffected in case a manager or employee changed position or left the organisation during the one-year longitudinal interview process. However, due to the fact that the research design included pairs of managers and employees, such changes could be handled to minimise the harm to the study and would even provide access to other interesting knowledge to the benefit of the study.

As it turned out, there were in fact no less than four out of six employees who got a new manager during the course of the study. In all four cases, it was internal transfers and by coincidence, in all cases the change happened after the second interview at midyear review. In line with the above design, I therefore invited the four new managers for the involved employees to participate in the study and luckily they all accepted and participated in the third interview round regarding final appraisal. Figure 3-1 below shows the distribution of interviews over the one year period with the six employees and nine managers.

Session	Interview 1 Target Setting	Interview 2 Midyear Review	Interview 3: Final appraisal
Workforce	Feb/Mar 2011	Sep/Oct 2011	Feb/Mar 2012
R&D	E1+M1	E1+M1	E1+M6
Staffs	E2+M2	E2+M2	E2+M2
	E3+M3	E3+M3	E3+M7
Production	E4+M4	E4+M4	E4+M8
Sales	E5+M5	E5+M5	E5+M5
	E6+M5	E6+M5	E6+M9

Figure 3-1: Distribution of participation in three interview rounds¹⁹

¹⁹ Abbreviations used: E=Employee, M=Manager

Figure 3-1 illustrates how the interview audience was stable over the first two interview rounds, while there were four changes at the third interview round. In R&D, the employee was transferred to another organisational unit under another manager. In Staffs, one of the managers swapped jobs with a manager colleague in a sister department. In Production, the employee continued to have the same job content, but the job was transferred to an operational rather than a project oriented unit. In Sales, one of the employees had a new manager due to a re-organisation. For all four affected employees, the new managers took over the responsibility to finalise the 2011 PA cycle, so that four out of six employees were finally appraised by another manager than the one with which they had set the performance targets and held midyear review.

Although my design was prepared for organisational changes, it came as a surprise to me that such a large proportion of my interviewees were affected by organisational changes over just one year. The average time a manager or an employee holds a position varies a lot across the organisation, but a change of manager in four out of six cases was a surprise to me. There are both methodological upsides and downsides of these dynamics which I see as inherent in longitudinal studies in dynamic organisations. As an upside, it gives me a good impression of how the PA process is affected by such changes. The downside is of course that the trust and common understanding built over the first two manager interview rounds would not be present to the same extent at the third interview, where the manager had changed.

3.4 Coding: Digital Analysis of Qualitative Interview Data in nVivo9

In all, I conducted 34 interviews with pairs of managers and employees and 4 with representatives from corporate HR and so I ended up with 38 interview transcripts, each at a length of 8-15 pages. To be able to maintain an overview of this amount of qualitative data through a consistent coding and analysis phase, I decided to conduct a digitally supported analysis of the qualitative interview transcripts. This means that I documented the interview transcripts in a format importable to Nvivo9 and that I conducted the coding and analysis in this tool. I defined my code categories, code relations and code hierarchy in the tool, including a coding taxonomy specifying

which codes were descriptive, analytical or merely theoretical markers. The tool also served to document the text searches I conducted during my analysis, thereby allowing me to be more systematic in the analytical steps conducted.

While I believe that the right use of software to support my analysis of qualitative data has supported me with structure and methodological rigour, it will never bring any guarantee against missing important points in the material or overseeing important bias in the data or in my own translation of the data. The same kind of methodological reflexivity is required for an analysis conducted with the help of analytical software as without it. To support me in this process I decided to follow Dahler Larsen's (2002) three rules for using and displaying data. First, I have done my best to quote as precisely as possible. I have done this by being particularly thorough in making interview transcripts and by using the analytical software to cut out text sequences without any editing.

One concern here is that most interviews were held in Danish, so that they needed translation before quoted in this thesis. It is of course critically important that this translation did not change the meaning of the quote and I have done my best to secure this. Second, I used the principle of inclusion wherever possible. This means that I worked to make sure that conclusions are based on an exhaustive analysis and display of the underlying data. Third, I aimed for transparency in the sense that data displayed is explained, figures' elements and dimensions should be unambiguous to the reader, data sorting in presentations is intuitive and easy to understand and so forth. By this, I hope the reader will find the argument compelling and the logic reasonably easy to follow.

One sometimes observes that researchers use 'quantitative language', while doing qualitative research. The risk of using quantitative labels in a qualitative context may be even more observable when using digital tools for qualitative research because the software can tend to 'objectify' the data. In my opinion, researchers using qualitative methodologies should be very careful with referring to numbers because not much scientific value can be attributed to them.²⁰ The purpose of qualitative

²⁰ As such, the occurrence of something, e.g. a particular attitude towards PA, can in some sense be perceived as a matter of frequency (one occurrence rather than no occurrence). However, my point here is that it is not valid within my design to attribute more explanatory power to a statement confirmed by two interviewees, than to a statement only confirmed by one interviewee.

research is not to give the impression that one can say anything about the distribution of variables across a population.

Parts of the coding process I used had much in common with the principles of grounded theory as presented by Strauss and Corbin (1990). I started out by having three high level categories of codes in nVivo9 called "Process", "Reflections" and "Theory". The two former categories were later renamed to "descriptive" and "analytical" code markers, in alignment with the vocabulary of Strauss & Corbin (1990). These two categories were further developed through an open coding strategy where code categories and sub-codes were defined as they presented themselves through the reading and analysis of the texts. In contrast to this, a different approach was used for the theoretical code category, which was defined up front, with departure in Weick's (1995) definition of seven properties of sensemaking.

In the *descriptive*, category I wanted to capture how managers and employees described the process and the targets in their own words. In the *analytical*, category I wanted to capture how the managers and employees reflected on the process and its inputs and outputs. In the *theoretical* category I wanted to capture how theoretical sensemaking markers could be attached to interviewee statements, i.e. how statements related to Weick's (1995) vocabulary of sensemaking.

From here, I simply started reading the first handful of interviews and added subcategories of codes to the code categories as they appeared in the texts. After the first 3-5 interviews had been coded this way, I had built a much more developed coding structure which kept quite stable through the rest of the study. At this stage I started all over again by coding the first interviews based on this more developed structure so that all interviews were coded almost based on the same coding structure. When I use the term "almost" it refers to the fact that a few coding categories only with relevance for the final appraisal interviews, were not added until at a late stage of the study. This is a consequence of the longitudinal design, since some categories did only become relevant when reading the last group of interviews which concerned the final appraisal event.

I wanted to keep my coding structure as "flat" as possible, i.e. without so many layers that it would become too complex to work with. The final coding structure had only three layers. The first layer was the 3-dimensional split between theoretical markers,

3. Methodology and Data

descriptive markers and analytical markers. A full coding overview can be seen in Figure 3-2 below.

Level 1: Marker category	Level 2: Code category	Level 3: Sub code	Number Of Sources Coded	Number Of Coding References
Descriptive	Process characteristics	Feedback	20	43
		General meeting perception	21	37
		Input from others	12	17
		Manager handover	12	40
		Meeting duration	13	15
		Meeting organisation	18	32
		Memory bias	7	7
		Other bias	9	15
		Preparations	27	59
		Privacy	13	17
		Rating Scale	17	29
		Relation to IDP	19	35
		Transparency	17	28
	Target characteristics	Are targets ambitious	9	16
		Behavioral targets	19	29
		Manager vs employee influence on targets	16	38
		Number of targets	8	9
		Target prioritisation	16	23
		Target quality	4	7
		Target realistic	14	16
Analytical	Attitude	Attitude	22	49
	Fairness	Fairness	20	35
	Improvement suggestions	Improvement suggestions	19	26
	Researcher's impact	Researcher's impact	5	6
		Career implications	21	37
	Impact	Impact on behavior	4	6
		Impact on compensation	19	35
		Impact on cooperation	3	3
		Impact on innovation	7	12
		Impact on motivation	28	51
		Impact on performance	4	7
		Impact on target setting process	1	1
		Performance Assessment	Business performance	27
	Calibration		22	75
	Dependency on others		18	29
	Impact of behavior		22	49
	Personal relations		10	13
	Relative vs absolute evaluation		11	16
	Other analytical	Strategic considerations	4	11
		Background info	2	22
Purpose of 3P		11	17	
Theoretical	Sensemaking	Action - focus on why or what in meaning creation	15	22
		Event selection - careful selection of examples	15	25
		Extraction of cues - control of event selection	13	31
		Identity construction - who am I, what is my role	22	75
		Plausibility rather than accuracy - a good story is required	8	16
		Retrospectivity - definition of individuals part in collective	14	21
		Social dimension - credibility standpoint	10	16
Total			38	1277

Figure 3-2: Full coding structure for qualitative interviews coding in nVivo9

Figure 3-2 shows the three layers of codes. All text coding is done at level 3, so that levels 1 and 2 are hierarchical categories made to order the detailed coding for analysis purposes. The table's last two columns show the number of sources (interviews) in which each code has been utilised and the total number of occurrences of each code across all sources. For example, the code "Calibration"

under the code category "Performance Assessment" under the "Analytical" marker has been used 75 times in all, and in 22 interviews. All in all, 1277 text pieces in the 38 interview transcripts have been assigned to 48 level 3 sub-codes, all in nVivo9. While this surely is a lot of coding, the nVivo tool made it easier to work with as I could with just a click select all code references for a single code, and with just another click jump to each of the reference texts themselves at the exact point where the code was assigned. Below I will explain the code definitions and present some reflections on pros and cons of the approach.

I used the descriptive marker category to capture two things. First, I coded for process characteristics as they were described and reflected upon by the managers and employees during the interviews. This included how the recent PA meeting, be it target setting, midyear review or final appraisal, was *organised* and prepared including time and spatial perceptions. It also included how the meeting was generally *perceived* by the participants, how *performance* and general *feedback* was included in the meeting and whether and how *input from others* was collected by the manager prior to the meetings. Further, it included perceived *bias* in the process, general perception of *rating scale* and its link to *rewards*, relation between the PA process and the separate process around maintaining an *individual development plan*. Finally, it included the activities around *manager handover* of PA in case of organisational changes, the *privacy* and confidentiality around PA and considerations regarding *transparency* of the PA process.

The second descriptive code category was *target characteristics*. Within this category I coded text elements that were used to describe the targets against which the later appraisal was supposed to be made. This coding was made by using a number of sub-codes. I soon found that there was a lot to be said by the interviewees about the *number of targets*. Some interviewees saw it as a critical issue if there were too many targets because this would be a sign of lack of prioritisation and focus. Generally, *prioritisation of targets* was a key concern to both managers and employees, so I coded for this as well. Another element that kept coming up was whether the targets set were *realistic*. Sometimes this came up in connection with examples of too ambitious targets. But the opposite also occurred so that targets were sometimes seen as fostering an un-ambitious tick-off process. So I also coded for statements about the level of *ambitions* in target definition, as well as the broader concept of

target *quality*, capturing whether the target formulations could hold relevance over the course of the full performance year. Finally, the texts were coded for quotes on the balance between management and employee *influence* on target formulation.

The analytical marker category was used to capture reflections of the interviewees and myself. I captured elements of the interviews where the interviewee revealed his or her *attitude* towards the PA process. In a separate category, I captured expressions about the perceived *fairness* of the process, be it either considerations about whether the interviewee found the process fair to herself, but also broader considerations about the general fairness of the process. *Improvement suggestions* were captured because I found this an interesting way to have the interviewees reflect indirectly about what they saw as most challenging with the process. I simply asked them what they would change if they were given the authority to redesign the process. In a *researcher's impact* category I captured quotes where the impact of my questioning was directly addresses by the interviewees. My study was not interventionist as such, but I found it interesting to capture thoughts on the impact of my study, typically effects in the form of a more reflectional approach to one's own participation in the process.

A number of analytical code categories were further broken down into sub-codes at level 3. *Impact* was a level 2 category capturing the perceived impacts of ratings by the interviewees. One of these codes was capturing what *Career implications* ratings have in terms of e.g. promotion, dismissal, opportunities to get interesting tasks or attractive inclusion into talent programs or high status training programs etc. A second code captured expressions about the *impact on behaviour* of receiving different ratings, i.e. considerations on how ratings is not only an outcome based on behaviour, but is also affecting behaviour as if it was almost a structuration process, to put it in Giddens' (1984) terms. A third code was capturing the *impact on compensation* of receiving different ratings. This could be in relation to bonus, but also to base salary. A fourth code *impact on cooperation* captured that some interviewees expressed thoughts on how a relative evaluation of colleagues could affect their motivation to cooperate. A fifth code captured the impact on the motivation to be *innovative* in relation to the set targets that form the foundation of the PA scheme. A sixth code captured the many thoughts expressed about the *impact on motivation* that ranking people through a PA scheme has. A seventh code

captured reflections on the impact on future *performance* of receiving a certain rating. Again, the process could be seen as circular in that the rating is given for past performance but may have effect on future performance as well. Eighth and finally a code captured how one interviewee reflected on the rating's *effect on the future target setting* process.

Performance Assessment was another analytical level 2 code category capturing quotes related to the process of providing ratings of the employees. This category contained the following codes: *Business performance* captured expressions about how the strict business performance affected the assessment process, while *impact of behaviour* was moderating the picture by capturing the softer, behavioural inputs that went into the assessment process. The code *dependency on others* captured reflections on how the individual performance was related to dependency on others, e.g. colleagues, vendors, supervisors, subordinates etc. A special *personal relations* code was made to capture expressions of how such relations affected the assessment process. Quotes on the *calibration* process were numerous and in fact the single most used code together with the identity construction code under the Sensemaking code category mentioned below. Finally, a *Relative vs. absolute evaluation* code captured reflections on the degree of absoluteness vs. relativity of the assessment process.

The last level 2 analytical code category was called *other analytical* and was broken down into three sub-codes. First, a code called *Strategic considerations* capturing more strategic thoughts behind the PA process as presented by corporate HR in the early informant interviews. Second, a code was capturing *background information* about the organisation and its development in relation to PA. This code was used for coding the interviews of HR staff. Third, a code called *purpose of PA* capturing how some interviewees by their own initiative put forward direct or indirect reflections on the purpose of the PA scheme in the organisation. The differing perceptions of purpose revealed by this made me add an extra question in the last interview round at final appraisal: I asked each interviewee to explain in their own words, what they saw as the purpose of the process seen from the organisation's side.

The theoretical marker was used to capture statements that were specifically relevant in a sensemaking analysis context. This coding category and its sub-codes were made up front based on Weick's (1995) seven conceptual properties of sensemaking

and was thus not a result of an open coding process. The first sub-code was *action* where I captured statements focusing on the action part of the meaning creation process, i.e. the 'what happened and why?' elements. I used the sub-code *Event selection* to code for statements where the interviewees gave examples of good or bad performance. In close relation to this, the sub-code *extraction of cues* was capturing statements about control of performance event selection for inclusion in appraisal decision. This of course is a process containing a lot of vested interests and potential power struggles. Not all performance events will be evaluated, so making sense by having the right ones selected becomes crucial for managers and employees. The *identity construction* sub-code was used to capture statements that indicated self-perception around questions like who am I? And, what is my role? *Plausibility rather than accuracy* was used as a sub-code to capture the 'good stories' that are sometimes required and often beneficial, when a complex activity like assessing a person's overall individual performance over a full year needs to be communicable. Further, I used the sub-code *retrospectivity* to capture statements about how the individual's share in any performance event could be limited and defined when the work context of all the interviewees was heavily interdependent on systemic and collective factors. Finally, I used the *social dimension* sub-code to capture statements about the influence on the PA process of an individual's credibility standpoint within the organisation and particularly in relation to the manager/employee relationship.

Generally, the coding structure worked well so that most of the time I found it rather easy to decide which code to use for a specific piece of text. However, I did experience some challenges, which I will elaborate upon below.

First, the text pieces did not always present themselves so unambiguously that they fit into one code definition only. Some text pieces had relevance to two or even three different codes at the same time. Particularly, I found the theoretical code category to often overlap with the analytical and descriptive categories. This is hardly surprising since the theoretical code category can be seen as a separate analysis dimension laid down on the text. Or it could be seen as another lens through which the text is read and analysed. The below example, can illustrate this.

Manager: "Regarding the ratings I would really like if I could have the feeling that I was the one who decided. That I could give the rating that I wanted to."

3. Methodology and Data

The manager is referring to the fact that ratings are calibrated and in other ways affected by other actors than the manager herself. In that sense the piece of text could be marked under the *calibration* sub-code. However, looking at the quote from a sensemaking perspective, this could be seen as an element of *identity construction*. The identity as manager is built around influence and the power to make decisions including assessing the performance of the manager's employees. The calibration process limits that power in a way that obviously annoys this manager. To make the picture even more blurred, the quote could also be coded as an *improvement suggestion* since it could be translated as a proposal for increasing managers' rights to freely appraise as they please.

One of the advantages of using digital tools for coding is that it is so easy to drag and drop pieces of texts into code categories, so I simply decided to allow myself to code the same piece of text into multiple sub-codes. This way I avoided losing information at the sub-code level, but I also at the same time introduced a second problem: I ended up with a very large number of codes - more than 1200 codes were attached to the 38 interviews that went through the coding process. Had this been done in the old, manual way, it would have been very difficult to avoid getting lost in this volume of coded text pieces. Again, the digital analysis tool makes it easy to extract all text pieces attached to a code at a mouse click. The number of coded pieces, although time demanding in the analysis process, gave me a richer foundation for the analysis.

The third challenge I faced during coding was related to the two already mentioned above. I coded the texts with 48 codes. My intention was to keep the coding tree as flat and simple as possible to avoid the coding process drowning in complex code hierarchies, definitions and interrelations. However, 48 sub-codes is still a lot. Looking at Figure 3-2 one can see that there are a number of sub-codes that were rarely used. Codes that are only used a very few times across a large number of interviews should in my mind at least raise concern if the code is critically important, or if it can be either dismissed or incorporated into another code definition, to keep the coding structure as simple as possible. As an example, the sub-code *Impact on target setting process* was created while coding one of the early interviews and at that time seemed relevant. As I went through the coding of the other interviews, it

turned out to not be something others were touching upon.²¹ Still, I decided to keep the code, because I found the reflection interesting.

The fourth challenge I experienced in the coding process was in a way in opposition to what I have just expressed. The challenge was not that 48 codes were too many but rather that it was too few. When open coding is used, and it is at the same time a guiding design principle to avoid too many codes, there is a risk that the researcher is twisting quotes into sub codes where it would be fairer to the quote to attach it to a new and separate code definition. An example can illustrate this dilemma. The quote below was coded as *preparations* because it occurred in the context of discussing how a PA meeting with the manager was prepared.

Employee: "I believe all managers in the organisation are well-prepared for these meetings". Interviewer: "Why do you think it is so?" Employee: "Because they have to. It's governance. They cannot allow themselves to go red on this one."

Why not create yet another code called "governance" or "manager success criteria"? In this example I chose to stick to the *preparations* code because I thought that a discussion of governance criteria as a motivator for manager meeting preparation would fall naturally in the context of analysing the meeting preparations. However, I admit that it could easily be argued to do otherwise. I found it important to keep the code structure as simple as possible while still capturing the most significant text perspectives in a structured manner. Keeping this balance comes through compromises and the important thing then is to be consistent in the use of the coding structure through the coding of all texts.

3.5 Procedures for Research and Documentation

Besides the above methodological considerations, the research protocol also laid out the more practical procedures for conducting research. These included procedures to secure confidentiality and ethics, establishment of templates, preparation of invitation letters, principles for keeping track of the large volume of established data in a research database and finally procedures for storage and backup of data. Although this part of the research protocol was very stable over the course of the study, some

²¹ I see this as a side effect of using a grounded theory approach to coding.

changes were required along the way as I learned more. The procedures will be presented below.

Confidentiality and ethics

It was an unbreakable principle of mine that informed consent must be obtained from all individuals who were interviewed. Thorough information about the study purpose and the role of the interviewees were delivered both in writing before the actual interviews as part of the invitation letter but also verbally in the beginning of the first interview of every interviewee.

Initially, I was considering the use of the terms confidentiality vs. anonymity. I ended up using the former because the purpose would be to protect individuals from exposure or in the worst case even sanctions, for participation in the study. It would not be possible to reach anonymity in the meaning that it should not be possible to identify who had participated as informants in the study. If someone was really interested, of course it would be possible to find out who was interviewed. In fact, the managers all knew exactly which of their subordinate employees I was interviewing, because it was a central part of the study design that the interviews were focused on a specific one-to-one relation between a manager and an employee. This design had a lot of advantages in that the interviews became very concrete and because it allowed me to study how views on the same performance and meeting events differed between managers and employees. But in a confidentiality context it of course also presents some tricky concerns about how to avoid disclosing managers' and employees' more controversial statements, particularly where they directly concern each other in a critical way.

Generally, protecting especially vulnerable persons, particularly ones who have controversial views or behaviours that may affect their opportunities if recognised is an important obligation for the researcher. As the interviews progressed and we got to know each other better, most became less cautious and more open. This revealed views that were not always supportive of the current conduct of PA in the organisation. These statements were given under confidentiality and it has been of critical importance to me to make sure to keep that confidentiality so that no individual will be exposed due to their participation in the study. This may limit my ability to link certain quotes to specific workforces because it would be too easy to

identify the interviewee quoted. However, this is the price for keeping confidentiality and for staying loyal to the agreement made with the interviewees. And in most cases, the quotes can still be used, unlinked to workforce, without missing significant value to the study.

The PA process is surrounded by a strong culture of privacy. Neither appraisal ratings nor appraisal process perception is something the interviewed managers and employees commonly discuss with their peers. For some of the interviewees this secrecy seemed to work as a lid on a boiling pot. One employee told me:

Employee: "...it is a bit funny that the only one I can share these thoughts with, is a stranger like you. We are sort of muzzled, because they don't like us to measure ourselves up against each other. They want to do that themselves in secrecy."

As the example shows, my own role as the person lifting that lid, letting some of the steam out, raises rather strong requirements for staying loyal to the promised confidentiality. One of the ways I have tried to live up to this requirement is by consistently using the female form when referring to any manager or employee, no matter their real gender.

Finally, the data collected is not to be registered with *the Data Protection Agency* since it falls out of scope for this as defined by this institution.²²

Letters and templates

Letters were prepared to fulfil multiple purposes. Letters to HR partners had the purpose to identify relevant interviewees within the four workforces. Letters to participating managers and employees had the purpose of briefly informing about the project and to present the terms of confidentiality under which the interviews would be conducted.

Templates were prepared to assist me during the data establishment phase. For the longitudinal interview cycle, I initially only prepared the template for the target setting event. The templates for the midyear review and the final appraisal events were prepared later, closer to conducting the interviews. This allowed me to include

²² See: <http://www.datatilsynet.dk/erhverv/forskere-og-medicinalfirmaer/private-forsknings-og-statistikprojekter/>.

learnings from the previous data establishment phases into the preparation of the later phases.

There were separate interview templates for managers and employees, so I had prepared six interview templates in all. The templates were basically lists of questions that I could refer to if the interview would benefit from this. I did not cover all the prepared questions in all interviews but rather let the interviews develop based on the input and stories told by the interviewees. However, in most cases we covered most of the template, although sometimes without having to explicitly formulate the questions. After the first few interviews in the first interview round, I modified the template based on the learnings of the first interviews, so that it found its final form. One could ask if it is not a problem for the validity of the study that the interview template developed over the course of interviews. In my opinion it was not. In a quantitative study it would of course be a problem if respondents were not asked exactly the same questions. In a qualitative study based on semi-structured open interviews, this is on the other hand almost by definition the case. In my opinion, it would in this context rather be an issue if one did not use the learnings from prior interviews to improve the foundation for acquiring richer input from the following interviews.

At the first interview round regarding target setting, I handed out a diary template to each manager and employee. The purpose of the diary was to help managers and employees better capture relevant performance events through the year so that the process was better supported by documented facts. In theory, such documentation could be added directly in the PA IT system, but I knew from personal experience that close to no one ever did this.

The diary templates included a short introduction to the diary purpose, some hints as to what could be relevant to note in the diary and a three-column scheme where notes could be documented by date, event description and personal reflections on event. The templates were only a little different for managers and employees, in that the manager template included hints about documenting broader activities like calibration meetings regarding PA, bonus and salary etc, where this had relation to the PA process.

Of course, I thought it could be interesting to be able to compare the diaries of managers and employees to see both which performance events they logged as relevant in the PA process context but also to see what reflections they made, to see if there were differences in their evaluations of and reflections on the same performance events.

Naturally, it was up to each of the interviewees to decide if it made sense for them to use the diary. And it was up to each of them to decide if they wanted to share the diary with me. Most were positive at start, but as it turned out, only two employees systematically used the diary all through the performance year. They both shared their diaries with me. In a way, the diary offering could be seen as a sort of experiment to investigate the willingness in practice of managers and employees to do a small extra effort to increase the quality of the PA process. In this way the lack of use of the diary, despite the expressed initial interest and belief that it would be valuable, was in itself an interesting observation which will be dealt with at more length in a later analysis chapter.

Besides these templates and letters, an interviewee management document was maintained. In this I logged all interviewee's contact information, planned and held interview dates, pending actions and general comments. This way I made sure to keep a central overview of the overall interview progress.

Building a research database

When conducting a study involving both a fairly large literature review of 125 journal articles and a longitudinal, multi-method empirical study, one will have to handle a huge volume of material and documentation of various kinds. To be able to keep track of all this documentation I built a research database in MS Excel.²³ The database contained four elements.

First, an overview of all scientific articles used in the context of the study. This included the 125 articles selected for my literature review and an additional number of articles used as background information and other sources. The 125 articles could be filtered on a number of dimensions on which they had been coded.

²³ For a thorough argument for the importance and benefits of working with a research database, see Yin (2009)

Second, an overview of all key archival documentation files used in the research. This overview could also be filtered in accordance with a number of descriptive codes and further indicated whether the files contained confidential information.

Third, an overview of all interviews conducted including details about the interviewee, timing, purpose etc. The interview overview did not include the coding of the interview itself as this was kept within an nVivo9 database.

Fourth, an overview of other notes, e.g. notes from participation in academic courses and conferences, notes from meetings with my supervisors, notes from participation in own and other PhD students' Work in Progress seminars etc.

Storage and backup of research data

To make sure everything is stored in an orderly way and kept safe from a data protection perspective, I built an archiving structure that matched the different elements in my study. All materials have been saved in electronic format on a network drive with nightly backup all year around. Access to the network drive is limited to me personally.

To avoid any risk of disclosure of confidential material, I have simply avoided having any physical printouts of interview transcripts or participant observation notes. Keeping data only in electronic format and in a limited access manner is in my opinion the best way to avoid unwanted data disclosure.

3.6 Summary

In this chapter I have presented my empirical methodological approach as it was first laid out in a detailed research protocol and as it developed during the course of the study.

The key component in establishing empirical data was 34 qualitative, deep dive interviews held with pairs of managers and employees over a full year's PA cycle, from target setting to final appraisal. I have argued for why I found it so useful to take advantage of a longitudinal design due to reasons of the process construct itself, the nature of sensemaking processes, the importance of time proximity to performance events, the heavy institutionalisation of the field and the necessity of building trust,

reflection and an open dialogue. And finally because my literature review had already clearly proven how longitudinal studies have often been recommended but very rarely used by scholars.

The transcribed material was digitally coded and analysed in nVivo9 through a combined process of open coding for descriptive and analytical markers and closed, theoretical coding for text pieces relevant in a sensemaking analysis perspective. More than 1200 code attachments were made across the 34 manager and employee interviews, and 4 additional interviews with representatives from corporate HR.

Because of my own privileged position as an insider at the organisation and in corporate HR, I also used participant observation to document my own observations and experiences. However, the use of participant observation was more limited than initially expected. One reason for this was that I found there were better ways to establish data to support a sensemaking analysis. But it was also because I found there would be too many ethical and practical issues.

The third empirical method I pursued was analysis of archival records and documentation. First of all I used this method both to get a good overview of the developmental history of the corporate PA scheme in the organisation and to analyse the organisation's macro level experiences with PA from a quantitative standpoint. But I also used the material together with the input from my own observations and interviews with people from corporate HR, to support the initial thought process required to prepare the research protocol including final refinement and formulation of my research question and my full approach.

As part of the research protocol I had defined very strict procedures for research and documentation which I have been keen to adhere to through the full duration of the study. These included procedures to secure confidentiality and ethics, establishment of templates and preparation of invitation letters, principles for how to keep track of the large volume of established data in a research database and finally procedures for storage and backup of data.

I think it would be fair to overall label my approach as belonging to an intersubjectivist epistemology. My own views, questions and general perception developed extensively during the process and through the interviews - as did the views of some of the interviewees. I find this to be only natural when working with longitudinal,

3. Methodology and Data

qualitative studies through open interviews. Having an open mind towards research questions, methodology, and theory is in my opinion a requirement for being able to conduct interesting qualitative research. If one wants to test a specific issue via closed questions and without the opportunity to adjust along the way, one would be better served by a quantitative approach. But investigating complex individual meaning creation in a highly institutionalised field, does not call for closed questions or quantitative methods. My Interviewees did not see things in one uniform way through the three interview phases. They developed their views and arguments - sometimes even to contradicting positions. And so did I.

Conducting a qualitative study in my own organisation requires mastery of balancing the obvious advantages of being an insider with the according risk of prejudice. By spending so much effort and time preparing a detailed research protocol, I have tried my best to make my approach to this dilemma as transparent as possible to the reader.

While a chapter on methodology naturally focuses on epistemological questions I find it most fair to the reader to finalise the methodology chapter to be transparent about my ontological standpoint. I consider myself to be a critical realist in the sense that there is a material world outside our minds, but it is complex and difficult to understand, and we therefore need theories to help us understand more, although still not all.²⁴ In a PA context there are real things at stake: real projects and activities that are to be evaluated. There are also real performance ratings and real rating consequences. But the evaluation of an individual's contribution is highly subjective and partly irrational. In fact, the past 20 years of research into evaluation bias and other problems in PA makes this quite evident.

²⁴ See e.g. Archer et al. (1998). Although there are real things at stake, the pre-defined corporate PA process is not uniformly implemented and operated across the organisation. It is mediated by and influenced by sensemaking processes driven by a need for creating meaning. I.e. a need for rationalising what has been done, rather than doing what is rational.

4 Case Introduction

PharmaComp²⁵ is a large Scandinavian pharmaceutical company. The company is specialised in research and development, production and marketing of medical drugs. Headquartered in Scandinavia, the company has a worldwide presence with production sites and sales affiliates all over the world.

The present case study overview is based on preliminary studies of internal documents, my own background knowledge and interviews with key stakeholders in the implementation and operation of PA in the company. These informants have held varying roles over the years since PharmaComp set out to implement PA. For example, an informant was managing the implementation project, one is a central HR statistician utilising the new opportunities for analysing aggregate data and another is responsible for the continued improvement of the global PA process.

Over the past 10 years the company has experienced a strong growth rate in terms of sales and also a strong increase in number of employees. The strongest growth markets both in terms of sales and number of employees are outside Europe.

In 2004 there was no corporate PA process in the company. Instead, local practices had developed over the years at the discretion of local management. A survey initiated by the corporate HR function in 2004 showed that there existed at least 25 different practices of PA in different parts of PharmaComp. The practices differed in terms of:

- Standards (e.g. was there a rating scale or only a free text evaluation? How many grades were there in the rating scale? Was a recommended or ideal rating distribution defined?)
- Documentation (e.g. was a template used and how was the template designed and stored? Was data aggregated at higher organisation levels?)
- Process (e.g. who were in scope for PA? How many and what steps were included in the process? How many compulsory yearly meetings and appraisals were there?)

²⁵ In the following I will use the term PharmaComp as a pseudonym for the case organisation.

The consequence of this state of affairs was that there was no global overview and strategic direction in terms of PA. No one had a clear overview of how to identify talent across the global organisation. There was no IT system to support alignment of the PA process although this was giving way to increasing challenges due to an accelerating internationalisation and a subsequent increased mobility. Thus, performance evaluations were generally not transferable or comparable across organisational borders. In addition, from an HR efficiency point of view, there was an inefficient use of resources since so many different templates and processes were maintained locally and simultaneously.

In August 2004 Executive Management decided to launch a corporate PA process. The intention was stated as to:

- Establish a common baseline for PA across PharmaComp
- Support the strategic focus on internationalisation and mobility
- Improve performance data overview and alignment of processes
- Secure compliance with the corporate management principles in all units
- Create corporate transparency on individual performance
- Prepare for an IT system roll out

While choosing this approach, PharmaComp relied heavily on a 'best practice standard' as defined in the People Capability Maturity Model (SEI/CarnegieMellon). According to this model PA should support the realisation of the unit's targets and individual performance objectives should be documented, performance should be regularly discussed to identify actions that could improve it, performance problems should be managed, outstanding performance should be recognised or rewarded and PA practices should be institutionalised to ensure they are performed as managed processes (Curtis, et al., 2002).

The plan was to first establish a companywide paper based process for PA of all managers and up, in phase 1, and later to implement a supporting IT system in phase 2. The process was named the "People Performance Process". It was decided not to include the relation to other areas such as rewards (salary increase, long-term incentives, and short-term incentives) in the scope of the first two phases, but leave these more advanced relations to a future phase 3. In Figure 4-1 below, the PA

journey is illustrated as it appeared in the project documentation at the outset in 2004.

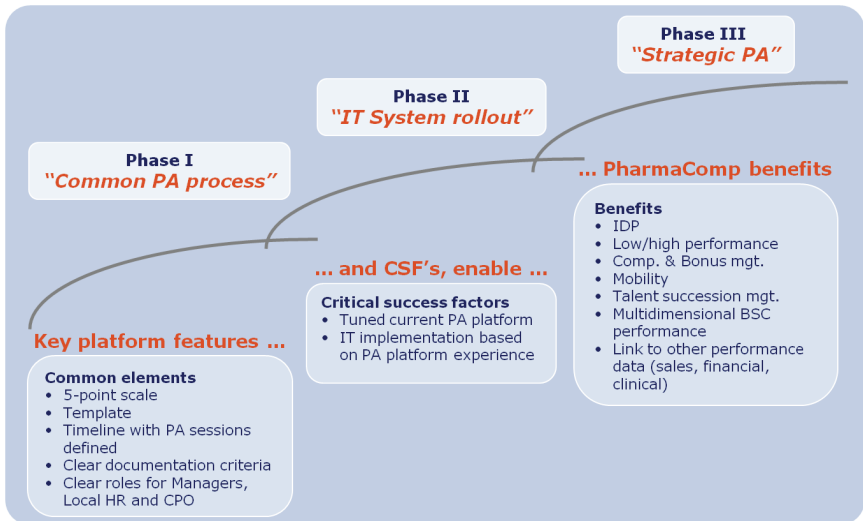


Figure 4-1: PharmaComp's PA journey as set out in 2004.²⁶

Figure 4-1 illustrates the three-step implementation process for implementing a corporate PA scheme in PharmaComp. Below the three phases will be explained in further detail.

4.1 Phase 1: Establishing a Common PA Process (2004)

In phase 1, a set of minimum requirements were defined: a) A common target setting, review and appraisal paper template, b) A common 5-point rating scale, c) A common process and timeline for target setting, review and appraisal (three mandatory yearly events), d) Documentation requirements related to the process and e) Clear roles and responsibilities for managers, local HR and Corporate HR. Below, each of these five core characteristics will be further clarified. Blue collar workers were, not included, partly because of union resistance but also because it was seen as less relevant to manage individual goals in this way for people doing more routine tasks on work shifts.

²⁶ People Performance Management ver.02 (2004)

A common template

Corporate HR decided to collect all existing local templates as inspiration for the preparation of a new standard template that supported the corporate objectives. Figure 4-2 shows the new template.

3P PEOPLE PERFORMANCE PROCESS					
People Performance Process (3P)					
Year: <Year>		Rating scale O: Outstanding AE: Approaches expectations and goals EE: Exceeds expectations and goals NM: Does not meet expectations and goals ME: Meets expectations and goals N/A: Not Applicable			
Employee: <Name/Initials>					
Manager: <Name/Initials>		Use the 3P toolbar to add, delete and rate goals. ↓			
Balanced Scorecard Perspective ↓	Goal setting (December 1 st to February 29 th)		Mid-year review (June 1 st to August 31 st)	Year-end appraisal (December 1 st to February 29 th)	
	Balanced Scorecard CSF/KPI/Targets ↓	Weight (%) ↓	Comments ↓	Comments ↓	Rating (see scale above) ↓
Perspective 1	CSF: KPI: Target:		Employee: Manager:	Employee: Manager:	
Perspective 2	CSF: KPI: Target:		Employee: Manager:	Employee: Manager:	
Additional goals	CSF: KPI: Target:		Employee: Manager:	Employee: Manager:	
Supporting Competence Goals	CSF: KPI: Target:		Employee: Manager:	Employee: Manager:	
Notes			Conclusive review comments ↓		
			Employee: Manager:		Manager conclusive rating ↓
			Conclusive appraisal comments ↓		
Employee: Manager:			Employee: Manager:		
Employee signature/date:		Goal setting:	Mid-year review:	Year-end appraisal:	
Manager signature/date:					

Figure 4-2: The Corporate PA template

The template in Figure 4-2 contains a number of information points to be tracked through the PA process. There was also a reference to the “balanced scorecard perspective” of the business area. The balanced scorecard structure of PharmaComp operates within four perspectives: Financial targets, business process targets, customers/society impact targets, and people and organisation targets. On top of the four balanced scorecard perspectives, two additional perspectives were added; a general “other goals” category and a supporting competency goals category. Then, for each of the three yearly sessions there were distinct columns. In the goal setting column the targets were codified according to a specified format relying on the definition used in the breakdown of the corporate balanced scorecard.

First, a Critical Success Factor (CSF) should be referenced. Critical Success Factors are those factors that are most critical for the achievement of business success. CSFs identify the significant things an organisation or an individual must do well to succeed. Thus, the corporate balanced scorecard definition of CSF states that:

A CSF does not describe a current situation, but an aimed objective. The objective should be challenging, yet realistic. The balance and priorities set in the CSF are particularly important. Each perspective usually contains 3-6 primary CSFs that state a key aspect of the strategy to be achieved over the next 2+ years.²⁷

Critical Success Factors are the most general level of targets. Examples of critical success factors at the corporate level would be e.g. *Realise growth in Operating Profit* (Finance perspective) or *Enhance winning culture* (People & Organisation perspective). The purpose of defining the link to critical success factors and generally following the vocabulary of the balanced scorecard process is to create a strong link to the strategic direction set by senior management.

Second, for each target a KPI should be defined. The role of a KPI is to specify a CSF and the KPIs are part of PharmaComp's management reporting system. The corporate definition of how to create KPIs state:

The definition of the KPI should demonstrate how and when the CSF will be achieved, and must be measurable. One or two suitable indicators should be defined for each CSF, demonstrating its achievement. The indicator can be displayed in absolute terms (e.g. \$000m), or using other valuation criteria (e.g. employee satisfaction index).²⁸

Examples of KPIs are Operating Profit Growth (Finance perspective) and Customer Satisfaction (Customers & Society perspective).

Third, the concrete targets are measuring points for the KPIs, which ideally provide management and employees with a clear idea of what to strive for, and by when. Referring to the KPI examples above, targets could be *operating profit growth = 12%* or *Customer satisfaction = 4.0 on a 5-point rating scale*. Each target definition would also require a due date for the target to be measured and evaluated.

Fourth and finally, a target weight could optionally be added to codify the relative prioritisation of targets.

²⁷ Corporate definition of Balanced Scorecard elements, Corporate intranet 2011

²⁸ Corporate definition of Balanced Scorecard elements, Corporate intranet 2011

The midyear review column should hold the employee and manager comments as discussed through the midyear session. Besides a follow-up on the status of target realisation, the midyear comments could also include rationale for adding new goals, changing existing goals or even deleting goals that had become obsolete during the time passed since target setting.

The year-end appraisal column holds the final evaluation by the appraisee (self-appraisal) and the appraiser. The appraiser has the final word in rating each target on the corporate 5-point rating scale (see below), but the appraisee must firstly note her self-appraisal as input to the discussion between the two. As for the midyear review, there is also for the final appraisal a table box that holds the conclusive appraisal comments from both the appraisee and the appraiser. But this time there is an additional field where the manager documents the overall rating of the appraisee's performance as measured on the corporate 5-point rating scale.

For each of the three appraisal sessions there are signature fields for both manager and employee to document that the process has been conducted timely and correctly. The appraisee does not sign to her agreement with the appraiser's evaluation, but to her own comments, and to verify that she has read and received the appraiser's feedback.

A 5-point rating scale

One of the most difficult steps in aligning the many different PA schemes existing prior to the corporate standardisation initiative from 2004 was to establish a compromise for a corporate appraisal rating scale. One of the key people involved in the 2004 standardisation initiative recalls:

It was a mixture of pragmatism and best practice. There were all these many rating scales – 3-point, 4-point, 5-point and 10-point scales. There were most that used 5-point scales, and also, our later IT solution provider recommended a 5-point scale as most normal. So it became a 5-point scale.

But how was the decision to even have a rating scale, i.e. a quantitative rating of employees, taken in the first place? An HR person close to the 2004 initiative explains:

It was a given that there should be a rating scale. It should be possible to compare across PharmaComp to be able to deal more systematically with low and high performers. What did a 3 mean here and a 10 there? Nobody could compare across organisational entities. We needed to be able to track our top performers.

So, the result was a 5-point scale ranging from *does not meet expectations and targets* to *Outstanding*. Below each of the five levels are defined in accordance with the original rating scale definition from the 2004-initiative:

Outstanding. An outstanding contribution to the company's performance. The person delivers substantially more than is required. The performance leaves little or nothing more to be desired. The way to accomplish the target(s) has been significant compared to standard in unit. The reaching of the target(s) will be used as a best practice in the unit.

Exceeds expectations and targets. A significant contribution to the company's performance. The person has exceeded expectations – faster and/or better than set forth and with desired behaviour. The person has during the reaching of the targets anticipated and proactively solved critical problems.

Meets expectations and targets. A solid contribution to the company's performance. The expected results have been achieved within agreed timeframes, available resources and with the desired behaviour.

Approaches expectations and targets. The person should improve his contribution to the company's performance. Performance fails to consistently meet expected results. All targets have not been met and currently, accomplishment of targets is below the standard of the unit.

Does not meet expectations and targets. The person's contribution to the company's performance is not acceptable. Targets are far from met and behaviour is considered poor or even negative.

In addition to these five categories, there was a sixth called *not applicable*. This was reserved to situations where targets were postponed or not relevant. From the outset, there was no support for enforcing a forced distribution of ratings on the 5-point scale.²⁹ A corporate HR representative involved in the 2004-initiative explains:

Where we did not succeed was to build support for a forced distribution. Personally, I was in favour of forced distribution but without absolute definite descriptions in the rating scale. I thought we should be clear: 'We rank our people, it is a race and someone is first'. But there was no support from top management for a relative scale, probably because they favoured a more cautioned approach with focus on getting basics right first: Is the processes followed and are timelines met? We in HR also had ambitions concerning quality, meaning how things were documented and audits of the process outcome, but we did not get through with that at that stage.

So, everyone in scope for PA was rated against the 5-point rating scale, but there was no forced distribution across PharmaComp and no systematic follow-up on differences in rating levels.

²⁹ Forced distribution is a process according to which a specific share of employees must receive a specific rating. E.g. that 5% of all employees must be within the lowest rating category. The purpose of forced distribution is to be "the antidote to the problems of inflated rating and the failure to differentiate that many organisations have installed to help bring the truth into the performance management process" (Grote, 2005)

The yearly process cycle of eight steps and three sessions

An 8-step process was defined and implemented around a three-session structure, as illustrated in the original process map in Figure 4-3 below. Except for a few minor adjustments, this process has been in operation in PharmaComp since its implementation in 2005.

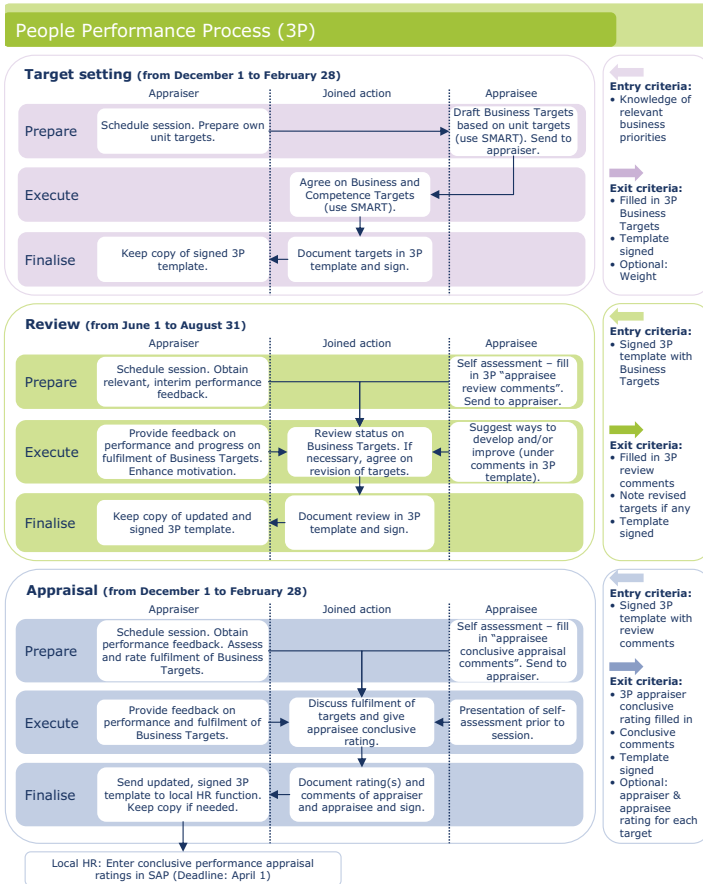


Figure 4-3: Initial PA process definition³⁰

While the appraisal period is defined by the calendar year, the target setting starts on December 1 and ends February 28, after the appraisal period. This is to allow

³⁰ People Performance Management ver.02 (2004)

business areas time to define targets before the appraisal year starts, finish the appraisal year before evaluations are done and generally have time to conduct meetings etc. The process requires at least three appraiser-appraisee sessions per year.

Below, I will go through each of the three phases and the eight steps in more detail, in accordance with the official documentation held by the company and presented to the managers and employees through training, communication and other material.³¹

Target setting

Step 1: Employee creates PA draft targets

To begin target setting the employee prepares a number of targets reflecting the main business objectives for the performance period. Targets must be defined according to SMART³² principles, in this context defined as Significant, Measurable, Ambitious, Realistic and Time bound.

Targets can be created and aligned with business objectives in different ways. Targets may be created individually by the employee or be cascaded to the employee from the manager or any other employee. When business targets have been prepared, the employee needs to decide which competence targets may support the business targets, and include these in the PA form. The use of competence targets is optional, unless local guidelines stipulate otherwise.

The employee has the opportunity to attach documentation – for instance an Individual Development Plan. Weighting targets is an optional way of aligning appraiser's and appraisee's view on the priority of the specific target. Whether a target is weighted high due to the time it takes or due to importance, or both, is a matter of dialogue between appraiser and appraisee. The managers and employees were encouraged to consider the following if using weights:

Weights should be defined at target setting in order to align expectations. Weighting is measured in percent and should equal 100%. Weights should be used as guidance in reaching an overall conclusive rating at year end.³³

³¹ People Performance Management ver.02 (2004) and corporate website

³² Doran, G. T. (1981). *There's a S.M.A.R.T. way to write management's goals and objectives*. Management Review, Volume 70, Issue 11 (AMA FORUM), pp. 35-36.

³³ People Performance Management ver.02 (2004)

If weights are not used, alignment of expectations has to be done in other ways. Such considerations should also be documented in the template.

Step 2: Manager approves PA targets

Manager reviews the employee's Business and Competence Targets in the PA form and provides comments on the form. If relevant, the manager can suggest revisions or additional targets. When the PA form fully reflects employee and manager mutual agreement on targets and prioritisation, the manager and employee approves the target setting for the coming performance cycle by signing the PA form.

Mid-year review

Step 3: Employee enters mid-year review

During mid-year review the employee evaluates progress on fulfilment of business and competence targets year-to-date. The employee optionally enters review comments for each target, and conclusive review comments.

Step 4: Manager enters mid-year review

The manager evaluates employees' progress toward fulfilment of business and competence targets. The manager reviews the employee's mid-year review comments and enters own review comments for each target, and a conclusive review comment on the employee's performance to date. The manager signs the midyear review.

Step 5: Employee signs mid-year review

The Manager and employee meet to discuss the midyear review. At the meeting the employee accepts the mid-year review by signing the PA form. By doing this the employee does not give her consent to the manager's evaluations, but rather to the fact that a review meeting has taken place.

Year-end appraisal

Step 6: Employee enters year-end appraisal

The employee evaluates her own fulfilment of business and competence targets and enters self-evaluated year-end appraisal ratings for each target. Optionally comments can be added for each target. The employee enters conclusive year-end comments and sends the form forward to the manager.

Step 7: Manager enters year-end appraisal

The manager reviews the employee's progress on fulfilment of business and competence targets. The manager enters appraisal ratings and comments for each target, and provides overall conclusive rating and comments. If applicable in the unit, conclusive rating is at this stage calibrated against the ratings of other employees.

Step 8: Manager and employee signs year-end appraisal

At a meeting between manager and employee, the employee reviews and signs the year-end appraisal. By signing the PA form the employee confirms that the year-end appraisal, feedback and comments, has been discussed with the employee.

Documentation requirements

With the new uniform process and rating scale it suddenly became possible for corporate HR to conduct reporting at the corporate level. Although there was no systematic follow-up on correlation to rewards or other areas, even the ability to compare process and deadline fulfilment and outcome in the form of appraisal rating distributions between units or levels, was new and interesting to corporate HR. But as long as the process was only paper based, the condition for accessing this data was that all units remembered to enter the final appraisal for each appraisee in the corporate ERP system. Thus, besides storing the PA template locally, each unit was also responsible for performing this update within certain deadlines.

Since the scope of the PA process covered a defined group of employees, corporate HR could police the process fulfilment by reviewing the degree to which a final appraisal had been stored in the corporate ERP system at any given date. This also meant that the final appraisal soon became the key indicator of process fulfilment – if there was a recorded final appraisal rating for each employee in scope for appraisal in any given unit, the unit was in process compliance. No further follow-up was at this stage made from corporate HR in terms of quality of documentation.

Corporate HR did follow up on appraisee's satisfaction with the appraisal process in an indirect way in that new questions were added to the yearly employee satisfaction survey held each year. However, the questions posed at the midyear term were distant in time from the critical target setting and PA meetings held in January or February.

Responsibility definition

An important element in having a globalised process was to establish clear definitions of roles and responsibilities within the PA process and across PharmaComp. Four key roles were identified and defined: the appraiser (the manager), the appraisee (the employee), local HR and Corporate HR.

The appraiser is the person who assesses and rates performance against targets. Her responsibilities are to plan and conduct the PA sessions, i.e. the meetings for target setting, midyear review and yearly appraisal. The appraiser prepares own unit business targets that will be cascaded to all or selected employees within the unit. The appraiser is responsible for sharing informal and formal feedback to the direct reports, at least at the three yearly sessions. The appraiser was also recommended to practice good leadership by providing constructive feedback on a more daily basis.³⁴ A very important responsibility of the appraiser in the appraisal phase is of course to ensure objective performance assessments of direct reports. It was also the responsibility of the appraiser to ensure audit trail through documented PA sessions by keeping an archive of appraisal forms for the unit and to submit the final, signed appraisal documentation to local HR for local storing.

At the same time the appraisee, i.e. the person who delivers on targets and gets appraised, was holding a shared responsibility. She should contribute actively to the PA process by participating constructively in all PA sessions and by keeping herself updated on the unit's business targets, and on the timeline and deadlines of the PA process. She should also prepare a proposal for own performance targets based on her job description, the targets for the unit and the task portfolio as it was known at the time of target definition. In addition, the appraisee should conduct self-assessment of performance before the appraiser was giving her appraisal input.

Local HR/administrative manager's role was to ensure and support PA processes directly at the local sites. Generally, local HR was to ensure that local PA practices met the corporate requirements of PA, i.e. the process, the timelines, the template, the archiving and that appraiser and appraisee were living up to their responsibilities. Local HR should also train the appraisers and keep them informed about new guidelines or tips and tricks on PA. When final appraisal was finalised, it was the

³⁴ People Performance Management ver.02 (2004)

responsibility of local HR to enter the final ratings in the corporate ERP system, to make corporate reporting on rating distribution possible.

The role of Corporate HR was to ensure that the company's PA standards on a global level were maintained and adhered to. This was to be done through supporting local HR in their role by providing training and guidance to local HR and by providing general management training in PA. Another key role for corporate HR was the operation and analysis of aggregated performance data to be able to evaluate the degree to which the global intentions were realised.

4.2 Phase 2: Implementing the IT system (2006)

In 2006 the paper based process had been rolled out to the white collar organisation and a full year's PA cycle had been conducted. It was now time to proceed to the next phase: The implementation of an IT solution. The IT solution went live in December 2006, covering the 2007 performance year cycle. However, only the top layers of PharmaComp were included in the scope for the IT system implementation, namely senior vice presidents (SVP) and two management layers down. A corporate HR representative recalls:

Implementation of a common IT system had been part of the plan from the beginning... There were concerns for the risks in a big bang approach so instead it was decided not to go deeper than SVP minus two. The implication of this approach was that the top layers of PharmaComp now documented their PA sessions in the IT system while other groups continued to use the paper template. This gave way to some confusion and demand for further rollout of the IT solution. Soon after, we opened up for further rollout to all managers and specialists who reported to a vice president. Since then, scope has somehow been a bit confused. Many professionals and even administrative employees are now using the IT system, but there is no clear picture of who is on and who is not. It does not seem to be a controlled development.

All in all approximately 2.500 managers and employees were using the IT solution through the 2007 performance cycle year. With the introduction of an IT system, no paradigmatic change in the PA concept was introduced in PharmaComp. It was still the same template, the same rating scale, the same process, the same procedures for archiving and the same roles and responsibilities. The main difference was that it now became much easier and faster for corporate HR to keep track of process progress in the local units. Still, the rollout of the IT solution was met with significant organisational resistance, as an HR staff recalled:

When the process was rolled out as an IT solution, it was like reality suddenly became clear in PharmaComp. Until then, I think many had just seen the PA process as 'this year's management lingo'. Now they were suddenly forced through each step in the process with full transparency for Corporate HR to see what was going on.³⁵

The organisational resistance was very much directed at the user friendliness of the system. Corporate HR mitigated this by undertaking a larger initiative in 2007, where many improvements were implemented. Since then, there has been quieter around the PA IT system. Although yearly user satisfaction surveys still show some dissatisfaction with the system, it is not easy to judge from this data whether this is due to the system or the process itself.

The IT system for non-mandatory users is funded by the organisational units of the users themselves. This means, that for any manager or employee outside the SVP-minus-two group that use the system, there will be a yearly local cost. This cost is the same no matter where in the world the user is located and is thus relatively higher in low cost areas like China or India. For this reason, the IT solution coverage has from the start been much broader in high cost areas like North America and Europe than in the rest of the world. However, with the increased user base over the years, the average cost per user has decreased significantly and now to a level where also the low cost areas start to show interest in joining the IT solution.

4.3 Phase 3: Aiming for Strategic Performance Management (2007-)

Phase 3 was from start looser defined than the two first phases. Whereas phase 1 delivered a uniform process and phase 2 delivered a supporting IT system, phase 3 was intended to include the more long term, strategic development of the process. In this phase focus should be on harvesting the benefits of the overall initiative. It was envisioned that this could for example be in terms of improved basis for establishing quality individual development plans, identifying high and low performers and making sure compensation and bonuses were distributed in accordance with performance, supporting the mobility candidate identification process, supporting the talent identification process and the succession planning process and more.

³⁵ Interview with Corporate HR representative involved in the 2004 standardisation initiative.

All this did not happen at once. Nor as one coordinated effort, as did the first two phases. In fact, it is a bit misleading to describe it as a "phase" at all, since there was no defined end to this stage and no clear scope or success criteria. It would be more correct to describe phase 3 as a series of initiatives aimed at improving satisfaction with the PA process and improving the opportunities for harvesting further strategic benefits from PA. The initiatives included both an on-going further rollout and refinement of the IT system supporting the process and initiatives aimed at increasing the business value of the process. The most significant of these initiatives are briefly introduced below.

Further rollout and improvement of the IT solution

Considerable resources have been put into further rollout and improvement of the IT solution since it was stabilised during 2007. Further rollout has both been conducted on an ad hoc basis (units can decide to join at any time they choose) and as two separate corporate projects, one in 2008-09 and another in 2009-10. These two projects were named "PA phase III" and "PA phase IV". These phases related to the first two phases as described in the present chapter, and thus illustrates that the original "phase III" term was never activated in its original meaning, which aimed much broader than improving the IT solution. The phase III IT project was a further rollout of the IT solution to areas that requested so. The phase IV IT project included a further rollout of the IT solution to additional requesting areas, particularly within R&D and the global sales organisation. Phase IV also included the introduction of new functionalities to support a high priority request from R&D to support matrix organisations, where the primary feedback providers would not be the line manager, but project managers.

Establishment of global minimum benchmarks

From 2010 a new set of "global minimum benchmarks" were established and rolled out through the global organisation. The rationale was that despite the fact that the corporate PA process had now been in operation for several years with significant efforts spent on the process throughout the organisation, and despite the increasing share of employees using the IT tool, there was still huge differences in the employee experience with PA across the global organisation. Rather than re-enforcing all elements of the process evenly, corporate HR decided to identify some critical focus

points that units could compare against. These were named "global minimum benchmarks" and were characterised by the following design criteria: They should be auditable so that it was clear what was required. They should be valuable, so that it was clear how the company benefited from the adherence to them. And they should require minimal extra work for line managers.

Five global minimum benchmarks were defined and for each, measuring opportunities were established. The five elements were: A) All employees should follow the PA process. The rationale for this element was to ensure that all employees (except where PharmaComp was bound by legal requirements, e.g. union agreements with blue collar workers) had business targets and that they received performance feedback at least twice yearly.

B) All employees should receive an introduction to PA. The rationale for this element was that in the past the only formal introduction of employees to PA in PharmaComp consisted of two quiz questions that new employees were made acquainted with through the onboarding process. Corporate HR wanted to secure that a basic employee understanding of the process was established with the present and future employees so that expectations could be set and so that it was signalled to all employees that PA is taken seriously in PharmaComp.

C) All managers should be trained in PA. This element was crucial since line management was the owner of the process and in practice had the tough role to identify, document and translate the performance of employees into feedback and ratings.

D) All management teams should calibrate their ratings. This element became relevant because it was clear that there were very different practices. In some units, calibration did not take place at all. In others, calibration took place, but in very different ways. Corporate HR saw calibration not only as a method to improve the transparency and fairness of rating distributions across different areas, but also as a learning opportunity for management, who could coach each other, learn from each other and even give input to the rating of each other's employee ratings.

E) PA appraisals should focus not only on outcome but also on how those results were accomplished (behaviour). This element is much in line with the strong emphasis on values that top management enforces and which is part of

PharmaComp's culture. Delivering strong business results is crucial, but it must not be at the cost of business ethics. Therefore, corporate HR wanted to make sure that all managers and employees had a dialogue on how performance targets could be combined with explicit targets for behaviour.

As the reader can imagine it is not a trivial matter to roll these five basic elements out to a global organisation with more than 30.0000 employees in 50+ countries and keeping the elements alive onwards with challenges like employee turnover, intensive growth in employee base and promotion of new managers and coupled with the fact that for line managers there basically are other and higher priorities than HR process compliance. All through 2011 and still into 2012, significant efforts were invested in corporate HR to anchor the global minimum benchmarks.

Increased statistically based knowledge and diffusion of use

Over the years since PA was implemented in PharmaComp, corporate HR has made different statistically based studies to investigate the PA practice across the organisation. I will go into more details on this in the following chapter. Here, it will suffice to mention that such analyses have produced interesting knowledge about the relation between rating and variables like geographical location, workforce, hierarchical position, bonus payout, retention ratio per performance rating, etc. Such studies have not only uncovered many examples of variation in practice across the organisation, they have also provided an educated platform from which to kick off new corporate initiatives like the above mentioned Global Minimum Benchmarks' initiative.

Another factor that has worked to increase the effects of the PA process has been what I would call the diffusion of use. As soon as ratings become available in broader volumes, diffusion to related processes starts mushrooming. This tendency becomes even more pronounced where the data can be automatically incorporated into an IT system, because reuse then becomes much more cost effective. This lowers the entry barrier for reuse compared to the value provided by the reuse. In PharmaComp, PA ratings are now available not only in the PA system itself but also in the corporate ERP system, in the corporate data warehouse, in the corporate manager self-service portal and in the corporate succession planning and talent management system. Most likely, this development of diffusion of use will continue

over time, so that PA ratings will be made automatically available to all processes where the data is experienced to provide value to the business.

4.4 Summary

This chapter introduced the history and process of PA in PharmaComp. PA started as a corporate project in 2004 with the aim to establish one uniform, global process to substitute the at least 25 local processes that existed at that time. The purpose of PA was to establish a common baseline for performance management across PharmaComp, to support the strategic focus on internationalisation and mobility, to improve performance data overview and alignment of processes, to secure compliance with the corporate management principles in all units and to create corporate transparency on individual performance.

PA implementation went through several phases, starting cautiously by defining and implementing a 3 stage, 8 steps paper based process in 2005, moving on to implementing a supporting IT tool in 2007 which in the beginning were aimed primarily at the top layers of the organisation. This was followed by a series of initiatives from 2008 and onwards that further rolled out and improved the IT solution and aimed at increasing the quality and global alignment in the practice of the PA process across the organisation.

Lately, corporate HR's focus concerning PA has been to rollout and nurture Global Minimum Benchmarks to secure that there is a solid, common and global foundation for conduction PA across PharmaComp.

5 The Macro View

The purpose of this chapter is to further present and discuss the state of affairs of PA in PharmaComp from a macro view to set the stage for analysing local adaption and meaning creation across workforces. Thus, in this chapter I will present some of the empirical data established during the study of PA in PharmaComp. Particularly, I will draw on a rich quantitative dataset included in the archival material to which I was kindly given access by the Corporate HR Rewards group. The quantitative data set made it possible to draw macro level conclusions about the relation between performance rating and a number of variables like workforce affiliation, hierarchical level, retention and more. Due to limitations in the data set, some parts of the analysis will only be presented at an aggregate level while other parts are broken down into the four workforces in focus of my study.

In the analysis of the data I will also draw on the interviews I conducted with representatives from corporate HR early in the study and connect some threads to qualitative interviews with managers and employees where these concern the macro level design of the process in each workforce. Besides helping me in the work to interpret the quantitative data, this material will help to illuminate some of the macro level work that has been done in parts of the organisation to handle various issues and challenges that local management has faced with the PA process. For this purpose, I will also draw on a number of other archival materials from HR departments outside Corporate HR to illustrate local policies within certain parts of the PA process. Finally, along the way I will relate the findings in PharmaComp to the scholarly research agenda as it was previously uncovered in my literature review.

The present chapter serves at least two purposes. First, the chapter will further build on the previous case introduction chapter to provide a richer and more detailed picture of the situational circumstances under which local adaption and meaning creation occurs, as it will be analysed in the following chapters.

Second, the chapter will serve to illustrate the point that there is no reason to believe that PharmaComp is in any significant way deviant from what existing research has already uncovered through multiple studies in other organisations regarding macro experiences with PA processes. As such, the case is exemplary and holds relevance beyond the boundaries of PharmaComp. This is also important because this

recognition was one of the reasons why I turned my attention towards local adaption and meaning creation.

The case organisation provided me with access to: a) an analysis report from 2012 of various interdependencies between performance rating and organisational impact (Halse, 2012) and; b) a data extract of all centrally and electronically stored PA ratings across the global PharmaComp organisation. The dataset covers a time series of PA ratings from 2006 to 2011.³⁶ The data volume and quality grows over the period as: a) more employees are covered by the PA process and b) PharmaComp's Corporate HR has strengthened its governance concerning collection and electronic aggregation of data. Hence, most of my quantitative analysis below will be made on 2010 ratings, where coverage and data quality is highest.³⁷ Thus, regarding 2010 PA ratings, it is reasonable to say that I am conducting analysis on what is essentially close to a full population dataset.³⁸ One advantage of this is that I can conduct the analysis by using descriptive statistics, without having to cope with the uncertainties of inferential statistics (e.g. sample representativeness and statistical uncertainty), as would have been the case with a randomised sample dataset. Another advantage is that I will have occurrences of all ratings in all workforces which would be difficult to get in a smaller sample since some ratings, particularly the lowest ratings, are rare.

The downside of using the full population data set is that I am not only analysing data for the Scandinavian part of the organisation in which I conducted interviews. Though the major parts of R&D and Staffs are located in Scandinavia, the situation is different for Production and Sales. Production has large sites outside Scandinavia and in Sales almost all personnel is located outside Scandinavia. Thus, some caution need to be taken concerning the conclusions made in this quantitative analysis since I have not controlled for possible geographical effects.

³⁶ Each rating corresponds to performance for the year before so that the rating received by an employee in 2011 concerns the employee's 2010 performance. For the remainder of this thesis, when using the term "2010 rating", it means the rating concerning 2010 but communicated to the employee in 2011.

³⁷ Typical data quality issues are: a) some units did not report their PA ratings in due time to Headquarter, b) some units reported incorrect data for some employees, or c) some units reported correct data but changed the rating after reporting to Headquarter without reporting the change. Particularly in the first years of the data set, issue (a) was present. All three mentioned data quality issues are rare regarding 2010 ratings.

³⁸ The 19,843 PA ratings registered for 2010 should be compared to a total number of more than 29,000 employees at the beginning of 2010. From the latter should be drawn employees who left during 2010, employees who were on leave and did not have any PA targets set for 2010 and employees generally not included in the PA process (e.g. blue collar workers). Hence the conclusion that the data set almost covers the full population.

Below, I will go through the macro analysis in seven steps. First, I will present how the overall number and rating average of employees included in the PA process and whose ratings have been reported to Corporate HR, has developed over the period covered by my dataset. Second, I will break down the dataset to the workforce level and analyse differences. Third, I will present and analyse findings regarding the relation between hierarchical level and performance rating. Fourth, I will present and analyse findings regarding the relation between PA rating and employee retention. Fifth, as the dataset did not include any information about salary increase and size of bonus, I will present and discuss the findings of a previous study in a part of PharmaComp that investigated this relation. Sixth, I will look into the policies for balancing business results against personal behaviour when determining the final appraisal of an employee. Seventh and finally, I will summarise the main points of the quantitative analysis and draw conclusions.

5.1 Performance Ratings at PharmaComp Over Time

In 2010, the PA ratings of 19.843 employees across the global organisation were electronically recorded by Corporate HR. As Figure 5-1 below shows, this is the culmination on a trend of growing numbers of recordings.

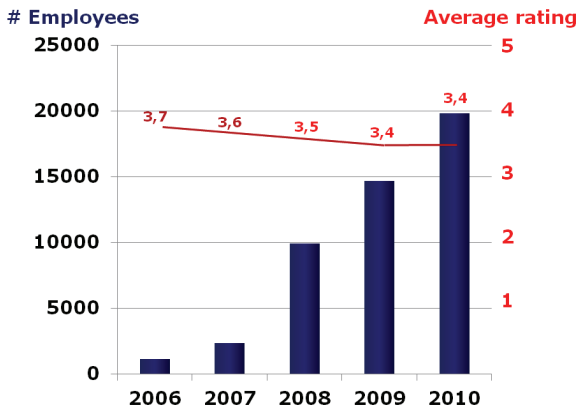


Figure 5-1: Number of employees in dataset and average PA rating, by year

Figure 5-1 shows the number of employees and the overall average PA rating per year for the period 2006-2010. The blue bars illustrate the dramatic increase in

number of employees included in the PA process and whose PA rating is electronically available for analysis due to improved data quality. The electronic PA system is one important data source. As previously explained, the PA process was introduced as a paper based process in 2005 while the electronic PA system was first introduced in 2007. Besides pulling data from the PA system, from 2008 Corporate HR also started to collect data from the paper based PA process from the global organisation. This was done by requesting the units to forward spreadsheets to Corporate HR with PA ratings for employees not using the PA IT system. Over the years this process has been capable of catching a growing share of all ratings and from 2010 the coverage has become really good.

The red line in Figure 5-1 illustrates the global average PA rating by performance year. We see a persistent decrease over the years from 2006 to 2009 where after the trend is flat to 2010. Below, I will show how the higher levels in the organisation systematically receive higher ratings than the lower levels. It is therefore no surprise that the average rating decreases as more and more lower level employees are added to the PA process and the systematic collection of ratings. However, over time the top layers in the organisation will constitute a still smaller share of the total population of ratings recorded, and this effect will be relatively smaller, which may explain the flattened curve from 2009.

5.2 Performance Rating Distribution by Workforce

Corporate HR is not enforcing or recommending any specific distribution of ratings across the 5-point rating scale: Outstanding (O), Exceeds Expectations (EE), Meets Expectations (ME), Approaches Expectations (AE) and does Not Meet Expectations (NME). However, many business areas define their own guidelines regarding rating distribution. For example, in one part of Staffs the following distribution is used as a guideline: O 0-5%, EE 5-20%, ME 50-75%, AE 5-20% and NME 0-5%. In another illustrative example, Production in 2009 analysed the rating distribution of 2008 and Production Management recommended a different distribution for the following year.

5. The Macro View

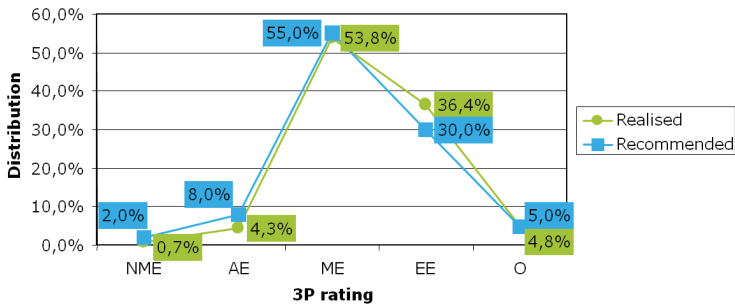


Figure 5-2: Recommended & realised rating distribution in Production, 2008/09

Figure 5-2 shows the recommended 2009 and realised 2008 PA ratings in Production. The figure shows several things. First, in 2008 Production rated half the share of employees into the two lower rating categories, compared to what Production Management recommended for the following year. So Production Management was not satisfied with the limited use of the lower rating categories. Second, the realised curve was somewhat twisted to the right in the diagram compared to the aspiration for the following year, so that the Exceeds Expectations category had more occurrences than preferred by Production Management.

Third, the figure also illustrates that Production Management was very precise in its distribution recommendation for 2009. Although PharmaComp does not operate within a 'forced distribution' paradigm (Grote, 2005), Production Management's recommendation should serve as guidance for calibration meetings in each business unit/functional sub-area.³⁹ Production Management found that the analysis had shown

...great differences in ratings which could not be accounted for in the department/function performance itself.

Management saw better alignment of performance ratings across Production as key to minimise rater bias (hard raters vs. soft raters), ensure a uniform understanding of

³⁹ According to a Production HR presentation, the recommended performance distribution for 2009 was "based on PA rating distribution in other areas in PharmaComp compared with overall performance in the area, historical performance in the area (excellent performance = right skewed distribution) as well as books/articles on performance distribution and performance calibration practices."

the rating scale and to improve the value of the process in terms of creating performance improvement. Management stated that the:

...recommended distribution is not forced but managers should use it for challenging/discussing direct reports and whether the rating distributions for each manager make sense based on overall performance. I.e. a low performing unit having more right skewed performance distribution than a high performing unit should probably be challenged.

The terminology used by Production Management is cautiously avoiding the term "forced distribution" (Grote, 2005) and instead use the term "expected distribution". But with a still stronger enforcement to meet the expected distribution, managers and employees sometimes see the difference as subtle.

One larger Staff area in 2009-10 conducted an initiative to improve PA. An analysis had shown that there was a significant gap between what senior management in the area termed an "ideal distribution" and the actual distribution of ratings, as illustrated in Figure 5-3 below.

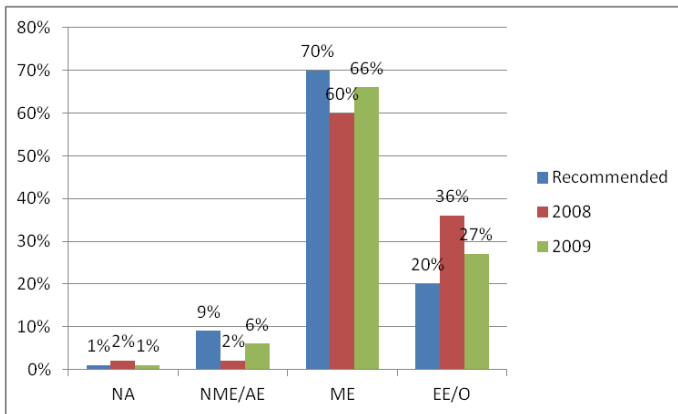


Figure 5-3: Recommended and realised rating distribution in Staffs area 2008/09

Figure 5-3 shows the recommended and actual distribution for the Staffs area in 2008 and 2009. The ratings have been grouped into three main clusters; the low-end ratings NME and AE, the medium rating ME and the high-end ratings EE and O. The analysis of the 2008 distribution had clearly shown that managers were very reluctant to use the low-end ratings while they at the same time were using the high-end ratings almost twice as often as senior management wanted. By focusing strongly on PA throughout 2009, management succeeded in bringing the actual distribution much

closer to the recommended distribution in 2009, although the tendency still was an underweight in low-end ratings and an overweight of high-end ratings.

In Sales, the process is somewhat different in that PA targets are primarily containing target indexes for realised sales. A certain percentage index realisation within one target is attached to a specific rating of the target. In theory, everyone could be Outstanding. However, that is very theoretical, since targets are adjusted quarterly based on sales forecasts. When a product has been in market for a while, forecasts are very good and it is hard for salespeople to significantly over perform. When products are new in the market, there is more uncertainty in forecasts, and salespeople can sometimes over perform with several hundred percent. Though, in the following quarters the forecasts will of course be adjusted again to level out the variation seen over the full year. For this reason, and because bonus and PA are not interdependent in Sales as they are in most other parts of PharmaComp, it has not been relevant to Sales management to establish strict recommendations regarding rating distribution.

In parts of R&D, management has gone furthest in aligning the relation between distribution and area business performance. Hence, based on the overall performance of an area, a pre-defined average rating with a precision of one decimal, and a predefined number of ratings within each of the five categories on the rating scale, must be met within the area. For example, in one area within R&D, management was told that based on the performance of the area, the average PA rating for all employees should be 3.7. Further, all units should make sure to hit a standard distribution across the rating scale within different job categories. In other words; a forced distribution approach. While this obviously provides management with strong control mechanisms in the rating process throughout the organisation, this would also be a controversial design in the view of scholars presented in the literature review who took a critical stand towards PA schemes, based on the argument that such schemes have a build-in tendency to develop to become bureaucratic and control focused (Piercy, 1994).

In Figure 5-4 below, the distribution of PA ratings across the four workforces is shown.

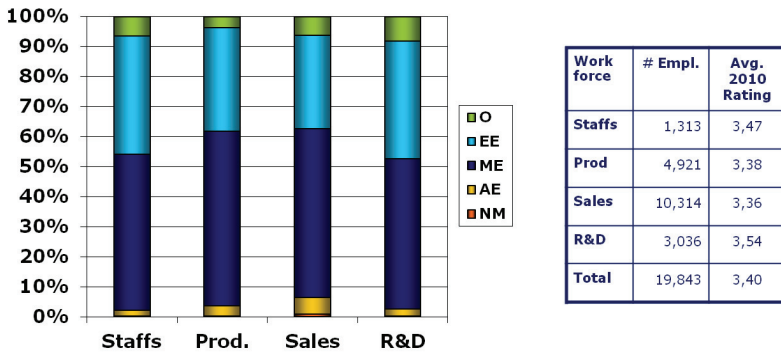


Figure 5-4: 2010 performance rating distribution by workforce

The table to the right in Figure 5-4 shows the number of employees and the average PA rating for employees included in the PA process within each of the four workforces. In 2010, R&D and Staffs had higher average ratings than Production and Sales. However, as the numbers are at a globally aggregated level, there may be significant local deviations hidden within each of the workforces.

The graph to the left in Figure 5-4 shows for each of the four workforces the proportions of employees that were rated within each category on the 5-point rating scale. In 2010 R&D rated almost 50% of their employees in the two upper categories *Outstanding* and *Exceeds Expectations*. Another interesting observation is that the lower two categories on the rating scale are rarely used, particularly in Staffs and R&D. Generally, 89% of all ratings in 2010, were either *Meets Expectations* or *Exceeds Expectations*, so although there are differences within each workforce, it is clear that ratings in the lower two rating categories or in the *Outstanding* category, are rare. Only 101 employees out of 19.483 got the lowest rating *Does Not Meet Expectations*. This corresponds to 0.5% of the employees and thus makes the lowest rating exceptionally rare in PharmaComp.

Since the start of PA in PharmaComp the central tendency bias has been a concern among those responsible for the PA process in Corporate HR. For example, HR saw an interest in improving the documentation foundation for dismissal cases and saw

PA as a straightforward place to produce such documentation when employees did not live up to management's expectations. However, this hope has generally been challenged, as one Corporate HR expert explained:

We wanted a better relation between rating and situations where managers wanted cases brought up for dismissal consideration. When we in HR were asked to support a dismissal case, I always started by asking the manager why she had given the employee an ME-rating and a bonus and now suddenly wanted to get rid of the employee? It was a tough way to try and train managers to be more consistent in their rating behaviour.

In one area of Staffs the central tendency bias was a key subject in a 2010-initiative together with Corporate HR experts to revitalise PA in the area. Area Management simply decided to enforce a broader use of the scale through a process that aimed at moving the weak MEs to the AE category:

We wanted to really push people into the 2 category. Because one of the problems was that a 3 means nothing if you don't have 2s. But then we also had to frame the 2s – in the past it was like if you get a 2 you are out of here. The 2 really has to be a 2. That you have a great opportunity for improvement. You have a performance plan, you get coaching... This means you got to improve and we are going to help you.

The process created a lot of discussion and some noise in the organisation, but it did increase the share of AEs in this particular area. One participant in the process recalls:

Basically the message was: Look at your people and see who is the worst one. He gets the 2. So, you know, they didn't like that. But on the other side, they found those 2s. I coached some of the managers before giving 2s – for some of them it was hard because it was the first time... We went closer to that ideal distribution – we were pretty much on target after a year.

In 2010 Corporate HR also tried to provoke a discussion based on statistical data on issues with the current use of the rating scale in areas in R&D, but here the audience was less attentive. Contrary to the Staff area that set a revitalisation initiative in motion, the R&D areas challenged the data reliability and questioned the relevancy.

5.3 Performance Rating and Hierarchy

The dataset also allowed analysis of the relation between hierarchical level and performance rating.

5. The Macro View

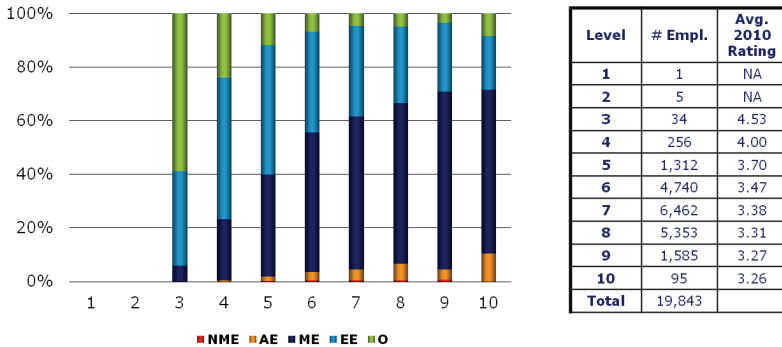


Figure 5-5: 2010 performance rating distribution by hierarchical level

Figure 5-5 shows the 2010 PA ratings distributed over the 5-point rating scale. The first column in the table at the right of the figure shows ten different hierarchical levels. Level 1 is the CEO, level 2 is the EVP level, and so forth. The number of hierarchical levels differs depending on which part of the organisation one looks at. In rare cases there can be up to 10 levels, but the most commonly found is 6-8 levels. Column two in the table lists the number of employees at each hierarchical level. Column three in the table lists the average PA rating at each level. PA rating data is not available at hierarchical levels 1 and 2.⁴⁰

Linear regression analysis shows a strong correlation ($r^2=0.81$) between hierarchical level and average performance rating. Level 3 employees on average get 4.53 in performance rating and the average rating decreases proportionally with the hierarchical distance to the top level of PharmaComp. Employees at level 10 got an average PA rating of 3.26.

The graph to the left in Figure 5-5 shows the distribution of ratings at each of the hierarchical layers (again, without levels 1 and 2). Considering what we just concluded about the average ratings correlation to hierarchical level, it is not surprising to see how top ratings are more and more common the higher in the hierarchy we look. For example, 60% of the employees at level 3 got the highest rating (O) and only one of the level 3 employees got the medium rating (ME), which was the lowest rating given to any at level 3. Not until we reach level 5 and beyond

⁴⁰ Levels 1 and 2 are absent due to confidentiality reasons.

do we start to find any in the lowest category (NME), although this rating is so rarely used that it is hardly visible in the graph at all for any level.

If we compare the graph in Figure 5-5 with the performance rating distribution used by Production in 2008 mentioned above, we can see that none of the hierarchical levels meet the Production recommendation of having 2% in the lowest rating category (NME). In fact, even for the second lowest rating category (AE) it is only at level 10 we see that the 8% "recommended" rating distribution in Production would have been met if it had been applied company wide. On the other hand, for hierarchical levels 3 to 5, we see that top ratings (O or EE) are given to 60% or more of the employees.

The question is in which direction the causal arrow points. Are people getting higher ratings because they are nearer to top management? Or are they nearer to top management because they are high performers that get high ratings? Most likely the causal arrow is pointing in both directions. On the one hand, it is trivial to conclude that the chance of being promoted to higher levels in the organisation increases with the level of performance and performance rating. That was a key point in installing the process in the first place and the reason why performance ratings are mandatory gate criteria in talent programs and used in promotion cases. On the other hand, the performance rating is related to the degree to which a person lives up to the expectations that PharmaComp has to someone working in a specific position. It was never the intention that the performance of an SVP at level 3 should be compared to the performance of an IT help desk supporter at level 9. An SVP's performance should be evaluated against the high expectations to an SVP, the IT help desk supporter should be evaluated against the more modest expectations set for an IT help desk supporter. So, why is it that SVPs are generally outperforming the expectations set to them, but employees at lower levels generally do not?

One Corporate HR informant suggested that the strong corporate culture within the higher management layers of "always keeping all KPIs green"⁴¹ may have as a side effect that these levels are capable of negotiating targets that are generally only moderately ambitious.

⁴¹ A "green" KPI means a KPI that has been fully met.

Alternatively, we may look for possible explanations identified in the literature review. Some important rater bias that was identified in the literature was interpersonal effect bias and bias caused by manipulation, impression management and politics (Varma et al., 2005. Cook, 1995. Poon, 2004). Generally, it would be reasonable to assume that higher ranking managers have more potent opportunities for leveraging their political and personal power base to affect their managers to give them higher ratings, than do lower ranking employees. Also, while there is no corporately enforced performance rating distribution which would affect the distribution of ratings of SVPs, I have already shown how the layers below have in many areas over the years taken initiatives to control the rating distribution within their own area. So, both because of reasons already indicated in previous studies and because of PharmaComp specific circumstances, it is no surprise to see that higher ranking employees receive higher performance ratings, as Figure 5-5 so clearly demonstrates.

It may be in line with the expectations of PharmaComp's corporate HR that higher hierarchical level naturally get higher performance rating. The company is very successful and high ratings generally seem to be appropriate for that reason alone, if for nothing else. But it still is an interesting background for the meaning creation processes of managers and employees lower in the organisation.⁴² Employees may ask themselves why targets and expectations are systematically outperformed at the higher levels while targets and expectations seem to be more realistic in the lower end of the hierarchy. One HR informant was concerned about this lack of vertical congruence in performance rating distribution:

For this [PA] to work well, we need to be consistent horizontally and vertically. Particularly regarding the latter, we aren't there. It creates this notion in the top layers that "there are special rules for us", while further down in the hierarchy, everyone is an ME. But they have another job description you know, and of course rating distribution should be the same down through the hierarchy.

Employees may also wonder why there is talk about normal distributions and calibrations to avoid inflationary tendencies at the floor level, when those concerns seem to be non-existent when it comes to ratings at the top levels. In this respect, the hierarchical dependency aspect fits nicely into the discussions about fairness

⁴² That is, of course, to the degree managers and employees are aware of the correlation between level and rating.

perceptions, top-down dominance in the process, difficulties with realism and ambition levels in target setting and other general issues identified in PA schemes that were discussed in the literature review (Thurston and McNall, 2010. DeNisi and Pritchard, 2006. Lee and Son, 1998. Coens and Jenkins, 2002. Austin, 1996).

5.4 Performance Rating and Retention

Based on a 2011 Corporate HR analysis it is possible to compare retention ratios by performance rating over the first year period after employees had received a certain rating. The comparison is depicted in Figure 5-6 below.

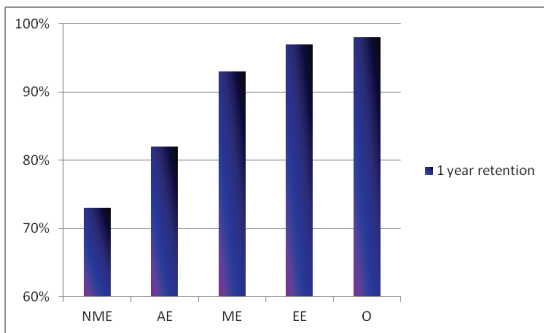


Figure 5-6: One year retention ratios per performance rating, 2008 ratings.

The figure shows the retention aggregated across all workforces and geographies and includes all available PA ratings from 2008. Linear regression analysis shows that the correlation is strong ($r^2=0.92$). While 98% of the 2008 employees who received the highest rating (O) were still with the company one year after the rating was conveyed, the same was only true for 73% of the employees who received the lowest rating (NME). The graph shows that there is a rather high attrition rate within the first year after having received the lowest two rating categories.

It can of course be discussed what these numbers mean to PharmaComp. One might question if it is satisfactory that 2% of all outstanding employees have left the organisation one year after receiving their O-rating? Or, if it is satisfactory that 73% of all NME rated employees and 82% of all AE rated employees are still with the company after one year?

The first question is the easiest to answer since PharmaComp has a stated balanced scorecard target for keeping highly talented employees from leaving the organisation. The target is that the retention rate of high performers, defined as employees rated as outstanding, must be at least 95% year by year. Seen in this light, 2% attrition over a one year period is acceptable to PharmaComp.⁴³

Whereas high retention is always positive concerning high performers, it is less obviously so for lower performing employees. Generally, PharmaComp distinguishes between wanted and unwanted leave, and unwanted leave varies a lot by geography, in line with the experiences of other companies. E.g. India has a high attrition rate compared to Europe. Thus, the second question about whether PharmaComp would prefer a higher or lower retention rate among employees with low performance rating is a more difficult one. There are no explicit targets for how many of the employees in the lower rating categories PharmaComp would prefer left the organisation. Nevertheless we can get an impression by the below statement from an HR representative:

In the past it was like if you get a 2 you are out of here. The 1s – you are still out of here. But the 2 now became a 'needs improvement'.

So getting a "1" (NME) according to this informant is basically a ticket out of the organisation. However, as we can see from the numbers, this does not hold true for three quarters of the employees rated NME, at least seen over a 1 year period after the rating has been provided.

According to the definition of the rating scale introduced in the previous chapter, the second lowest rating AE contains the wording "performance should be improved". In most units the AE rating is followed by an action plan that the involved manager must evaluate at latest before the next appraisal, to investigate if the employee is the right person for the job. As already shown, the lowest NME rating is very rare. The terminology of the NME definition is also quite harsh:

Contribution to company performance is not acceptable. Fails to reach goals and behaviour is poor or negative. Has not responded to corrective action.⁴⁴

⁴³ The attrition rate of O-rated employees was 2% the first year, 6% (accumulated) the second year and 8% (accumulated) after 2.5 years.

⁴⁴ Corporate HR PA rating scale definition.

This terminology taken into account, a retention rate of 73% over a 1 year period (the number is 61% of all NMEs if we observe over a 2.5 year period) in fact seems surprisingly high. However, the number does not say anything about whether some of the employees still with the company have changed position to lower ranking, less demanding positions that the employees may be able to fill in a more satisfactory way, or if the employee have improved and received a higher rating the following year.

If the numbers are broken down to a regional geographic level, some differences between the areas start to pop up that may provide input to possible elaboration. There seems to be an effect of differences in local management culture on the tendency to get rid of low rated employees. For examples, the data shows that while only 42% of all 2008 AEs and NMEs in China were still with the organisation after 2.5 years, the same figure for Japan, where mutual loyalty between employer and employee is a stronger cultural trait, was 82%. However, since the numbers particularly in the NME category, but also for some regions in the AE category, start to get really small when broken down to this level, statistical validity gets smaller. Because of this, and because it is not the objective of my study to get into any deeper analysis of the question of retention of low performers at the macro level, I have not pursued this path any further.

5.5 Performance Rating and Wage Increase & Size of Bonus

The dataset did not include any information on wage increases and bonus. However, in a previous study of two geographies in PharmaComp (Headquarter and North America) Halse, Smeets and Warzynski (2011) concluded, that there were positive relations between performance rating and both wage growth and size of bonus. The authors concluded that there were differences in the correlation between geographies, although the relation was neither consequential nor static:

Higher evaluation is positively related to higher wage growth, but the sensitivity is much higher outside Europe.

And;

Higher performance is associated with higher bonus, but the sensitivity is different between countries []. Interestingly, the sensitivity has diminished quite a lot in the US, while it has increased in the domestic country. (Halse et al., 2011: 10)

Unfortunately, the authors do not go into further analysis of the reasons behind the geographical differences or the dynamics of the relation. One reason could be that North America generally had a larger proportion of variable pay than Scandinavia. And thus also the ability to significantly differentiate payouts based on performance. Another reason for the difference could be that the compensation processes has in the past been much more streamlined and standardised in PharmaComp's North American region than in Scandinavia. This has lately been a focus area in the Scandinavian organisation, where more emphasis has been laid on 'fact-based' calibration sessions where performance ratings play an important role. Both reasons relate to previous studies mentioned in the literature regarding the relation between performance ratings and rewards and the importance of cultural and situational factors (DeNisi and Pritchard, 2006. Woods, 2003. Kohn, 1999. Coens and Jenkins, 2002).

One Corporate HR informant explained how the link between PA rating and rewards has increased the challenge of getting managers to use the lower end of the rating scale:

Since the link to rewards has become more explicit,⁴⁵ there have been many more fights between managers and employees in cases where an employee is near the borderline between an ME and an AE rating. You are not only saying AE - you are also saying zero bonus.

But it is not only the managers and the employees who see the increased challenge affected by the more explicit link between performance rating and reward. One Corporate HR expert explains:

Maybe we went out a little too fast with promotion of the link between rating and bonus. I have sometimes asked myself if it is worth the efforts. All the energy that we put into rating, calibration, run-back, etc. It creates much more pressure on managers and on the after all limited [bonus] funds we have. Before, the bonus process was much more mechanical based on payroll.

The Sales organisation is somewhat different from the other three workforces since Sales is not coupling PA rating and bonus directly. In Sales, bonus is allocated based on sales targets and on a quarterly basis. Since sales targets form the main part of sales representatives' PA, although not the entire PA, there is of course a relation

⁴⁵ The link between PA rating and Reward started to emerge on the Corporate HR agenda from 2007-08 when data availability made it possible to document the challenges in current practice.

between the sales performance and the PA rating. But it is not the PA rating as such that is used to allocate bonus.

In other areas, PA ratings are important input parameters to the bonus calibration meetings held by local management teams. As we shall see later, there are parts of PharmaComp where bonus is calculated automatically based on a combination of performance rating and job category. So if there had been no correlation to be found between rating and bonus in PharmaComp, this would generally have been very surprising.

For the purpose of the present study, it will at this stage suffice to conclude that there is a correlation between performance rating on the one hand and sizes of wage increase and bonus on the other. This forms part of the background against which local adaption and meaning creation takes place.

5.6 Performance Rating and Corporate Culture

PharmaComp has a long tradition for working with target setting, measurement and evaluation, not just within people performance management but also in a broader context. This has to do with a least two fundamental aspects of being an innovation-driven company within the knowledge intensive pharmaceutical industry.

First, the research and development process can best be illustrated as being organised as a funnel of evaluation gates where a large number of new ideas through the discovery and development process is continuously evaluated for stop/go decisions. In the end, it is only a very small fraction of all ideas that make it to becoming new products on the market. Bringing ideas through the entire funnel to full-scale product launch can easily cost hundreds of millions of dollars and typically takes no less than twelve years from the first and initial idea to product marketing approval by the authorities, e.g. the FDA in the US. By far most of all ideas will be discontinued along the way through the funnel. The further a project moves through the funnel before eventually being discontinued, the more resources the company will have spent without having generated any revenue. Put simply, the sooner non-viable ideas are discontinued, the better for the company's performance, since fewer resources will have been wasted. Therefore, setting clear targets and tough evaluation criteria is of key importance in this process. This strictness in evaluation is

part of the corporate DNA in many innovation-led companies, and in PharmaComp it has over time diffused to areas outside R&D so that similar processes have become dominant also there. This is the case particularly within the area of project governance across business areas. Similarly to the gates in the R&D funnel, stage-gates in the project governance methodology segregate the better projects from the less promising. Similarly, the PA process can be seen as a target definition/evaluation/consequence funnel for segregation of talent.

Second, the pharmaceutical industry is a highly regulated and compliance driven industry. To be able to take products to markets, companies must adhere to very strict procedures and documentation standards that are continuously audited and evaluated by external and internal auditing bodies. The ability to comply with legal regulations and industry standards set by such bodies is a fundamental requirement for being in business at all. Thus, the entire part of the organisation that is involved in the development, production, distribution and sale of a pharmaceutical product must be available for an on-going monitoring, inspection and evaluation by internal and external auditors who will adhere to strict approval criteria. This affects almost every single employee and manager in PharmaComp either directly as they or their departments or projects are inspected, or indirectly as they are affected by the outcome of such inspections. Thus, compliance monitoring and evaluation of ability to comply with defined standards permeates the entire PharmaComp organisation and culture.

The importance of evaluation and segregation of successful ideas, projects and talents is also high on the agenda for executive management in PharmaComp. They have issued 'The PharmaComp Way' which is an important document that describes "who we are, where we want to go and the values that characterise our company". In short, "The Way" contains executive management's perception of what is the core company identity differentiating it from competition. Thus, "The Way" is intended to support the organisation in being able to maintain a coherent corporate identity and culture across geographies and units in an accelerating globalisation process. Therefore "The Way" is mandatory reading for anyone employed in PharmaComp and significant resources are used on training and examining managers and employees worldwide in their understanding and fulfilling of "The Way".

Included in “The Way” are ten so called ‘essentials’ that define the key characteristics of how Executive Management wants employees to work and behave. No less than two of these essentials are directly relating to PA. The first is “We set ambitious goals and strive for excellence”. This essential holds the following interpretation guideline; “Unit and individual goals are ambitious, clear and understood and link to PharmaComp’s overall goals. Business challenges are identified and addressed with short and long term performance indicators being clearly defined and monitored. Goals are dealt with by balancing competing demands to ensure sustainable business success. The unit strives for excellence by defined benchmarks or other relevant measures. And, a systematic and transparent performance management process is in place.”

The second essential dealing directly with PA is “We focus on personal performance and development” and holds the following interpretation guideline; “Individual measures of performance and expectations of behaviours are clear. Employees receive clear, specific and timely feedback on performance. Development opportunities for individuals are identified based on current competences, performance and company needs. All employees address their relevant improvement needs and development opportunities.”

By letting PA play such a key role in “The Way” executive Management sends a strong signal to all managers and employees about the importance of PA, and how it is an integral part of working in PharmaComp that one engages into the PA process.

In conclusion, employees in PharmaComp are used to measurement and evaluation in many dimensions; as part of units being evaluated on their collective performance, through the R&D idea and project funnel process, on projects in general through stage-gate evaluations, and personally through the PA process. In this way measurement and evaluation have become integral parts of the PharmaComp culture and the way ‘things are done here’. People are being measured and evaluated on a daily basis in multiple dimensions individually and collectively, to a much higher degree than would likely be the case in other organisations less driven by innovation, compliance and projects. This is an important part of the PharmaComp culture and likewise an important element to keep in mind when analysing how managers and employees make sense of their experiences from the PA process.

5.7 Performance Rating - the Balance of Results and Behaviour

One of the issues with PA that was identified in the literature review is that the process relies on extrinsic targets that has as a side effect that they divert employees' focus from adding value to reaching targets (Bowen, 1995. Kohn, 1999). Countering this negative side effect by including the behaviour of the employee as a supplement to the performance rating of business results has been a topic for discussion in PharmaComp since the introduction of the PA process. An HR informant explained that behaviour was actually the key problem in most cases where low ratings were considered:

The situations where I have encountered issues with employees it has usually been attitude rather than competencies that was the root cause. Competencies can be lifted quite easily through training. It is more difficult to change behaviour.

Not until 2012 was there any corporate guideline for how to do this. But locally in the organisation there have been initiatives to create more structured approaches to the balance between business results and behaviour in relation to PA appraisal. Particularly within Production, work has been done since 2008 to improve the practice of appraising behaviour by providing central guidance to Production managers. For example, in 2008 Production Management communicated that:

It is recognised that there are two dimensions of performance – results and behaviour (or what you deliver and how you deliver it) which both need to be balanced in a PA rating. Results are more easy to measure whereas shown behaviours to a greater extent will be based on management judgement.

So, according to Production Management, not only are there two fundamentally different dimensions on which appraisals need to be considered. One of them, the behaviour part, is also considered more difficult and more subjective than appraisal of results. This statement from Production Management is very interesting because it can be seen as a clear recognition of the challenge to the objectiveness in PA schemes.

The importance of this recognition of course depends on how much weight behaviour was intended to carry relative to business results. Therefore it is interesting that on the same occasion, Production Management introduced a model that managers in Production were requested to use when balancing behaviour with business results. The model is shown in Figure 5-7 below.

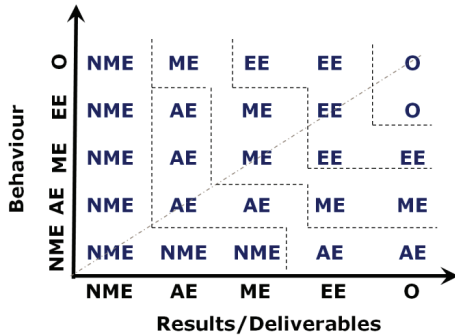


Figure 5-7: Results and behaviour converted to PA ratings, Production

Figure 5-7 shows the graphic guideline to Production managers. The guideline operates in two dimensions; behaviour and results. Conceptually, the model suggests that each employee is rated on the corporate 5-point rating scale separately in the two dimensions. Subsequently, the model can be used to consolidate the two dimensions into one final appraisal rating. For example, if an employee earns an EE rating on business results, the final rating can end anywhere between AE (if behaviour is rated NME) to EE (if behaviour is rated ME or higher). The model is constructed so that business results weigh more than behaviour. For example, if business results are NME, then the final rating should be NME no matter the behaviour. And if business results are rated O, then final rating should be O even if behaviour is "only" EE. Still, the model introduces the evaluation of behaviour as a powerful moderator of business performance in determining the final rating of an employee in Production. Thereby the entire measurement logic gets subjectified in the sense that it opens up for interpretation and varied practices which are not directly connected to tangible targets and measurements.

To further assist the managers in giving the right rating to their employees, Production HR also established a guideline for how to use the 5-point rating scale in the two dimensions. The guideline is summarised in Figure 5-8 below.

5. The Macro View

Rating	Outstanding	Exceeds Expectations	Meets Expectations	Approaches Expectations	Does Not Meet Expectations
Results/Deliverables (Best used for assessing performance against key goals/job responsibilities)	Distinguished Extraordinary Exceptional Outstanding Role Model	Superior Commendable Very Good Well Regarded	Competent Good Solid Desired behaviour	Inconsistent Improvement Need guidance Adequate	Unacceptable Unsatisfactory Ineffective
Behavioural Frequency (Best used for assessing performance against competencies)	Always Invariably	Consistently Regularly	Often Usually	Sometimes Occasionally	Rarely Never

Figure 5-8: Production guideline for evaluating business results and behaviour

Figure 5-8 shows terminology associated with each level on the 5-point rating scale in the two dimensions; business results and behavioural frequency. Note that the terminology for behaviour is focusing on the frequency of the desired behaviour. For example, to exceed expectations in the results dimensions, terms like "superior", "commendable", "very good" and "well regarded" must be associate. In the behavioural dimension, getting an EE requires that desired behaviour is demonstrated "consistently" and "regularly".

There have not been similar initiatives to establish manager rating guidance for the entire R&D, Sales and Staffs areas for the 2011 cycle. However, I have come across several examples of local guidelines in units within Staffs, Sales and R&D. For example, one unit in Staffs set up four key elements for individual evaluation; individual business target realisation, team target realisation, customer focus and "the way we work". The latter category was capturing the behavioural dimension and was further broken down into six key terms: positive attitude, respectful, professional, change prepared & flexible, initiative & drive, and knowledge sharing.

In 2010, Sales also applied a model to guide managers in balancing behaviour and results when determining final appraisal rating. While Sales copied the terminology guideline for appraising in the two dimensions from Production as illustrated in Figure 5-8, Sales Management twisted the conversion matrix a bit, as illustrated in Figure 5-9 below.

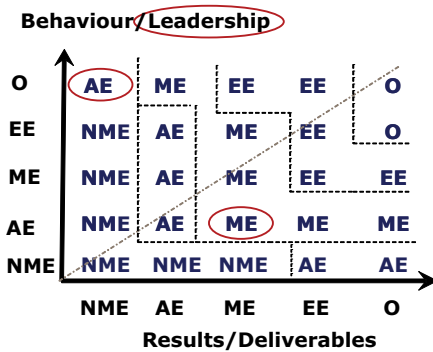


Figure 5-9: Results and behaviour converted to PA ratings, guideline for Sales

Figure 5-9 shows the Sales conversion matrix for aggregating behaviour and results ratings into a final rating. Although it looks a lot like the model used in Production, the red circles indicate where there are differences. First of all, Sales has added the term "leadership" to the behaviour dimension, probably to explicate that initiative and drive are key elements in high performance behaviour for a sales representative who works very independently, every day on the road to visit doctors and hospitals. The second change is that the impact of behaviour has been changed a bit in Sales compared to in Production. First, employees who do not meet business targets but show outstanding behaviour will not get the lowest NME rating as in Production, but the second lowest rating AE, which in practice is an increase in impact of behaviour. Second, employees who meet their business targets (ME) but only approach the desired behaviour (AE) will still get the ME rating, whereas in Production they would get AE. This is a reduction in impact of behaviour compared to Production. Thus, the changes are somewhat ambiguous and basically reflect a slightly different weighting of behaviour and results in Sales Management than in Production Management, but without changing the overall picture, which is that behaviour is a powerful moderator of results when determining appraisal rating.

With the 2011 update of PharmaComp's defined corporate values, renewed emphasis was put on the significance of PA. In 2011 Corporate HR started a project to communicate the new global PA minimum benchmarks as described above and to re-enforce good PA practices across the organisation. In this context, the significance of making the behavioural evaluation more transparent was communicated as an

element in a global communication and training effort. All units were recommended to include behavioural targets into the employees' PA. It is up to each unit to decide which behavioural targets they want to emphasize in their area.

Again, so far it has been in Production that the most standardised approach has been utilised, as Production decided to set two pre-defined behavioural targets valid for all employees included in the PA process across the global Production organisation. The two behavioural targets are mandatory for all, but it is up to local management to decide if they want to add more behavioural targets. The two mandatory behavioural targets for 2012 in Production were:

We optimise the way we work and strive for simplicity. We work according to the standards, and we solve problems systematically. In this way we ensure continuous and sustainable improvements.

We never compromise on quality: We always put the patient first and never compromise on quality. We see and communicate problems and potential quality issues and act timely. When in doubt, we elevate. We share failures as well as successes, to learn.

In other workforces, behavioural targets have typically been implemented at local initiative lower in the organisation and thus with greater flexibility in terms of local possibilities to decide what behavioural targets to include. However as a result of the Corporate HR minimum benchmark initiative, there is a clear tendency from 2012 towards more centralised and standardised approaches. For example, in one part of Sales where behavioural goals had so far been more or less unknown, the following three behavioural targets were cascaded to all sales representatives for the 2012 PA cycle.⁴⁶

Customer-centric: Focuses activities on how to create value for our customers. Builds long-lasting and mutually beneficial relationships with customers based on understanding their needs and on offering appropriate solutions that increase satisfaction and strengthen PharmaComp's position and reputation. Note: Customers include patients, health care professionals, payers, governments/authorities and other stakeholders.

Focused on what matters: Focuses on priorities for business growth and applies them, making choices and working efficiently and effectively, minimising bureaucracy. Takes initiatives to improve ways of working and shares best practices with others, optimising the use of resources and monitoring progress.

⁴⁶ The idea was that the each sales team in the area should then select two of these to focus on through the 2012 performance cycle.

High-Performing: Strives to achieve ambitious goals in line with PharmaComp's strategy and feels accountable for results, whether individual or team. Develops own and others' skills to enhance performance. Focuses on contributing to the business and plans work to ensure excellence and strengthen PharmaComp's leadership.

Also throughout Staffs and R&D behavioural targets are from 2012 more commonly defined and at higher levels, although of the four workforces it is still Production that has gone through the most consistent standardisation.

5.8 Summary and Conclusion

In continuation of the previous chapter where PA in PharmaComp was introduced, the present chapter has sketched out important elements of the foundation upon which local adaption and meaning creation takes place. The analysis was made based on archival material in the form of presentations and communications from corporate HR and business units, interviews with experts in corporate HR, quantitative studies made by corporate HR and analysis of a data set covering all electronically stored PA appraisal ratings in PharmaComp over the period 2006-2010.

The analysis showed how the share of PharmaComp's employees engaged in the PA process and the corporate availability of data has grown dramatically from only covering a few thousand from the top layers of the organisation in 2006 to cover almost 20.000 final ratings for 2010.

Data documented how all four workforces struggle to use the entire rating scale. The rating categories O, AE and NME are so rarely used that almost 90% of all ratings fall into the ME and EE categories. It was also documented how local management teams have tried to react to this by establishing rating distribution recommendations that aim to increase the number of employees particularly in the NME and AE categories. There were also differences identified in the rating distribution between the four workforces, so that R&D and Staffs had more employees in the O and EE categories than did Sales and Production.

It was documented how hierarchical level and performance rating correlates, so that hierarchical level is in fact a quite strong predictor of performance rating. It was discussed how this may affect the justice and fairness perception of employees and

form part of the foundation for meaning creation processes of managers and employees in the lower end of the hierarchy.

A strong correlation between performance rating and retention indicated that high rated employees stay with the organisation while the tendency to leave (voluntarily or involuntarily) grows reverse proportionally with performance rating. It was discussed if current levels of retention were satisfactory seen from PharmaComp's point of view, particularly why such a large proportion (three quarters) of the lowest rated employees were still with the company a year after having been rated NME.

Referring to a study by Halse et al. (2011), it was concluded that there is a correlation between performance rating and wage increase and bonus allocation. Since performance rating is a key input parameter in local management teams' wage and bonus calibration sessions, this was not surprising. But it still forms part of the foundation for meaning creation in that it emphasises that it does matter what rating one gets, also financially.

I have argued that the PharmaComp mode of operation also outside people performance management was one characterised by widespread and frequent use of measurement and evaluation. This has to do with the nature of having a structured approach to managing innovation and project processes as well as it has to do with the nature of the pharmaceutical industry being so intensely inspected, audited, monitored and evaluated by internal and external authorities in ways that have a direct impact on the work life of most employees and therefore also on the corporate culture. In PharmaComp, executive management has explicitly selected people performance management to be at the heart of no less than two out of 10 "essentials" that are intended to spell out the essence of what characterises and differentiates PharmaComp.

Various attempts to balance business targets against personal behaviour when appraising employees were discussed. Particularly in Production a structured effort was made already in 2008 to align the approach across the entire area. More recently, the approach has been extended with identification of two behavioural targets mandatory for all employees in Production and tools that support the managers in their evaluation and feedback roles. Also in other areas, behavioural targets are coming to the agenda, so that e.g. parts of Sales, R&D and Staffs from

2012 will operate more or less centrally defined behavioural targets that will be cascaded to all employees within certain areas.

The present chapter has indirectly demonstrated why Management and Corporate HR values the increasing diffusion of the PA process and the increased ability to aggregate rating data, since this can be used as documentation of the quality of the HR work done in the organisation, in the form of e.g. key performance indicators on key talent retention, identification of areas with deviating rating distributions or identification of units with mismatch between level of business performance and average performance rating.

The empirical findings in the form of correlations, bias, issues and challenges fit nicely into the debates that have been going on in academic journals about PA over the last 20 years. In that sense, the above findings in PharmaComp are interesting input into existing scholarly debates but it does not break much new ground. PharmaComp is generally facing the same issues and challenges as other big companies face with their PA schemes. Those issues and challenges that my literature review has documented have been dealt with from so many different empirical, methodological and theoretical angles. This, however, is not diminishing the value of the present study. On the contrary, exactly because PharmaComp faces similar challenges with PA as do so many other organisations, PharmaComp is an exemplary case where we can expect that analysis of local adaption and meaning creation processes will also have relevance beyond the PharmaComp organisation.

6 Meaning Creation and the Task at Hand

In this chapter I will introduce the analytical toolbox I will use in my analysis of local adaptation and meaning creation in PA. I will explain how sensemaking theory (Weick, 1995) is a particularly relevant theoretical framework for studying meaning creation in PA across workforces, because it directs attention towards how specific environmental circumstances, through enactment, affect individuals' translation of past actions and thereby form the stories they create to make meaning of their experiences. I will also dig further into the differences between the four workforces that form the object of my study. I have called this section "The task at hand" because it concerns how the four workforces differ in terms of their work tasks and how this in turn can be expected to affect meaning creation in PA and each workforce's expected fit to the company PA scheme.

In some of the most celebrated writings on sensemaking, Karl E. Weick has used the perspective to analyse extreme situations characterised by life and death drama, intense pressure and dramatic change. For example, he takes departure in the Mann Gulch disaster (Weick, 1993), the Columbia Disaster (Weick, 2005), severe over-death rates in medical care at the Bristol Royal Infirmary (Weick & Sutcliffe, 2003), the Tenerife flight disaster (Weick, 1993) and the recognition of severe physical child abuse in the 1950s (Weick, 2004).

This may partly be because such dramatic settings appeal to the researcher's interests. But it is also because sensemaking is most extreme, dynamic and visible in high intensity situations where existing meaning structures are most challenged. The Mann Gulch disaster is a good example of how sensemaking processes occur when the situation quickly develops in a way where training, experience and genuine perception of self no longer fit the circumstances. The Mann Gulch story is also a good illustration of how sensemaking is not a rational process and it is an example of how action fuelled by intense sensemaking processes can be fatal. Sensemaking materialises through action and is generally about how the individual and the group create meaning out of a situation. However, sensemaking does not infer that this 'meaning' would be particularly smart from a rational point of view.

While the extreme cases of Weick are very pedagogical because they put sensemaking at its edge in life or death situations, in this study I will use

sensemaking theory as a tool to analyse 'ordinary life' in an organisation. Weick is not against the use of sensemaking analysis in more ordinary contexts, he just does not practice it a lot himself. However, Weick does now and then leave small hints to where we should look for sensemaking in more mundane situations. For example, Weick (2009: 140) suggests we look "where everyday situations sometimes present us with either too many meanings or too few". As we shall soon see, this description fits nicely with what we see when we dig into meaning creation of managers and employees in the PA process.

In my view, meaning creation through sensemaking is a daily phenomenon; it is an unavoidable element of social life. It takes departure in how individual's frames of mind interplay with each other and how people develop sensible stories that bring meaning into their often complex and confused realities. Using sensemaking theory as a tool to analyse daily operations of organisations will lack the drama and the catch-eye of the more extreme cases, but it will also help to broaden the scope and value of sensemaking theory as an alternative perspective on life in modern organisations.

Sensemaking as a theoretical approach is a constructionist view on social reality (Weick, 1995: 36). When organisations are analysed through a sensemaking perspective, one is not looking for rational decision making or empirically observable cause-effect relations. Rather, the processes of making sense have much more to do with the on-going process of inventing images of social realities that make sense to the individual.

In sensemaking, action comes first. Sensemaking is by and large a hindsight activity where people make sense of their environment, the actions of others and not least their own actions. Through sensemaking, people do not only translate or interpret 'what really happened'. Rather, they invent an image of what happened, but not an arbitrary one; they invent the image that is most meaningful to them. Other people with other frames of reference could create another story of the same event, simply because it would be more meaningful to them. In a rational sense, both stories cannot be true. But in a sensemaking perspective, both stories are true in the sense that the actors perceive them as such and act accordingly as if they were, so that the stories again produce true effects on their continued behaviour.

An example of how sensemaking can be used to analyse social reality is Garfinkel's study (Garfinkel, 1967 in Weick, 1995: 10) about the process by which jurors work. What Garfinkel found out through an analysis of a concrete example was that jurors in their mental process first decided the verdict and after that focused on how to make sense of that verdict by developing the 'facts' required to justify the verdict. Thus, action comes first and sensemaking starts shortly after, even when the concern is one's own actions. This is what Weick has formulated in a famous quote: "How can I know what I think till I see what I say?" (Weick, 1995: 12).

For a PA system to remain legitimised, it must be possible for managers and employees to be able to create stories that show how the process is meaningful. In fact, some have even argued that the entire function of HRM systems like PA is to create shared cultural-symbolic meaning, rather than to produce tangible outcomes such as improved organisational performance (Alvesson & Kärreman, 2007).

PA events materialise through action but sensemaking of such events should not be expected to be clear-cut reflexions of a causal sequence covering from target setting to PA. Rather, we have already seen in the literature review how PA processes would be expected to contain a multitude of ambiguities like unevenly distributed information, conflicting interests and interpretations, different values, emotions, lack of transparency, time pressure, unclear roles and responsibilities, measurement challenges, etc. Therefore, when looking at PA in a sensemaking perspective, one would have to pay much attention to the individual reality inventions through activation of important properties of sensemaking (Weick, 1995: 17-62). In the present chapter I will explain how Weick's (1995) '7 properties of sensemaking' model can be understood and utilised as a tool for analysing meaning creation patterns in PA.

As my literature review revealed, PA schemes are widely used in larger Western organisations. Through the spread of 'best practice' models and supported by the standardising effect of consulting firms assisting organisations with implementing PA schemes, ideas and concepts around PA have been institutionalised.

Today, most large corporations have implemented and maintain standardised PA processes. Already 10 years ago surveys indicated that 75-89% of all US business organisations had implemented a formal PA process, with an increasing tendency

over time and positively correlated to organisation size (Coens and Jenkins, 2002: 35). Fuelled by the fashion of New Public Management, the institutionalisation has for two decades now also expanded to large parts of the public sector. The institutionalisation of PA defines the overall playing field for the ways companies design and conduct PA.

Institutional theory is often portrayed as a macro perspective being powerful in explaining how structures reproduce themselves and bound the behaviour of individuals by rules, norms, traditions etc. On the other hand institutional theory has often been criticised for its lack of explanatory strength regarding agency and change.⁴⁷ In my opinion it is important to understand that sensemaking is not going on in an institutional vacuum. Institutions matter and they also matter to individuals and to the meaning creation of individuals. Institutions frame what is 'normal' and 'how we do things'.

In my view, institutions are not substituting sensemaking. Institutions and sensemaking live side-by-side and affect each other such that macro and micro interacts. Weick himself combines sensemaking theory and institutional theory within the concepts of tradition and 'industry ideology' (Weick, 1995: 124). PA as an industry ideology is a frame that combines beliefs about cause-effect relations (e.g. the target-performance-reward sequence), preferences for certain outcomes and expectations for appropriate behaviour (Weick, 1995: 110).

Institutions and ideologies simplify the world which is also a crucial dynamic in sensemaking processes. Institutional theory would likely be a fruitful way to analyse the spread and dominance of PA as a management technology at the field level creating an increasingly homogenised field. In PA, ideas and concepts have been institutionalised, so that institutionalisation becomes a resource for enactment and thus for meaning creation. In that sense, consultants and management fashion can be said to provide market cues to the organisation about 'best practice' which the organisation translates into local procedures. After all, "Seeing what one beliefs and not seeing that for which one has no beliefs are central to sensemaking" (Weick, 1995: 87).

⁴⁷ See e.g. Scott (2008), March & Olsen (1989) and Powell & DiMaggio (1991)

Weick points to the importance of information technology in future studies of sensemaking, since this affects work life more and more (Weick, 1995: 177). But besides the IT system, PA is in itself a management technology that follows a normative decision model created to invent a meaningful ordering of employees into objective, comparable performance categories. This 'corporate story' has the potential to contribute to a shared company culture, e.g. 'we are a community of competitive high performers'. The more advanced the technology is thought to be, the more likely people are taking it for granted and refuse alternatives without reflecting. This is what Weick calls 'The fallacy of centrality' (Weick, 1995: 3). But as described above, this homogenous meaning is challenged by the fact that although people may share collective PA experiences, they do not necessarily share a collective meaning of these experiences.

I will leave it to other studies to dig into a possibly fruitful institutional analysis of the spread and persistence of PA schemes at the macro level and the hegemony of the dominant paradigm within the international research discourse on PA.

Below, I will explain how Weick's model can be understood and made operational in the context of PA. This will be done by exemplifying each of Weick's seven properties of sensemaking with particular aspects of the PA process. By using examples from my interviews with managers, employees and HR experts at PharmaComp I will illustrate how each of the seven sensemaking properties can operate as different lenses for closer observation of meaning creation in PA, thereby building the argument for using sensemaking theory as an analytical tool when analysing meaning creation in PA.

6.1 Performance Appraisal as Identity Construction

First, sensemaking in PA is grounded in identity construction (Weick, 1995: 18). All people have a fundamental need to find an answer to the question 'who am I?' This need does not exist in isolation from the social world; it is highly dependent on ones perception of other's perception of oneself. Every employee and manager through the PA process will ask herself questions like 'what kind of performer am I?', 'how does my performance constitute me as a person and a capability in relation to my manager, my colleagues and myself?' and 'what is my social position when I act and

talk like I do?' Additionally, beliefs about the identity of the firm are part of the individual's mental model in the sense that each employee deals with the question 'what kind of firm am I working for' and 'what is my role in it?' while constructing their own identity. Let us start by an illustrative example:

The special thing about this company is that we are so much characterised by a consensus culture. Because we are a research based organisation, where all arguments are listened to.

The quote is taken from a Corporate HR expert participating in the establishment of PA in PharmaComp. It is one of many examples where people I interviewed directly refer to collective and individual identity construction. In the quote the interviewee reproduces a specific aspect of corporate identity (the term "*consensus culture*") with relevance not only for the organisation as such, but also for each individual (the term "we" and the statement "all arguments are listened to"). Finally, the quote also constructs an explanation to why PharmaComp's culture is allegedly characterised by consensus (the statement "*we are a research based organisation*").

The traditional academic way to analyse a quote like this would be to challenge the claims. Is it true that PharmaComp is a more "consensus based" organisation compared to other organisations, considering e.g. that it has implemented and enforced one global PA process across geographies and workforces? Is it true that all "arguments are listened to", considering the apparent need for higher levels of management to enforce strong control measures to avoid rating inflation? And is it true that PharmaComp employees generally perceive themselves as being part of "*a research based organisation*", when the number of employees working within R&D constitutes only a small fraction of the employee base?⁴⁸

However, trying to provide answers to these questions would miss the purpose of my study. I am not looking for one objective truth about PA in PharmaComp, as if any such thing should exist. Neither am I on a mission of falsification of managers' and employees' would-be misperceptions. On the contrary, the purpose of my study is to understand the meaning creation of managers and employees as it is embodied in sensemaking through their involvement in the PA process.

⁴⁸ In 2010, R&D employed approximately 12% of all employees in PharmaComp. However, that includes all employees within R&D, also all staff who did not have research related tasks, e.g. administrative employees.

One thing is the direct references to identity construction made by managers and employees. If we further add examples where identity construction was indirectly referred to by interviewees, we have a large and rich set of documentation of how the PA process affects identity construction.⁴⁹ Let us look at an example of such indirect reference to identity. Below, a manager reflects about what she would change in the PA process, if she had the opportunity:

If I could, I would allow more subjectivity into the [PA] process instead of so much focus on governance. I would like it to be me as a manager who decided if it should be a 3 or a 4. And then it should be up to me to handle any issues with my own bonus pool. I am not in favour of this idea of an ideal distribution of ratings.

What constitutes the identity of a manager? In the quote, the manager is not directly referring to identity. But when one looks closer, it is not difficult to see how the statements are related to perception of role and self. In the statement "I would like it to be me as a manager who decided..." the manager refers to a perception of the role as manager which includes the power, or freedom, of decision making within certain boundaries. This power (and freedom) is threatened by "so much focus on governance" and by the current perception that the manager does not have the discretion to decide if the rating of her own employees "should be a 3 or a 4". The manager realises that too many 4s may be an issue in relation to the limited bonus pool and the organisationally desired link between rating and bonus. But the manager's self-perception dictates that "it should be up to me to handle any issues with my own bonus pool". Notice the embedded ownership expressed in "my own bonus pool". By installing strict governance on PA rating distribution and accordingly on bonus allocations, the process threatens what the manager essentially sees as her domain; the power to take decisions regarding her own employees and budget.

According to Weick (1995: 23), individuals learn about their own identity through action: "People learn about their identities by projecting them into an environment and observing the consequences". In this view, identity does not pre-exist - it is something one learns through interacting socially. When a manager or an employee sees challenges to her role in the PA target setting process she senses a loss of

⁴⁹ Weick (1995: 18-24) particularly deals with identity construction at the individual level, i.e. the self-concept or self-image, symbolised through the metaphor of what the individual sees when he observes himself through the organisational 'mirror'. However, Alvesson & Kärreman (2007) also emphasises the organisational level identity construction, meaning the way organisations reproduce meaningful identities through the maintenance of particular management and control procedures, no matter the tangible outcome of these procedures.

meaning which must be filled. This can either be by sticking to the role as it is and elaborate new stories that justify unchanged behaviour or it can be by renewed behaviour supported by new stories contributing with new meaning. Weick (1995: 23-24) elaborates:

I make sense of whatever happens around me by asking, what implications these events have for who I will be... What the situation means is defined by who I become while dealing with it or what and who I represent.

Thus, sensemaking is characterised by being self-referential (Weick, 1995: 20). In a PA context the interesting question will be how process components like target setting, feedback, calibration and appraisal affect the identity construction of the individual and what that then does to the individual and the surrounding organisation. However, sensemaking is not just an individual endeavour. In fact, Weick (1995: 18-20) is quite clear about the importance of the social: "No individual ever acts like a single sensemaker... Identities are constituted out of the process of interaction".

In the present study I have through three rounds of interviews with manager-employee dyads from four workforces focused heavily on the interaction between manager and employee, as they constitute the key actors in the PA process. But they are not the only actors. Identity creation is not only about making sense of what one sees when one looks in the mirror, but also what one anticipate that others see when they look at you. Ring & Van de Ven (1989) formulate it nicely:

Sensemaking processes derive from...the need within individuals to have a sense of identity - that is, a general orientation to situations that maintain esteem and consistency of one's self-conceptions.

Esteem is about receiving respect from one's surroundings (Oxford Dictionary Online). It is a social and relational feeling. Consistency of one's self-conception requires translation of complex signals from the surroundings. Also in the case of PA, the picture may be complex. An employee may get too little or less transparent feedback. Or an employee may receive contradictory feedback from multiple sources. Nonetheless, the PA process ends with a single, one-digit rating summarising the performance feedback for an entire year. Although the rating may be based on a complex multitude of indirect measures of business results (Austin, 1996), interdependencies to others' performance (Cook, 1995), personal behaviour (Cardy et al., 1998), organisational politics (Cook, 1995), interpersonal relations to the manager (Varma et al., 2005), higher management level's rating distribution

policies (Piercy, 1994), system factors (Soltani et al., 2004), external factors (Furnham, 2005) and much more; at the end of the day it still ends up in a very simple expression: a number from one to five. This number communicates the formal organisational view on what and how the employee has been contributing over the past year.

Identity construction takes place all through the PA process but nowhere is it more at the edge than when the individual is confronted with this view of one-self by 'the others' in the form of a one-digit rating. Where this feedback is characterised by lack of congruence with the individual's self-perception, the effect will be loss of meaning to the individual and sensemaking processes must be activated to create new meaning.

6.2 Performance Appraisal as Retrospective

Retrospectivity is a key characteristic of the PA process, and in fact for almost any kind of appraisal. How do actions in hindsight become selected as 'performance events'? How does the individual's contribution in a performance event that is most often inherently collective become defined in hindsight? How do frames of sensemaking of past events differ between manager and employee and how does that affect their sensemaking of the same past events? An employee told me a story about a department meeting where there had been a discussion of what the department had achieved over the past year:

Employee: "We were discussing in the department what we had been doing over the past year. And I was thinking... eh, what did I actually do last year? And then, when I looked back in my calendar, then - oh yes, there was that little project in February and March. I had completely forgotten about that."

Weick claims that every manager is a historian (Weick, 1995: 184), but I think it would be closer to the heart of sensemaking in PA to say that 'everyone is a historian'. The employee will through her self-appraisal need to capture the essence of her own performance over a full year in a few sentences. If she has done nothing to write down performance events on an on-going basis as they unfolded, she must be able to remember them retrospectively. The quote demonstrates how an employee can forget even a two-month long project. What are the chances that the manager would have a better memory of the individual's contributions over the past

year, when she often has 10+ employees to cover and lots of other priorities on her plate?

Actions happen in moments but competing stories around action can live for years. PA is communicated as a rational, linear process; PA is preoccupied with a sequence covering foresight, prediction, planning, execution, evaluation and reward. But in a sensemaking perspective, the targets, performance events and appraisal will be under on-going invention inside and between the involved parties in their constant search for sensible stories that may or may not reflect the idealised rationality and linearity of the PA process.

Retrospectivity is not just about remembering and filtering performance events for inclusion in the evaluation process. It may be even more important to understand how retrospective rationalising establishes stories about the individual's part in the collective. If employee Z participates in a project and the project is less successful, what does that say about the performance of employee Z? Did the project miss because of the bad performance of Z? Did the project miss despite good performance of Z? Would the project have completely failed if it had not been for Z's outstanding contribution? Who defines the answers to these questions, and how do they do it? A manager explained to me how transparency in PA was limited due to the individual's dependency on others, due to targets only being partially covering the responsibility of the employee and due to factors that are important for evaluation but that are not documented in targets. I asked the manager where all that left the employee in terms of PA process transparency. The manager explained:

Well, what we do is we write some comments to each target. Really, it is the comments that are the important part, because that is the real story. That is what it's all about... That is where the transparency is.

The real stories are in the comments. Not in the numbers in the ratings. But it is the numbers that are left when a PA cycle is closed and it is the numbers that feed all the downstream processes of bonus allocation, salary adjustment, talent appointment, promotion etc. So why is this manager putting so much emphasis on the comments? Could it be because the comments are within control of the manager? Rating distribution policies and calibration procedures limits the freedom of the manager concerning rating. But no one interferes with the comments. The manager can write almost anything in the comments. This is where the true managerial evaluation can

be freely provided. Calibration may force the manager to give a 3, but the manager can formulate the comments as if it had been a 4.

However, where does this leave the employees? Employees are of course aware of the retrospectivity of the process and act accordingly. One employee could at midyear foresee that she would not be able to reach one of her targets. I asked her if she had been considering how to explain that to the manager. She replied:

Yes I have. It is some external factors that caused it, right. And because we have these artificial limits on budgets

I asked the employee if she thought the manager would accept that as an explanation, and she replied:

Well, I do not know my new manager as well as I knew my previous manager. But I think it can be accepted.

Retrospectively defining what *really* happened, is key in the PA process. And a story's viability is dependent on the receptiveness of others. One must look at the stories the actors create and promote, to analyse how they activate the past to make meaning of the present.

6.3 Performance Appraisal as Enacting Environments

According to the Oxford Dictionary online, enactment means "an instance of acting something out". To Weick (1995: 30), this "something" is often a part of the organisation one faces. Therefore, analysing demarcated parts of an organisation, such as different workforces, can be a fruitful way to build understanding of how characteristics of workforces plays together with patterns of meaning creation in PA.

Enactment of sensemaking frames is expressed through action. As Weick (1995: 125) puts it, "The moment action is performed, it ceases to exist... the only things that can be transmitted are images of action and beliefs". With its focus on images of past action, enactment is closely related to the concept of retrospectivity. In PA, managers and employees enact past actions bound by perceptions of their workforce environment, by developing stories that make the past more meaningful to themselves.

The enactment processes of managers and employees is not always congruent, which gives way to competing stories that sometimes can live in parallel, but

sometimes collide and re-emerge as either a consolidated story or new parallel stories that are more meaningful to the actors than the previous. A manager may conceive her environment to consist of a number of elements, e.g. workforce guidelines, distribution policies, involvement of higher management, calibration, etc. But also employee enactment is a key input provider to the manager's enactment through action. Often for the manager, the part of her environment that is easiest to 'move' will be the employee. For the employee, the situation is different, because the power base is different. But the employee still pulls on environmental factors in her sensemaking and by creating stories that are accepted as meaningful beyond the employee herself, the environment will be directly or indirectly affected.

Some managers are very aware of the problematic coupling of a rational PA evaluation vs. a much more narrative translation of action and environmental enactment. I asked a manager if she ever experienced that people reached their goals by leaning back and letting others do the work for them. The manager explained:

Yes, that is quite typical. It is not a problem to agree that something was down-prioritised because something more important came up. That is easy to accept. But it is true that it is sometimes difficult to measure people's contribution when other people have been involved as well. You know, the project succeeded and the target was realised, but what was really your contribution?

Sensemaking starts with action, but soon turns to enacting stories around this action. According to Weick (1995: 168) the stories can either focus at the level of commitment, i.e. why did the action occur? Or they can focus at the level of outcome, i.e. what did actually occur? It is the enactment and re-enactment of images of action through stories that maintains a meaningful perspective in the heads of managers and employees. Sometimes actors enact contradicting stories that challenge the meaning creation of one or more of the involved parties.

Thus, retrospective sensemaking can be observed through the enactment of actions of the actors through stories.⁵⁰ Executive Management defines frames for sensemaking by imposing a strict PA regime on the organisation, to which managers and employees have to relate. Through their actions managers and employees not only enact the PA process itself, they also enact variations that translate into stories

50 Here one could see a relation to the structuration theory of Giddens (1984) who emphasises the complex interplay between structure and agency, and the reproduction and development of structures through action.

about performance and about the PA process in a way that is meaningful to themselves.

One way to make this puzzle meaningful is by realising self-fulfilling prophecies which the PA process could be expected to be full of. The question an employee asks herself is not about exactly what actions and contributions she made, but how these meaningfully can be invented as relevant performance events related to targets defined many months prior to the action. The outcomes will so to speak be searching for targets (prophecies) for which a match can be established. In this perspective the PA process can be seen as the can in a garbage can model⁵¹ for employee appraisal decisions.

6.4 Performance Appraisal as a Social Phenomenon

The social dimension of PA is crucial. In modern working life, employees are almost always dependent on cooperation with others. Most evident is the role split between manager and employee and the social implications of this. The establishment of a job description and performance targets set socially defined boundaries for the employee's actions. Every individual in an organisation is always contingent on other individuals in the organisation and on individuals and structures beyond the organisation. As Weick (1995: 40) puts it: "Even monologues and one-way communication presume an audience".

Weick is concerned with what he calls a built-in minority problem in sensemaking, derived from the fact that minorities invent meaning images from a low credibility standpoint from which arguing over stories with higher credibility adversaries will be stressful. This could be interpreted as a traditional power dimension and can easily be recognised in the manager-employee relation which is also characterised by unequal formal power bases. In a power perspective, the three yearly PA meetings are occasions for maintaining not only the hierarchical relationship between manager and employee but in fact the entire organisational hierarchical structure. This is both in terms of power to appraise but also in terms of creating an image of consistency of objectives all through the organisation. This is what strategists would call 'strategic

⁵¹ See e.g. March & Olsen (1989) for a discussion of organisational choice as a garbage can model.

alignment' or 'strategic planning'. Hence, the meetings reproduce the ideology and the social relationship between different agents in the organisation and this gets even more substantiated when the meeting is objectified in a physically signed PA form.

However, also the environment external to the organisation plays its part in the social frame of the individual. One employee I interviewed was so embarrassed with a particular type of management follow-up activity that she would not talk about it with anyone from outside the organisation:

I cannot tell this to anyone. My best friends would never think that I would accept these things. That is actually the toughest part of it all.

We see in this quote how properties of sensemaking melt together in single sentences, as the socially defined self-perception collides with experience and thus challenge the employee's identity as a person. Here, the employee acquiesce by not telling her friends about the issues but in other cases the reaction of the friends could have been enacted to try to affect management to reconsider their approach. Or more modestly, persuade the nearest supervisor to modify the approach.

Besides being a power relation, PA is also a much more complex relation of interdependencies. If the manager's role is to coach and direct the employee, how can failure on part of the employee to meet a target avoid also being a failure of the manager? How does this interdependency affect how the manager and the employee make sense of performance events? And how do parallel social relations to project managers, colleagues, vendors and other people influence and frame sensemaking of the event?

6.5 Performance Appraisal as an On-going Process

Although work performance is evaluated in one-year cycles, Weick (1995: 43) directs the attention to the fact that sensemaking never starts or stops - it is a continuous process. Past experience and existing stories affect future action, enactment of action and meaning creation of action. However, not everything that happens is included as equally important in sensemaking. Sensemaking is selective and focus on meaning creation by careful selection of events:

To understand sensemaking is to be sensitive to the ways in which people chop moments out of continuous flows and extract cues from those moments.

In this sense, a performance event is in itself an abrupt or even artificial 'chop' in the performance flow of an individual through a full year. The event selection is in itself a construction made for the purpose of making sense. Both the sensemaking of each single event but also sensemaking of the larger context; what performance is expected of me? What performance have I delivered? What kind of employee and person do I become by that?

In the PA process, the employee and the manager 'chop moments', or let us call it pick events, out of a large amount of possible events. For example, one of the managers I interviewed had been disappointed by a specific behaviour of an employee, and picked examples of this behaviour as one of the key themes she wanted to discuss at a PA Midyear meeting:

It may be small things you note, but it is part of an overall impression... It is better to have the examples. Then you can decide how to use them.

The manager states that "*it is better to have the examples*", but what examples? What story is created by focusing on these examples instead of some other examples and what meaning does that make to the employee? In this case the examples the manager talks about are small examples where the employee's behaviour did not fully live up to the expectations of the manager. Unsurprisingly, the employee had not picked the same examples as events relevant for discussion at Midyear. And when the subject came up at Midyear, the manager and the employee did not agree on what meaning to attach to the examples.

The on-going character of PA can also be illustrated by its calendar year driven property. I asked a manager how the PA calendar year cycle fit with the nature of the employee's work:

It does not fit at all... It is the constant dilemma of PA that we live in a changing world

If a project or task is going on for years, or even just across a year change, but PA targets are set for calendar years, there is a lack of congruence between the ongoing nature of the reality of the task as opposed to the fixed start and stop of the PA target definition. And this incongruence gets even worse if the task at hand is not just relevant across a year change, but maybe even is dependent on performance on preceding tasks and/or itself a precursor for later tasks, to which there are therefore dependencies that should in theory be included in the evaluation of target realisation, to be able to capture the on-going character of reality.

While the above manager may be right that it is a constant dilemma that tasks, circumstances and priorities change in a dynamic reality; this in fact only increases the importance of the process of choosing which examples should be used to facilitate feedback and evaluation of the individual employee. Both manager and employee have their own agendas and maintain their own meaning creation of what has been going on over the past year when they meet in January for final appraisal.

6.6 Performance Appraisal as Extraction of Cues

One thing is selecting examples of performance as input to the feedback and evaluation processes. Closely related to this, is the concept of extracting cues from those examples. Performance events can be interpreted as cues and an important element is to understand for which cues there is a receiving audience. Since “sensemaking is everywhere” (Weick, 1995: 49), and performance is a continuous flow of events of which only a small minority will be selected as examples, it becomes important to understand how actors attempt to control or affect which PA events are allowed to generate cues that will serve as reference points in feedback and evaluation, and which are not.

Sensemaking is driven by schemas and seemingly recognisable order. Beliefs are based on and reproduce frames of reference which help employees to maintain meaning when they are met with uncertainty. As Weick (1995: 133) puts it; “Belief is seeing - to believe is to notice selectively.”

Pressure from workload, lack of time and general confusion and uncertainty encourages people to confirm their expectations. Darley and Fazio⁵² show how these expectations form outcomes, thus adding to the accounts of PA within an expectancy theory paradigm. Referring to Weick’s (1995: 54) famous story about how “any old map will do”, one could see performance targets as a map that will direct the manager and employee towards a specific location although the relation between the map (the targets) and the location (the performance) is sometimes much more decoupled than one would expect.

⁵² Darley & Fazio (1980) in Weick (1995: 150)

The process of extraction of cues starts already with the target setting process. In some areas targets are defined strictly top-down, but in most areas the employee has a more or less pronounced influence on at least the final wording of some of the targets. An important mutual expectation setting is going on in this phase where the employee may see herself as having an interest in setting the bar low, to be able to over-perform with less effort than if the bar had been higher. In this target negotiation, reference to previous performance, external factors, particularities of the task ahead and other elements can be enacted as factors determining the level of the bar. The manager's role in this negotiation is more challenging than it may at first look. In fact, it would be an incredibly difficult task for the manager to secure that targets were equally ambitious for all employees in her area. Even if all had exactly the same targets, differences in competencies, experience, salary, job class and other factors would mean that targets were in fact far from equally ambitious. On top of this, all the managers I interviewed claimed that the final appraisal rating of employees is not only measured on how they perform on their PA targets. Other factors are important too. The selection of such factors and the extraction of cues from them are keys to understand how a manager reaches a final rating for an employee. One of the managers I interviewed formulated this in its positive dimension:

I think it is about making people feel that they are not just evaluated on their [PA] targets. But also if they do something outside the targets.

It is nice for the employee to know that if they do something good and important outside their PA targets, it will benefit their rating. But what if the manager extracts other cues from such events than the employee? What if the manager does not see such non-planned performance as good and important? The manager's statement may be a comfort, but it may as well be the opposite. It depends on what performance events are selected and what cues are extracted from those events.

6.7 Performance Appraisal as Driven by Plausibility Rather than Accuracy

According to Weick (1995: 56-61) the sensemaking perspective will be directed much more by creation of plausibility than by creation of accuracy: "Accuracy is nice but not

necessary... action requires simplicity rather than accuracy". While the dominating economic view on PA is preoccupied with optimising accuracy by developing technologies for defining unambiguous targets, measurement methods and transparency, sensemaking takes a much more constructionist approach to truth. The entire critique of PA designs for being ambiguous due to indirect targets, measurement challenges, manager bias and all the other issues that we learned about in the literature review, somehow get turned upside down through the sensemaking perspective.

In a sensemaking perspective, the performance events are constructions, so why be obsessed with pursuing accuracy and objectivity? What is necessary is not a clear and undisputable determination of the individual's performance and its relation to an unambiguous target. This is nothing but an illusion anyway. And why should that be problematic when all that people really need is that the process and the outcome of it are plausible and meaningful, rather than accurate?

In a sensemaking perspective, what is required are good stories that satisfy the needs of the involved parties to create and maintain a meaningful conception of themselves and their environment. Stories convey values and meaning - and shared stories can convey shared values and meaning, although they do not necessarily do so. Shared stories can work as tools for diagnosing new events as they arise and require adoption into the individuals' sensemaking perspective. If we look at PA at PharmaComp in this light, we can see that many of the processes and directives from HR and Senior Management are in fact counterproductive to the opportunities of managers and employees in jointly creating such stories, because the ideal distributions, the average rating targets, the strict process governance and policies, the calibration and many other aspects of the PA process are all limiting the degrees of freedom for managers and employees to find common ground and reach meaningful common stories.

This observation is of course not a proof that the organisation is not benefitting from having installed a PA scheme. Rather, it is an observation of some of the challenges strictly governed PA schemes will be to sensemaking processes. These challenges do not mean that sensemaking will not take place. But the challenges will mean that sensemaking will take place in different ways than would otherwise be the case. Those ways will not only be about creating meaning out of performance events, it will

also be about how to create meaning out of the PA scheme itself, with all its governance, deadlines, rules and limitations.

6.8 *The Task at Hand - Performance Appraisal across Workforces*

This study is aiming at uncovering patterns of local adaption and meaning creation across workforces, based on the assumption that workforces can be seen as demarcations of environmental differentiation that affect meaning creation processes. Thus, it is relevant to consider what particularities characterise each workforce and how they differ from each other. In other words, what kind of work is conducted in the four workforces in scope for comparison, what are the main similarities and differences and how does this relate to theories of PA and measurement.

PA and the task at hand

Austin (1996: 89-101) distinguishes between two management paradigms that are mutually conflicting; measurement-based and delegation-based management. Measurement based management is relying on extrinsic motivation and relates strongly to the philosophy behind PA schemes, i.e. the establishment of clear, measurable targets and evaluation of employees through observance of tangible, preferably quantifiable, outcome in relation to these targets. In contrast, delegation-based management relies more on intrinsic motivation and is more concerned with the employees' "unsupervised effort allocation" through e.g. enthusiasm for the mission, personal satisfaction etc.

None of the two models are cost-free to the organisation. Austin lists a number of circumstances under which delegation-based management can be difficult to establish successfully. These include e.g. the impersonal environment of large modern organisations, cultural heterogeneity, infrequent interrelations among employees, low organisational prestige, low perceived level of mutual commitment and care between employee and organisation, etc.

Austin establishes a model of elements that affect the organisational costs of leveraging either the measurement or the delegation-based management approach. The cost drivers connected with the measurement-based management approach include: repetitiveness of the tasks (high repetitiveness means lower measurement

costs), complexity of tasks (low complexity means lower measurement costs), newness of task (old and well-established tasks mean lower measurement costs), specialised knowledge (low requirement of specialised knowledge means lower measurement costs), interdependence and separability of effort (high degree of individuality and effort separability mean lower measurement costs) and finally environmental covariance and noise (low degree of environmental effects means lower measurement costs).

The measurement costs listed by Austin are closely connected to many of the problems in PA we have previously dealt with. For example, the many different bias problems in PA (Cardy et al., 1998. Stark and Poppler, 2009. Lee and Son, 1998. Varma et al., 2005. Cook, 1995. Coens and Jenkins, 2002. Raymark et al., 1999. Bacal, 1999. Murphy, 2008) are all related to challenges in the measurement phase. The extrinsic/intrinsic motivation problem in PA (Kohn, 1999. Bowen, 1995. Piercy, 1994) is related to the choice of measurement vs. delegation management. And the quantity over quality problem in PA (Soltani et al., 2004. Deadrick and Gardner, 1999) is related to the focus on quantifiable measures in measurement schemes. Thus, Austin's model of measurement costs is a good way to group a significant number of the problems in PA into a simple structure against which we can compare the four workforces in this study to identify particularities and commonalities that may be expected to affect meaning creation within PA in the workforces.

Delegation-based management include the following cost drivers: organisational size (big organisations tend to be less personal and hence it is more difficult to foster organisational identification), cultural homogeneity (less homogeneity means higher delegation costs), duration of relationships among principal and agent (new relationships drive higher delegation costs), interaction frequency among employees (low frequency drives higher delegation costs), organisational prestige (high prestige fosters high employee identification and lower delegation costs), and finally perceived level of mutual commitment (the more employees perceive that the organisation takes responsibility for employees' life quality, the lower delegation costs).⁵³

⁵³ Austin operates with two additional categories of delegation cost driver that I have not included here. One concerns whether the organisation is public or private. However this separation builds on the assumption that public sector employees are generally less motivated than privately employed, which I believe is questionable. The other concerns "the degree to which employee needs are met in the organisation", which I think is overlapping with the "Level of mutual commitment" cost driver mentioned above.

Also the delegation costs listed by Austin are connected to multiple problems in PA. For example, the cultural homogeneity cost driver is connected to problems regarding cultural difference and global use of PA comparisons (Woods, 2003). The relationship between principal and agent cost driver and the interaction cost driver are related to many of the problems and bias in the relationship between managers and employees (Kohn, 1999. Varma et al., 2005. Cook, 1995. Raymark et al., 1999). The organisational prestige cost driver relates to the extrinsic versus intrinsic motivation dilemma in PA (Bowen, 1995. Kohn, 1999. Frey, 1997). And the perceived level of mutual commitment cost driver relates again to problems in the manager-employee relationship inherent in PA schemes (e.g. Waite and Stites-Doe, 2000. Langan-Fox et al., 1998. Prowse and Prowse, 2009. Kohn, 1999).

The task at hand and expected management by measurement cost drivers in the four workforces

In the pharmaceutical industry, the research process is rather long from discovery of a new molecule, over formulation, laboratory and animal testing, three phases of human tests and finally market access approval, before Marketing and Sales takes over. It can easily take 12-15 years from discovery of a new molecule to marketing of the related drug. The area in R&D I investigated was occupied with the earliest phases of the research process, i.e. discovery through laboratory work. This area is characterised by operating a large number of smaller pre-projects with the purpose of identifying molecule candidates that are promising enough to go into the pipeline of future development projects. Most of the pre-projects are shut down early because they prove to be less promising, so employees are often transferred to new projects and they often also work on multiple projects at the same time. Although the laboratory operational procedures are in many ways routinised to adhere to quality control standards, the opposite is the case for the content of the tasks.

The whole purpose of discovery is to do things differently than what has previously been attempted, so in that sense task repetitiveness is low and task newness is high. At the same time new molecule discovery is a very complex task requiring highly skilled and trained employees. Typically, employees are either employed as PhD students or are employed already having a PhD in science. For key positions, world leading researchers are hired to generate the molecules that will determine the R&D

project pipeline and thus the long term future of the company. Also, interdependence is high in the sense that discovery work within R&D is a teamwork effort where it is rarely the case that success or failure can be attached to the performance of single individuals only.

Lastly, the environmental noise in R&D is generally neither low nor particularly high. There are many environmental conditions affecting the R&D process, e.g. quality standards, authority inspections, patent rights and strategies, business ethics issues etc. But within these frames there is also a playing field where R&D units can operate quite freely without being constantly affected by external actors. One exception can be in cases where competitors suddenly launch new market changing products or ideas that affect the research strategy of other companies. But that rarely happens - research tends to progress more stepwise and over longer periods of time, mainly due to the long turnaround time from idea to market access.

The area in Production I investigated was responsible for delivering technical services to a number of production lines. In the period of my investigation, the implementation of an extra product line was on-going, as one of the major projects in the area. Although the implementation of product lines is fairly complex, it was also characterised by some degree of repetitiveness because most of the configuration was unchanged compared to the existing similar product lines. In that sense, it is reasonable to say that the degree of repetitiveness was relatively high while the newness of the task was in the lower end.

However, it still requires specialised and highly skilled resources to implement a new product line and there is a lot of interdependence between the people engaged in the project and people engaged in the on-going daily production, which makes it difficult to separate out exact individual contributions and responsibility for project success or failure. As for R&D, also in Production there are many quality regulations and strict procedural compliance requirements defined by authorities and the organisation itself which function as environmental conditions under which work takes place. However, also in Production the environment is fairly stable, at least on a short to medium term, since new products are typically known in the pipeline years before they enter large scale production.

The Staffs area is the most heterogeneous area I investigated, covering very diverse business processes. I was looking at the IT organisation which is responsible for both development projects and on-going IT operation. Hence, I chose to look at both functions, i.e. two separate departments in IT where one was specialised in operation of IT systems and the other was specialised in managing large IT implementation projects. Both in the operational and project oriented units the tasks are generally of a medium complex nature and with some element of repetitiveness. There are very strict procedural methodologies both for running IT projects and for operating large IT systems, and although the business process supported differs from system to system, the methodologies are so profound that employees often change across projects and systems without major impact on their performance. Therefore, it is also reasonable to assume that the newness of tasks is lower and that degree of specialised knowledge required is also lower than in e.g. R&D, such that system and project managers are often highly skilled generalists rather than specialists.

However, interdependence is still very high. It is not possible to be a successful project manager or system manager without the help of project participants, internal and external vendors, management support and so forth. In that sense, the areas in Staffs I investigated was also medium dependent on environments in that e.g. external vendors lived up to agreements and that other IT departments provided the required support to keep a system accessible for employees or to be able to deliver projects successfully on time.

Sales is the workforce that sticks out most in many ways. First, the repetitiveness of the task is fairly high. Although each dialogue with a customer of course can develop in different ways, it is a fairly standardised setup, which also explains Sales management's reliance on measuring effort by counting number of customer calls, visits etc. Although the amount of work that can be put into a sales job is limited only by what the employee is willing and capable of investing, the content complexity of the task is fairly low. The sales representative will meet a large number of customers and discuss one or a few target products to build awareness and knowledge that will ultimately make the doctor prefer to prescribe the company's portfolio rather than the competitors'. Newness of the task is also low, because even when new products are introduced, the task is the same. It is only the product characteristics that change, not the approach.

To be a good sales representative requires a rather solid understanding of not only the products but preferably also of processes within the health care sector in general. Further, it requires particular personal and social skills to be able to create trust with other people within the typical half hour session with the customer. While this may not be categorised as very specialised knowledge, it would be incorrect to phrase it as generic competencies, so I value it as medium specialised knowledge. For sales representatives, interdependence and environmental noise is higher than for the three other workforces. Sales representatives share geographies and it is practically impossible to track exactly what customer prescribed what amount of drugs attributable to individual sales representatives' customer visits. Further, there is a huge impact of the organisation's ability to forecast (too optimistic forecast will lead to unachievable targets) and to the changing pricing policies and marketing strategies of competitors, as well as changing regulations by governmental bodies. Sales representatives may have received a tough but achievable target, but if her fellow sales representative is a low performer or if a competitor suddenly decreases price on substitutable products, then it can quickly become impossible to realise targets.

In Table 6-1 below I have summarised the rating of the four workforces on each of the six task characteristics included in Austin's model for measurement cost drivers.

Task characteristics	R&D	Production	Staffs	Sales
Repetitiveness	Low	High	Medium	High
complexity	High	Medium	Medium	Low
Newness	High	Low	Low	Low
Specialised knowledge	High	High	Medium	Medium
Interdependence	High	High	High	High
Environmental noise	Medium	Medium	Medium	High

Table 6-1: Measurement cost drivers by task characteristics and workforce

Although the categorisation must be taken with some caution - after all they are very broad categories across very large, complex and diverse workforces - I still find it to be illustrative for how the workforces differ. Production and Staffs are probably the two areas where the tasks are most similar by nature. On the other hand, R&D and Sales appear to be largely opposites in terms of task characteristics, except that they

both operate with a high degree of interdependence. Based on the categorisation summarised in Table 6-1, we would expect the biggest measurement issues in R&D and the least issues in Sales, while Staffs and Production would position themselves somewhere in between these two extremes.

The task at hand and expected management by delegation cost drivers in the four workforces

It is more difficult to characterise the four workforces in terms of their characteristics in relation to Austin's management by delegation cost drivers. This has to do with how the "softer" cultural elements in this cost driver category are determined both by the overall company cultural traits and by the more particular traits in each workforce. Without going into a deeper cultural analysis of PharmaComp, it is not easy to separate the two factors affecting the delegation cost drivers. Nevertheless, there are identifiable and general traits across workforces that drive different delegation cost profiles which I will try to categorise and exemplify below.

First, the overall organisational size is of course one and the same for PharmaComp as such. Also, each workforce is in itself rather large, if we consider the entire global entities of Production, R&D, Sales and Staffs. In such a comparison, Sales and Production would be the biggest, followed by R&D and Staffs. However, this comparison holds little relevance to the environment that each manager and employee works under in daily life. Employees in Sales are managed at the regional headquarters and many guidelines, targets and policies relevant for Scandinavian sales representatives are therefore defined at the European level containing approximately 25 country level sales affiliates. Thus, as we will also later see, Sales managers and employees often refer to the European headquarters as the ones responsible for certain elements in PA particular to Sales. In this way, the sales managers and employees perceive they are part of a big organisational entity (large scale decisions means more for the practice of PA) than was the case in the three other units I investigated, where the perception was that their local area senior management was setting direction for PA.

Overall, PharmaComp is recognised for having a strong corporate culture due to its long history, strong performance track record and the continued influence of Scandinavian culture, which is in itself characterised by a high degree of cultural

homogeneity (Hofstede, 1983). Still, in terms of cultural homogeneity, R&D sticks out from the other workforces. In R&D, it is mainly the dominating academic background (i.e. science PhDs) and the scientific ethos these people bring into the organisation that forms a general cultural trait.

Sales also has traits of strong homogeneity; the archetypical background for Sales representatives is to have been trained as a nurse and to have worked for some years in the health care sector. Besides this, the culture in Sales is characterised by a strong focus on sales performance - sales is what matters - with less methodological and scientific process dogma than in particularly R&D. Nevertheless, the managers and employees I interviewed in Sales found that there was a huge cultural distance from their local level to the European level where policy and process design was made for Sales. Since this was a key source to challenges in PA in Sales, it is probably most accurate to categorise Sales as having a low cultural homogeneity in terms of PA. The Staffs and Production areas I investigated were in this respect more heterogeneous than R&D and Sales as they held a mixture of project and operation resources working with different types of tasks.

In terms of duration of relationships and frequency of interactions, the workforces also differ. In the area of R&D I investigated, employees worked closely together in small teams on individual projects. This work was characterised by very close daily cooperation, sometimes over a shorter period (where projects were terminated) and sometimes over extended periods of time. Where projects are approved for continued research, teams of employees can work closely together on investigating single opportunities for years. In contrast, Sales people work on a very individual basis divided by geographical focus areas and individual lists of customers to cater for. Also in relation to the customers, the frequency of meetings is fairly low; often defined to be 2-3 times per customer per year but in many cases in reality fewer, simply because it is the customer who decides how often they want to prioritise spending time with a sales representative.

In the area of Production I investigated, projects were also running over an extended period and while this was going on, interactions were fairly frequent among the participants. However, next project could be in an entirely different area, because of the generalist nature of the employees in the area and the varied project portfolio, while the specialisation of R&D employees in comparison tend to make them work in

more closed research communities on more similar projects. In Staffs, the heterogeneity of the tasks between project and operation makes it more difficult to generalise. In the project area, interactions are frequent through a project, but then the employee moves on to another project and interaction stops with the prior project's participants. In the operational part of the Staffs area, relations have longer duration but interactions are then less frequent.

In terms of organisational prestige, the overall level is medium. In general, the pharmaceutical sector is not particularly prestigious. Examples of bad press coverage and critical scientific scrutiny due to alleged business ethics issues are many, e.g. regarding biased drug prescription due to financial ties between medical doctors and the industry and other unethical funding and interest issues (e.g. Klitzmann, 2010), pricing policies in developing countries (e.g. Kremer, 2002) and high profit levels compared to other industries (e.g. Scherer, 1993). However, within this general frame of low industry prestige, PharmaComp is normally considered to be among the players who receive better press coverage.⁵⁴ And particularly within Scandinavia, the company has a good reputation of operating in accordance with high ethical standards and being an attractive place to work. This generally is perceived to have a positive effect on people working in the company, independent of workforce affiliation. However, if one workforce was to be picked as particularly prestigious it would be R&D. Not only is this the workforce that sets the highest formal requirements for educational standards, it is also considered to be the workforce securing the long-term viability and continued success of the company by inventing and maturing a strong pipeline of drugs that will make the company able to defend and develop its competitive position.

In terms of the last element in management by delegation cost drivers, the perceived mutual commitment, I believe it is difficult to draw any clear separation between the workforces. Across the organisation there is a strong and general focus on people well-being and the organisation is well-known for this. The company consistently finds itself being in the top of local and international employer surveys, e.g. by the Great Place to Work Institute.

⁵⁴ According to the Dow Jones Insight Media Index.

In Table 6-2 below I have summarised the rating of the four workforces on each of the selected six task characteristics included in Austin's model for delegation cost drivers.

Task characteristics	R&D	Production	Staffs	Sales
Organisational size	Medium	Medium	Medium	High
Cultural homogeneity	High	Medium	Medium	Low
Duration of relationships	High	Medium	Medium	Low
Frequency of interactions	High	Medium	Medium	Low
Organisational Prestige	High	Medium	Medium	Medium
Perceived mutual commitment	High	High	High	High

Table 6-2: Delegation cost drivers by task characteristics and workforce

Again, the categorisation must be taken with some caution. Also for the management by delegation cost drivers it is true that they are very broad categories across very large, complex and diverse workforces. But they are still illustrative for how the workforces differ. As we saw regarding the measurement cost drivers, Production and Staffs are probably the two areas where the tasks are most similar by nature in terms of delegation cost drivers. And again, R&D and Sales appear to be largely opposites in terms of delegation cost profiles, except that they both operate with a high degree of mutual commitment. Based on the categorisation summarised in Table 6-2, we should expect the biggest management by delegation potential in R&D and the least potential in Sales, while Staffs and Production would position themselves somewhere in between these two extremes.

All things else being equal, based on the task at hand as it has been categorised above we should anticipate most challenges to PA in R&D and least in Sales. It is with reference to this categorisation and in combination with the sensemaking toolbox explained above I will now step into the analysis of patterns of local adaption and meaning creation across the four workforces in PharmaComp.

At this stage it will be beneficial to revisit the research question:

How is managers' and employees' meaning creation affected by problems in PA and what patterns of local adaption and meaning creation can be observed across organisational workforces?

I have ordered the analysis into three chapters corresponding to the three main sub-processes of PA; target setting, on-going feedback and midyear review, and final appraisal. I have done this to stay in line with the structuring of problems in PA as they were presented in the literature review and in line with the data establishment methodology that was ordered around three rounds of interviews, one for each sub-process.

In the below analysis I have put labels on different stories told by managers and employees in the four workforces I investigated. It is important to highlight that these labels represent my interpretation of how particular circumstances affected local adaption, meaning creation and feedback loops related to problems in PA. The labels cannot automatically be extrapolated to the entire R&D, Sales, Production or Staffs areas in PharmaComp, let alone to organisations beyond PharmaComp. A fairly small, qualitative study would never be able to create that kind of generalisation to populations. Luckily, that was neither the intention. Rather, the intention was to generalise to theory by showing how particular local environments, here in the shape of workforces, can be understood as important moderating variables with significant explanatory power regarding dynamic meaning creation in PA and complex interdependencies between problems in PA.

Therefore I have deliberately called each of the stories of local adaption and meaning creation for "a story", not "the story". Hopefully, other researchers will take on the task to further build on my theory from different angles and by different means, including quantitatively based studies allowing statistically significant conclusions.

Sometimes, where statements are of a character where confidentiality can be important for the protection of the informants, I have used the term 'manager' and 'employee' without indication of workforce affiliation. But wherever quotes are less person-sensitive, I have referenced the workforce affiliation to allow comparison across workforces. As mentioned in the methodology chapter, I have systematically used the female form in all quotes to further support confidentiality.

I would like to remind the reader that in the context of PA, every manager is also an employee. In that sense the separation of the two groups is somewhat artificial. Managers will in their role as employees face many of the same challenges as will employees without management responsibilities. However, every employee is not also a manager, and therefore the two groups will have at least one characteristic that fundamentally differentiates them; whether or not they have management responsibility.

7 Local Adaption and Meaning Creation in Target Setting

7.1 Introduction

In the definition of roles and responsibilities in PA that Corporate HR offers to managers and employees, it is defined as the responsibility of both parties, to "*agree on business and competence targets*". Two important challenges in PA were identified in the target setting phase of the PA process in PharmaComp. The first concerns the challenge of establishing commitment in target setting. This is a broad category dealing with several problems in PA; how managers and employees are challenged by particularly the top-down problem in PA (Coens and Jenkins, 2002), the problem of PA being bureaucratic and control focused (Piercy, 1994) and the problem that targets only indirectly represent real value contribution (Armstrong, 1996). As the analysis will show, these and other problems challenge managers and employees across workforces in different ways, so that they subsequently enact different stories of what is going on concerning employee commitment in the target setting process.

The second challenge in target setting can at first seem narrower. It is about handling behavioural target setting. In all cases I investigated, behaviour had an impact on the final appraisal. Hence, one could intuitively expect it to be a natural part of the target setting process to explicate expectations about behaviour. However, the analysis will show that multiple problems in PA make it much more challenging for managers and employees to incorporate behavioural targets in a meaningful way than one might have expected.

Let us start by recalling that the purpose of the target setting process is to establish the targets for the coming year so that both managers and employees are clear about what the expectations are in terms of responsibility, focus, priorities, deadlines etc. Also, the purpose is to make it clear and transparent to the employees what they can be expected to be measured against in the later final appraisal process. As such, the target setting can be seen as the foundation, or premise, for being able to conduct a meaningful appraisal process at the end of the year.

7.2 The Challenge of Establishing Commitment through Target Setting

Below four different stories from the four workforces serve to illustrate patterns of local adaption and patterns of how managers and employees meet the challenge of establishing commitment through target setting. I have labelled these stories the story of commitment through manipulation in Staffs, distribution in Sales, involvement in R&D, and a fourth, middle position story from Production.

Commitment through manipulation

In Staffs, a manager explained how she handled a dilemma of having on the one hand an obligation to meet top-down defined targets on the organisational level and on the other hand providing employees with influence on target definition:

I have a, what should we call it, hidden agenda for what I would like the employees to have as targets. I ask them what they would like, and then there most often is a 50-50 match. From that, roughly speaking, I manipulate them to feel that they have a great deal of influence. Often though, it is my experience that they actually want the same things that I would have preferred.

This manager operates within a paradigm where employees have influence on the target setting. However, although the manager is not dictating target setting, she has a clear understanding of what will be necessary in terms of individual employee goals to be able to reach her own aggregate goals at the department level. The problem of PA being essentially a top-down process is a challenge to employee influence in target setting. How can real influence be provided if the targets are pre-defined? Meaning is established by moderating employee influence in a way so that the manager lives up to her departmental targets while trying to give the employees the impression that they have real influence on target definition. We could say that the influence of the employee is moderated and the moderating tool is manager manipulation.

The term manipulation has negative connotations, but in this context it is not this meaning that the manager refers to. The shared story the manager tries to establish is a story of balancing involvement and co-responsibility in target setting against the need of coordination so that the sum of individual goals in the department still meets the aggregate expectations to the department. The manager's degrees of freedom to let the employee decide her own targets are bound by department level targets, rules

and procedures. They are bound by the manager's own personal interest in having all top-down defined targets in her area covered in a suitable way.

It is in fact a tricky job to make these ends meet; real employee involvement vs. manager coordination and control of targets. The manager invites the employees to chip in so that hopefully the employees feel they can influence target setting in a meaningful way. This is the commitment, or leadership, element. But the manager also constantly has an eye on the larger puzzle and must make sure the end result is close to what the manager wanted in the first place. This is the control, or management, element. The manager has to master the art of manipulation.

Employees are aware that their influence on target setting is limited by factors out of their own as well as their manager's control, because the cascaded targets are predefined in everyone's PAs. The volume of cascaded targets thus becomes a strong indicator for the level of top-down enforcement and therefore also of the magnitude of the top-down problem in PA. As one Staffs employee saw this, this had implications for employee commitment:

When so many targets are defined from above... then the idea disappears, I think. Then I am left with 4-5-6 targets that are important to me and the rest are just something we all have. It annoys me that so many targets are defined from above.

It is not very meaningful for this employee to participate in a target definition process where it is procedurally up to the employee to define a first draft on targets, but where the majority of targets in reality are predefined and collectively distributed by senior management.

A similar experience was expressed by a Staffs manager, although the focus here was less on the top-down approach in terms of content but more on the top-down procedural restrictions, i.e. lack of process time to include the employees in the target setting process. The outcome, though, is the same; the employees are less included in establishing their own targets than the manager would have preferred:

We have to wait for the upper layers to fix their targets before we can set ours. I get my targets when my manager has her targets locked, and that is by the way rather late. That is a shame because I would like to have more time to discuss the targets with my people.

This manager reflects on the limitations to employee involvement that is embedded in the way the PA process is designed in her organisation. In a true top-down process, each layer has to wait with discussing targets until the layer above has settled and

locked their targets, and so forth all the way down through the organisation. The manager at the floor level will ultimately get her targets late. And therefore she will also have reduced time to discuss the targets with her department and with each employee individually. Due to top-down enforcement through target cascading and pressured timelines, it is simply not possible for the manager to reach the level of employee inclusion she would have preferred. Still, the manager sticks to the process deadlines and thus redirects meaning creation from content (i.e. establishing meaningful targets through employee inclusion), to process compliance (i.e. meeting formal requirements and deadlines). Again, the management by measurement principle takes prevalence over the management by delegation principle.

Another likely side effect of the above depicted process in Staffs is that there are also minimum opportunities for challenging targets coming from above. For practical reasons that would be very difficult to handle in such a "waterfall" target setting process. A Staffs manager reflected on the impact of this:

There have been targets where the employee did not exactly understand what was in it for her, but was just told to go and realise the target. We have received such targets from above this year. It creates huge frustration among the employees. So I feel we need to work a lot with that to make it tangible to the employees.

The manager's role is twisted between on the one hand the increased demand for alignment, control and transparency of targets across hierarchical levels as required by senior management, and on the other hand the perceived need of the employees to be able to affect, or at least have opportunity to discuss, the explication of targets at the floor level. We could say that senior management is pushing a management by measurement regime while employees are looking for elements of a management by delegation regime. As such, the 'commitment through manipulation' story is an illustration of how the task at hand in Staffs, with its intermediate position between the more extreme opposites of R&D and Sales, is connected to managers' search for a workable middle way between fulfilling the requirements of the measurement paradigm, while moderating the negative side effects by leveraging small elements of the management by delegation paradigm.

Commitment through Distribution

A more extreme example of the problem of target setting being a top-down controlled process is from Sales. In the area in Sales I was investigating, a manager explained:

7. Meaning Creation in Target Setting

I set the targets for my employees. I do not negotiate them, I communicate them.

This statement illustrates how different the practice of target setting can be in different workforces. The manager acts within a clearly top-down, controlled environment and practices a particular target setting approach, leaving no room for employee influence on targets. Sales targets are very tangible; you have a product, you have a local market, you have a \$-target within a defined period, you have a number of sales representatives among which to distribute the target. It is quite simple; if someone wants to have a lower individual target, someone else must take a higher individual target. It is easier - and may in fact also be fairer - to base the distribution of this kind of targets on mathematics (forecasts, historic performance etc.) rather than on an individual dialogue with each sales employee where they can each leverage their varying degrees of negotiation capabilities:

Sales Manager: "We know from the start that if we have to sell this much on product Y and product X, then we have to split it between area A and area B, where I am responsible for area A. From that we look at some trends and trend distribution between modules and then I get an area A target... then I simply distribute it among my employees."

This is an alternative example of local adaption illustrated by a story of target commitment through target distribution in the sense that there is a mathematical logic behind the target setting; a logic which is beyond discussion. We might say that although this manager acknowledges that the employees do not have any influence on target setting, this is not perceived as a challenge by her because it is quite meaningful to her that targets are mathematically distributed by management without any employee involvement. However, a sales employee had a different view on this:

[The targets] don't have much to do with me, I think, because I am not involved in any dialogue about them. Now I have become used to how things are - that I don't have any influence on it. But I still think that it is just something I do because the company says I must. So, I sign PA. But it doesn't mean anything to me, really.

While the manager sees the top-down defined targets as a premise for operation within PA, and therefore not a challenge to her meaning scheme, the sales employee struggles to see the target setting process as meaningful. The employee's meaning scheme is challenged by targets being defined without any involvement by her. But then new meaning is created in the shape of compliance; the employee participates in the PA process because the company says she must, but it doesn't mean anything to her. There is no commitment through PA targets.

For other workforces than Sales, the complexity of targets may be an argument for involving the employees in defining targets at the individual level. It may be argued that the manager for a project department in Staffs that runs multi-million \$ projects, to a higher degree will need the input of each of her employees to be able to reach the level of understanding required for settling targets across her department, while a similar involvement of the employees in the Sales manager's department would add rather than reduce complexity. At least that would be a convenient explanation. But if we look closer at the top-down approach and its lack of employee involvement in target setting, it seems to be built on the assumption that employee-defined goals would be less ambitious than targets defined higher in the organisation.

This assumption is relevant if employees were expected to act strategically to negotiate reductions in their targets if they were more involved in the target setting process. That would be a logical assumption within a rational perspective in the sense that lower targets would mean higher chances of target realisation and thus increased bonus and praise. In fact, if one is supportive of the positive motivational effect of extrinsic rewards, one should expect rational employees to work actively to reduce their targets to maximise their chances of success. This is an assumption supported by a strong management tradition with close ties to McGregor's (1960) theory X style of management. This constitutes a very particular story that can be meaningful to both manager and employee if they both subscribe to the above base assumptions about extrinsic motivation.

However, if managers and employees do not subscribe to the theory X style of management, they will perceive their meaning schemes to be challenged and sometimes maybe even that their role identities are challenged. If the practice is built on assumptions to which they do not subscribe, they will struggle to establish new meaning in the process, like the above quotes illustrate. Thus, at least since McGregor, it has been clear that there is an alternative assumption one could have about human nature in the workplace. Thus, a Sales manager reflected on her own experience in relation to the question of involving employees in the target setting:

The times when I have asked what they would have set the target to themselves, then many of them are in fact very ambitious. Often they set the bar higher than we [management] do... It is very top-down managed. I think we should show a little more courage.

This Sales manager's target setting role today is to distribute and communicate sales targets. But the manager is not certain that this is the best way. By using the positive term "courage" the manager indicates that stronger involvement would go against common practice in Sales in PharmaComp, and probably even against wider practice in sales organisations. But she also indicates that in her opinion it could possibly have positive effects on organisational performance and be more meaningful to employees.

Even in Sales, targets differ in nature and so does the role of the manager. In contrast to the hard targets on sales index realisation which in the cases I looked at constituted 75% of the total rating weight in PA, there are other targets where the manager potentially could be more open towards input from her employees. Also for these targets Sales operates a very clear and distributive approach to target commitment. However, at least for these targets a Sales manager experienced a demand for a less rigid top-down process than the current:

One thing that I think many in my team would appreciate is a bit more flexibility. Particularly regarding call plan and call plan compliance.

The call plan is a plan that defines the number of calls a sales employee needs to make to a pre-defined group of customers within a given period. The customers included in the target are updated bi-yearly so that any changes in the customer base cannot be reflected in the target outside these updates. The call-plan frequency targets are top-down defined at European Sales HQ level and cascaded to employee level. The Sales manager quoted above is right that employees would appreciate a bit more flexibility. In fact, this part of the scheme is perceived as a serious challenge by the sales employees I interviewed:

If I could change something, I would certainly pick the call plan targets. The call plan targets on doctor calls - that hurts much more than the sales targets... They can negatively affect the goodwill I have with some of my best customers.

I asked the same employee in Sales how she handled this dilemma. If she, for example, gave low priority to the call plan targets because she found them destructive to her customer relations:

Yes, I do. The relation to the customer comes before the realisation of call plan targets. They can say whatever they want - I will never destroy my relations.

We see here how meaning is created around a perception of call plan targets being destructive to customer relations. When call plan targets are destructive to customer relations, it is more meaningful for a sales employee who values customer relations higher than PA compliance, to abandon the call plan targets rather than to adhere to them. Thus, in this example the distributive approach to establishing call plan targets only has very limited commitment effect in that the employee find it meaningless and therefore create a story around why she does not have to commit to the call plan targets. What we see is a feedback loop where management's commitment by distribution approach kicks off an unintended feedback effect; alienation of the employee from PA and make her re-orient towards a non-committing compliance focus, with a negative effect on PA commitment.

Sales is the workforce I investigated that most clearly followed the principles of management by measurement, as was also the expectation I argued for in the previous chapter. The task at hand in Sales is such that there is a better fit to this regime than in any of the other workforces. Nevertheless, we also see that there are clearly costs associated with this approach to PA; commitment is becoming instrumental through a compliance perspective rather than an ownership perspective.

Let us now turn to a story from R&D, the workforce where the task at hand was presented as most different to Sales in terms of fit to PA as a management by measurement instrument.

Commitment through Involvement

In R&D, I encountered an example that seems furthest from the top-down problem in PA, as it was depicted by Coens and Jenkins:

R&D Manager: "I had prepared some mandatory targets for the department. It was up to each of my employees to consider how they could contribute, and then to write it down in personal targets. They have done that all of them. It is not my job to dictate targets."

In this example, the role of the manager in target setting is to set the boundaries in the form of overall departmental goals. It is then up to each employee to propose how they can contribute to that goal through individual targets. In this scenario, the manager must maintain an overview of what each employee suggests as their contribution, to make sure that the sum of individual contributions will ultimately meet the department targets. At least in larger departments this would be quite challenging

for practical purposes; it is unlikely that department goals will be fulfilled by the sum of individual goals if managers let it be purely up to employees to define their own contributions. However, taken literally, the quoted approach does not activate the problem of PA being top-down because it is up to employees to formulate their own expected contribution. This is effectively eliminating the top-down problem because it installs the employee as an agent who can not only affect, but in fact also formulate, her own targets.

As mentioned above, R&D is operating a large portfolio of research projects, from smaller discovery projects to large scale pre- and post- marketing product testing and follow-up. An R&D manager explains the target setting process as she experienced it for one of her employees:

She has written [her targets] down in her own language, which she is comfortable with. I think that is generally quite important. Or else, people would not relate to the targets. I prefer they reflect on the targets themselves. That also makes it much easier to evaluate later.

Notice that the manager in the quote proposes some assumptions about human nature that seem to be of a quite generalisable character and which have clear implications for the role of the R&D manager in the target setting process. The statements in the quote do not favour a top-down process. For this R&D manager it is meaningful to have a very active employee involvement in the target setting process because this is a pre-condition for employees' to be able to relate to the targets in the first place. Where the role of the manager in the former Sales example was characterised by a top-down management practice of target distribution, the R&D example is much more that of a bottom-up coordinator.

There is also a much less strict environment in terms of specific target cascading from higher management levels in the area of R&D I investigated. The R&D manager's degrees of freedom are wider than they were for the Sales managers I interviewed. And the R&D manager is able to enact her environment differently than the Sales manager did. But is the increased employee involvement somehow particularly relevant at a more general level in an R&D environment? At least, an R&D manager does see the R&D domain as defined by being of a particularly dynamic nature:

I think that many of the targets you set in the beginning of the year have to be altered during the year, because by then they would have become irrelevant. It is research,

7. Meaning Creation in Target Setting

you know. There are lots of changes. I would imagine that it was more stable elsewhere in the organisation.

If targets are likely to be changed, maybe even changed several times during the year, then the target setting can be perceived as less binding for both the employee and the manager. The top-down directive role of the manager can be softened a bit because there will be opportunities to update the targets later. The manager still has a role in making sure targets are relevant and support the department goals. But if there is a common understanding that there will be occasions for updates during the year, it is less hazardous for the manager to loosen control by allowing increased employee involvement. The manager is acting within a less controlling environment and is constructing a manager identity founded on facilitation through interaction and review rather than on top-down, directive measures. This could be assumed to be a more meaningful manager identity in a research culture, where concepts like supervision and peer review are common compared to more traditional control measures. A quote from an R&D employee confirms this story:

It seems a bit crazy that PA is for a full year. Things are changing a lot here, because we are here to try out a lot of different ideas. Things don't last long here. That's why I don't look at PA as something very settled. That wouldn't make any sense. I couldn't see how I should write down targets for a full year and then believe that it would also end up that way.

When conditions for operation are fluid like in exploratory R&D, management is reluctant to enforce top-down targets on employees because it would be very difficult to foresee what projects would become interesting to pursue and what projects should be stopped. We should be cautious not to take the argument too far, but based on the above R&D quotes we could see the R&D manager's enactment and role identity construction in target setting to be more of a peer reviewer rather than a traditional manager. As any scholar knows, academic peer-reviewers are still powerful and the employee would still need strong arguments to avoid incorporating review comments. But the target setting scene is much different from the one previously described in Sales.

The analysis in the previous chapter showed a lack of fit between the task at hand in R&D and the management by measurement regime that forms the foundation of the PA scheme. What we see in the story of commitment through involvement is that in the area in R&D I investigated, commitment in target setting was established by turning the management regime around towards a delegation rather than a

measurement paradigm. By loosening control of the target setting process the manager fosters a stronger employee involvement in target setting, through a more delegating management style. The cost is less management control in target setting, but this cost is clearly lower than it would have been in Sales, where a strict quantitative target should be met. Hence, the task at hand makes very different enactment possible in the two workforces.

A middle position - commitment in target setting in Production

Although R&D representatives perceive the research domain to be different, it is not only in R&D that managers perceive their environment to be more dynamic than the once-a-year target setting presumes. A Production manager had a similar perception of her area:

The PA yearly cycle does not fit the project lifecycle. She [an employee] had just started on the project so there was no project plan yet when we set PA targets. So her target became to follow the project plan, which did not exist yet... it is the constant dilemma of PA; the world is changing all the time. And we are trying to set targets for a full year, even though we know for sure that in a year's time, things look different.

In a previous chapter I have explained how Production is the most aligned of the four workforces in terms of having standardised PA processes across the whole workforce⁵⁵ and in terms of being first mover on a range of common decisions regarding policies for use of the rating scale, inclusion of behavioural targets etc. In the year of my study, a Production manager had experienced an increasing tendency towards enforcement of uniform target formulation from above:

This year the target setting process has been very top-down. I would have preferred to have my team more involved in the process.

The manager role is challenged by a decrease in opportunity to involve the employees in the target setting process. The manager enacts this environment by adjusting loyally to the increased control paradigm, but she feels uncomfortable with the reduced employee involvement and sees her concern to be confirmed by the feedback from her employees. It is important to emphasise that enactment here is understood as the individual's adjustment to the environment, but not only that; it is also about the individual's actions within this environment to creatively moderate and

⁵⁵ That is, the part of the Production workforce that is included in the PA process. I.e. not blue collar workers.

ultimately change it where she sees opportunities to do so. In the quote above the manager is in a situation characterised by less opportunity to moderate change in the environment (because of an increase in top-down control in target setting) and must adapt. Thus, sensemaking is redirected from the preferred scenario of more employee involvement to the current scenario of more top-down distribution of targets. Another way to phrase this is to say that enactment of the changed environment (more top-down) changes the focus in the manager's meaning creation from content, i.e. employee target quality and buy-in, to process compliance, i.e. procedures and directives.

It is important to understand that the manager does not necessarily experience a permanent loss of meaning, although she would have preferred to stay with the more inclusive target setting process. Sensemaking is an on-going process and new meaning is generated by the manager by enactment of the environmental changes. In other words; the 'old' way of making sense would not be meaningful under the changed circumstances, because there would be an unacceptable mismatch between practice and procedures to which the manager would stand responsible towards senior management. Target setting is still meaningful to the manager, but the story has become different.

Nevertheless, a Production employee still perceived the target setting process as at least partly inclusive:

I experienced it as a reasonably open process. Of course, there was a draft because targets had been cascaded down through the system to us employees... So it is more about the interpretation of the targets. How I can affect each of the targets. I can suggest what my personal contribution should be.

So, although the Production manager above saw a trend towards more top-down control of the target setting process, it was still possible to create an environment that the employee perceived as at least partly involving.

Discussion

From the analysis above, we see how, within the same company, different enactment of PA by senior management in different workforces creates different approaches to how rigidly target setting is interpreted and implemented. In fact, while the problem of PA being top-down is a serious challenge to the employees I interviewed in Sales and to a somewhat smaller degree in Staffs and Production, it appeared to be much

less of a problem in R&D. This pattern matches the categorisation of the task at hand in the workforces and their fit towards the two contrasting management regimes; management by measurement and management by delegation.

Particularly the contrast of R&D and Sales is illustrative. Sales, that was the workforce with the closest fit to the management by measurement regime was developing the story of commitment through distribution. Whereas R&D, which was the workforce with the closest fit to the management by delegation principles, twisted the PA scheme towards the management by delegation characteristics by developing a story of commitment through involvement. To draw it up sharply, we may say that in R&D, the standard PA scheme was locally adapted to become an exponent of the opposite of what it was developed to be. PA is by nature a management by measurement tool, but in R&D it is in the target setting phase twisted towards becoming a management by delegation tool. The result is a better fit between the task at hand and the use of the tool in R&D, but at the cost of management control in target setting and at the cost of substantial differences in the practice of PA across PharmaComp, despite the fact that PA has been installed as a uniform standard process in the company.

The four different stories of target setting from Sales, R&D, Production and Staffs can be described as a two-dimensional outcome of the balance between how involving the target setting process is and how top-down controlled it is. In an ideal world, a target setting process should be both involving, for the benefit of employee commitment, and controlled, for the benefit of strategic alignment across organisational units and individuals. However, none of the four stories place themselves as fully supporting both purposes.

It seems that there is a dilemma, or trade-off, between involvement and control. The involvement story from R&D and the distribution story from Sales constitute the two clearest cases of the combinations low involvement/high control and high involvement/low control. The manipulation story from Staffs and the mixed story from Production constitute the more mixed balance where there is a high degree of control but where the managers' strategy is to make the process appear less controlled by pretending there is less top-down control than there actually is. This increases employee commitment compared to the Sales story, but at the cost of transparency of the process.

This dynamic feedback element of different articulations of problems in PA as a consequence of the different patterns of local adaption can be illustrated in the meaning creation feedback model. Figure 7-1 compares the involvement story from R&D with the distribution story from Sales. Here we see how commitment to targets takes two different paths in the two workforce examples, both in terms of workforce filter, meaning scheme challenge, meaning creation activity and PA problem articulation or re-enforcement. The feedback loop was empirically visible in that the lack of involvement in Sales was generating alienation and a non-committing compliance focus on behalf of the Sales employee, thereby articulating other problems in PA, particularly related to lack of commitment, i.e. crowding out of intrinsic motivation. In R&D, the general transparency problem in exploratory R&D (targets are short-lived and the future unpredictable) gets reinforced because managers are not capable of installing a strong link between hierarchical levels' targets when they let go of control in target setting. So also here, the feedback loop was visible.

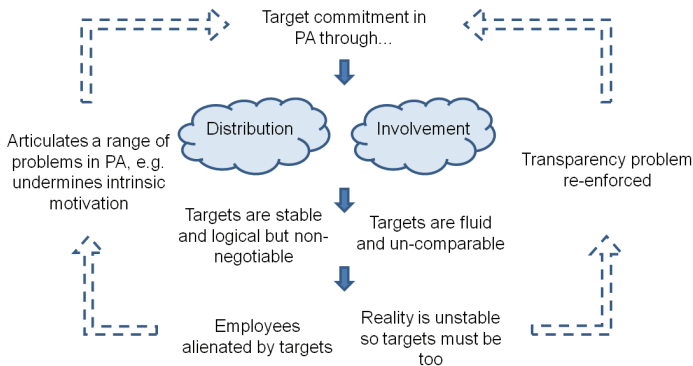


Figure 7-1: Distribution and involvement - alternative ways to commit targets

In the commitment through distribution scenario, the manager's role becomes more directive, as the enactment of the environmental constraints prioritises compliance over involvement. This is meaningful to the manager, because process compliance is an important parameter for the manager's own evaluation - more important than employee involvement in target setting. Or, in other words and with reference to a problem raised by Austin (1996); targets are almost always indirect representations of real value. Thus, the manager is measured on process compliance (does the

manager fulfil all target setting signatures before the deadline) rather than target setting quality (are the targets meaningful to the employees and valuable to the organisation), which is more on the agenda in the commitment through involvement scenario.

7.3 The Challenge of Handling Behaviour in Target Setting

One particular challenge that many managers and employees reflected on through the target setting phase interviews was how to handle behaviour. Let us start by revisiting the definition of behaviour as presented in the introduction chapter; by behaviour I mean *the way a person acts, reacts and interacts with others*. The definition is deliberately very broad because the concept of behaviour also is in the minds of managers and employees. A narrower definition of behaviour may hold analytical advantages but would not capture the vagueness of the concept in the practice of managers and employees in PharmaComp. The concept of behaviour is basically capturing everything that is not included in the term "business performance" which relates directly to outcome or results. Thus, there are roughly speaking two dimensions in rating employees; one is an evaluation of outcome (what), the other is the evaluation of behaviour (how).

Below, four different stories from the four workforces serve to illustrate patterns in meeting the challenge of handling behaviour in target setting. I have labelled these stories the story of meaning through relocation, taken for granted, concealment and adaption.

Meaning through relocation

Consider the following dialogue with a manager in Staffs who had just explained to me how behaviour had an important impact when ratings were ultimately decided:

Staffs Manager: "She did not have any behavioural targets." Interviewer: "But if you know it will have significant impact on her rating, why not include it in the targets?"
Staffs Manager: "We could do that. But it is incredibly difficult to measure it, you know."

For the 2011 PA process that I investigated, the Staffs area that this manager worked in did not operate with behavioural targets. The manager did recognise that it was part of her role to evaluate the behaviour of the employee as an integral part of

establishing a final rating, but in the manager's perception the inclusion of behavioural targets is challenged by lack of measurability. In the literature review we saw how the problem of measurability was raised by Austin (1996) who emphasised how complex, modern workplaces faced serious challenges with measurability because employees' daily work was characterised by complexity, dynamic change and interdependence. This also goes for behaviour which is in itself a dynamic concept. Because behaviour is about actions and reactions, it is also relational and characterised by interdependence.

It is not only managers who are aware of the measurement problem of including behaviour in target setting:

Staffs Employee: "I think it would be difficult to formulate in a PA target. I mean, what is the target? You know, it is also difficult to measure that a project is a success. But it would be more difficult to do it for behaviour:"

It is important to understand that the challenges in setting behavioural targets are not necessarily due to managers and employees being reluctant to have a dialogue about behaviour in general. Rather, a manager explained that she saw such dialogues as essential, but that she preferred to have them outside PA, in the context of the Individual Development Plan (IDP):

We have discussed if the employees should have behavioural targets in PA or in their IDP. Personally, I prefer the IDP - unless you really have a serious issue. I prefer to treat it in the context of development, because then it is more legitimate to talk about.

This manager refers to legitimacy as the issue at stake. As we learned through the literature review, PA systems can be conceived as reward/punishment systems (Kohn, 1999). In PA things are more serious, more committing and with tougher consequences than in the IDP. The manager acknowledges that employees can have a need to develop on the behavioural side, but it is "*more legitimate*" to relocate that dialogue to the context of the IDP which is not a reward/punishment system but a purely development focused process where only the employee and the manager are involved. This is in sharp contrast to the yearly appraisal in PA which generates ratings that have consequences and will be used across a number of other HR processes, with effect years into the future. By enacting the environment of the IDP rather than the environment of PA, the manager will be able to convey a sometimes difficult message in a less threatening way and thereby increase the chances of

creating a common story together with the employee. Such a story will stand a better chance of being meaningful to both. And that story will therefore have a chance of resulting in improved employee behaviour. A Staffs manager explained the inclusion of behavioural targets into the IDP:

X is very active in relation to internal activities. So we have discussed if we could allocate her to tasks where she could increase her visibility... She is the kind of person who does not like to sell herself. We have talked about that several times, so this time I have set some IDP targets for her about 'selling herself to management'.

As this example illustrates, we should not necessarily see the manager's concern with behavioural targets in PA as a sign of the manager being afraid of the difficulty that a dialogue about behavioural targets would incur. Rather, the manager has a more fundamental concern; whether a behaviour dialogue will have better chances of providing successful results if it is taken within a less committing context than PA. The manager is interested in improving behaviour and simply believes that this is best done by relocating the dialogue from PA to the IDP, unless there are serious issues with the employee's behaviour. In the latter case, PA is the better context, because this is the system upon which tougher sanctions rely. However, this leaves us with two important questions in relation to sensemaking in target setting.

First, if the manager believes that the rating element and punish/reward character of PA works detrimental to the purpose of improving behaviour, why would we then believe that it would be any different for business results? Logically, for business performance one could argue in parallel with what we just heard about behavioural development; unless there are "*serious issues*" with business performance, there is a better chance of a positive business improvement in a less threatening context than a reward/punishment system. This argument is in line with another problem in PA, suggested by Coens and Jenkins (2002), that PA systems are basically designed to handle the 5% low performers while the rest of the employees would be better off without it.

So why do some managers see it as more appropriate to include business improvement targets in PA, than behavioural targets? Likely, they do so because it is more meaningful to them. We have in the above quotes seen how the exclusion of behavioural targets is due to considerations about legitimacy and challenges with measurability. If behaviour is difficult to measure and it is therefore a problem to discuss it within a reward/punishment paradigm, then the task of incorporating

behavioural targets into PA is not meaningful to the manager. It will be more meaningful to create an alternative story where behaviour is something fundamentally different from business results, something more development oriented and thus something better handled in the context of the IDP. As long as behavioural targets are not enforced by senior management as being mandatory (which they would become from 2012), the manager can decide to relocate the dialogue on behavioural performance to the IDP.⁵⁶

Second, the transfer of the behaviour dialogue to the IDP context leaves us with a question related to what could be referred to as the more general problem of transparency in PA (DeNisi and Pritchard, 2006). The managers I interviewed generally acknowledged that behaviour plays a significant role when the final rating will be established. And we know from the rating matrices from Sales and Production which were discussed in a previous chapter that managers are directed to include behavioural evaluations as a significant component in the final appraisal rating. So if behavioural targets are not included in PA targets, where does that leave the employee in terms of transparency of what is actually going on in the PA evaluation?

The two questions in fact constitute challenges not just to the logic of how target setting and evaluation occurs but also to the role of the manager. Is the manager's role to work to improve business results and behaviour? This is what the Staffs manager is aiming at in the quote above and why she prefers to have the behavioural dialogue outside PA. Or is the manager's role to make the PA evaluation as transparent as possible? The latter is what the inclusion of behavioural targets in Production, and from the 2012 PA process also in Staffs, is about. Including behavioural targets in PA is an attempt to make it transparent that behaviour does affect the final appraisal rating, so that the employees understand that behaviour is not just a subject for un-committing dialogue with the manager, but an important element in reward/punishment decisions within the framework of PA.

⁵⁶ Kohn (1999) argued that any reward system is in reality also a punishment system, since in such schemes any lack of reward would be perceived as punitive. Generally, Kohn sees punishment based systems as deteriorating the dialogue between managers and employees and undermining employees' intrinsic motivation. Therefore Kohn would surely agree with this manager that a dialogue about behavioural improvement would stand much better chances of providing a positive outcome in the context of the non-punitive IDP context than in what he would see as the punitive context of PA. However, Kohn would see the same logic be valid for business targets.

Thus, we have two different imperatives, each meaningful on their own terms, about handling behaviour in target setting. We could call them the outcome imperative (focusing on how behavioural change is best achieved) and the transparency imperative (focusing on increasing transparency for employees). Both imperatives can be meaningful to managers and employees, depending on what cues they extract from the process and the way they enact the environmental frames each imperative refers to. As we have seen, the tendency seems to be that senior management favours the transparency imperative, so that from 2012 all workforces are enforcing the inclusion of behavioural targets into PA. The introduction of this new approach to behavioural performance appraisal in 2012 in some of the areas I investigated, although late in the data establishment phase of my longitudinal study, allowed me to analyse sensemaking processes where they are most observable; when environments change and actors enact new frames of reference to create new meaningful stories about the world in which they operate.

As we would expect, the inclusion of behavioural targets into PA is not without trouble. A Staffs manager explains how she perceived the enforced inclusion of behavioural targets into PA for the first time in 2012:

We have given the 3 behavioural targets 5% priority each this year. We had the three targets cascaded from above. They could easily have been consolidated into one. That would have been more elegant. But we were told that they had to be there as three targets. I would have preferred that they had let us work with it in our teams. I feel that it has all been decided for us.

The manager role identity is linked to influence. The top-down process of installing three identical behavioural targets into PA across a large area in Staffs alienates this manager from the behavioural targets. The manager would have preferred that she could have discussed behavioural targets internally in the team and set targets that she and her team would be able to take more ownership of. By enforcing uniform behavioural targets for all employees across an area, senior management in the area communicates a coherent message to all managers and employees about what they see as desired behaviour. However, at the floor level this challenges the manager's meaning scheme because she sees a need for local adaptation of the behavioural targets for them to be meaningful.

The inclusion of standard behavioural targets is a significant change and so is the challenge to the manager's meaning scheme. It was meaningful to the manager

when behavioural targets were relocated from PA and handled in the context of the IDP where a "soft" dialogue could be held. How can new meaning be created from a changed setup where behaviour is not only transferred to the "hard" dialogue context of PA but also in a way where target formulation is cascaded top-down, out of the hands of the manager? This brings us back to the top-down problem in PA that we have already dealt with in detail above. And again, as we previously saw, new meaning creation in this case has a tendency to move from an outcome oriented imperative (improving behaviour) to a transparency oriented imperative (improving transparency of the target setting-evaluation relation).

We may also ask if the new behavioural targets are after all really increasing transparency. The above described total priority of behavioural targets of 15% seems to be very low if we compare to the rating matrices used for rating consolidation purposes that we saw in a previous chapter from Sales and Production. In Sales and Production, behaviour was a strong moderator of business performance; more in the neighbourhood of 40% than 15%, determined from the matrices. So either behaviour means much less in Staffs than in Production and Sales, or the 15% priority is still a significant understatement of the real influence of behaviour in the rating process.

I will later dig more into the challenges to managers' meaning schemes in evaluating and providing feedback on behavioural targets. For now, we can conclude that in the target setting process, the introduction of mandatory behavioural targets in Staffs sent a message to all managers and employees about what behaviour senior management wanted to see. But at the same time it installed a new environment for the dialogue on behavioural improvement between managers and employees which gave way to new meaning creation processes establishing new stories around the same dialogue. And we saw how this for the managers I interviewed gave way to new challenges.

Meaning through taking for granted

I have above dealt at some length with examples from the Staffs area because they were in many ways exemplary for the challenges in including employee behaviour in target setting. If we now turn to an example from R&D, we see a rather different approach:

R&D Manager: "Besides the business targets, there is a whole lot of stuff about positive attitude, willingness to cooperate and things like that. We do not write that into PA, but we explain to the employees what it requires to get an ME... We do not write it into PA because, honestly, we expect that everyone behaves decently in this firm"

In this quote we see reluctance to incorporate behavioural targets into the PA process. But not only that, the manager is in fact taking the behavioural element for granted to the degree that she sees it as meaningless to incorporate it into the PA targets, since management "expect that everyone behaves decently in this firm". In this story the argument for not incorporating behavioural targets is not that the dialogue about behaviour is better suited by being relocated to the IDP. Here, the argument is that behavioural targets are superfluous. Again, we may refer to the previous analysis of the task at hand in R&D that demonstrated how R&D is a workforce characterised by a strong dominance of highly skilled science experts. Here, professional qualifications and results are in focus, not 'softer' dialogues about behaviour. Behavioural targets are ruled out as being irrelevant. The rationale is not lack of legitimacy but rather it is downplaying the potential issues with behavioural performance.

Thus, according to this story, it is sufficient to explain to the employees what behaviour is required, it is not necessary to write it down in PA. However, also R&D included general, cascaded behavioural targets from the 2012 process. From this we can conclude that at least senior R&D management thinks behaviour is in fact important when rating employees and that for transparency purposes behaviour should therefore also be incorporated into target setting.

Meaning through concealment

As we have already seen, in Sales business targets in the shape of sales index realisation dominates in PA. A Sales manager explains the relation between behaviour, target setting and evaluation:

We have not had behavioural targets but we will have that in the future. But we have had a rating matrix that also includes behaviour... One thing is the indexes, they are easy to measure. But then there is the behaviour. The rating matrix gives some guidance.

In the quote we see considerations much in line with what we saw in Staffs above. Behaviour does matter. It even matters so much that a rating consolidation matrix was in 2010 in use, although behavioural targets were still not applied. The purpose

of the matrix was to assist the managers in their difficult task of overcoming the problem of measurability in evaluating behavioural performance. As we saw in a previous chapter, like in Production the Sales matrix indicated a strong moderating rating effect of behaviour on business results. If behaviour can determine whether a medium business result of ME ends with anything from the lowest final rating NME to the second highest final rating EE, then behaviour is important. But in Sales, extra complexity is added to the picture, because there is a separate bonus system in which only sales results count. In PA, the Sales targets constitute 75% of the total PA priority, and most of the rest is constituted by process oriented targets (call plan establishment and call plan compliance). A Sales manager explains her reluctance to setting behavioural PA targets in such a context:

Personally, I see it as an advantage to have as few targets as possible... We weigh the targets by percentages. I do not see much value in adding targets that the employees would like but that do not have any weighting. Then you would have something that counts zero percent.

We have seen how senior management in Sales through issuing a rating matrix has installed a clear demand for including behaviour in the PA evaluation process. However, the Sales manager is challenged to create a meaningful story on the following premises: a) for sales employees, all bonus is based on sales index realisation; b) in PA, it has been decided by senior management that sales index realisation constitutes 75% of the overall target weight; c) the residual 25% is distributed across a number of elements of which call plan establishment and call plan compliance are the most important; d) senior management has distributed a rating matrix that indicates a strong importance of behaviour evaluation in the establishment of final ratings, but there is no target weighting left to reflect this importance in the targets.

How does the manager create meaning out of this delicate situation? She does so by creating a positive story around a strategy of "*as few targets as possible*". According to this story, the PA role of the Sales manager is to keep the focus on the employees' business targets realisation regarding Sales and call plan compliance. It is acknowledged that behaviour is important for PA evaluation but there is no room for prioritising it through PA targets within the current framework in Sales, so therefore it is better left out, which further has the benefit of reducing the number of targets to something that the employee (and the manager) can maintain an overview of.

I also asked a Sales employee if she would like behavioural targets to become included in PA. She confirmed this and explained her reflections on what such targets should cover:

It should be more about human values. How I have worked as a colleague. How I work with my customers. I know it is going to be difficult because it is someone else who is going to do the evaluation. But if you could just find a way to do it so that it was equal to all.

The employee would like to supplement the current focus on business targets with behavioural targets because this would provide her with a more adequate feedback on her contribution to the organisation. However, the employee is also well aware of the problem of measurability that the Sales manager referred to. And she couples this to the problem of comparability. The general problem of comparability was according to Coens and Jenkins (2002) derived from the measurability problem due to a false assumption that it is possible to provide objective, comparable evaluations of employee performance across large complex organisations. While Coens and Jenkins were referring to a general PA problem, the employee above refers more narrowly to a particular lack of comparability when behaviour is concerned. Lack of comparability is ultimately a challenge to the fairness of the process if people are rated relative to each other based on comparison.

The inclusion of behavioural targets in PA is challenged by the problems of measurability and comparability. But it is also challenged by the problem of transparency as it has been laid out by multiple authors, e.g. DeNisi and Pritchard (2006). No matter if behavioural targets are in PA or not, they play an important role in the later evaluation process. Leaving behavioural targets out of PA targets may be a way to avoid dialogue with the employee about the measurability and comparability problems in the target setting phase, but as long as behaviour is still included in the rating process, the measurement and comparability problems are equally present, although now supplemented by an increased lack of transparency.

In fact, if the Sales rating matrix is adhered to, behavioural evaluation comprises around 40% of the real weight in final appraisal. So why is there room in PA for representing call plan establishment by 5% in the targets, but no room for representing a 40% rating weight on behaviour into the targets? The struggle to handle this dilemma is probably the reason why Sales Management from 2012

decided to include behavioural targets in PA, just like Staffs and R&D. However, in Sales this was still without any weighting:

Sales manager: "We will not give them [the behavioural targets] any weighting the coming year. But they will be there and I am fine with that. I think everybody is. It is good to focus on behaviour. But again, targets must be measurable. Whether you are a good team player - that is a very subjective feedback."

Again, the measurability dilemma gets crystallised in this quote; behaviour is important for the final rating - so we include behavioural targets in PA. But behaviour is subjective and difficult to measure - so we do not attach any weighting to the behavioural targets. In a sensemaking perspective, we could argue that Sales management is here drawing on the important concept of plausibility rather than accuracy when creating a new story to convey to the managers and employees; that on the one hand behaviour is important so we include it in the targets, but on the other hand we are uncomfortable with the subjectivity of it, so we downplay its importance by avoiding any weighting of behavioural targets. We could call this 'symbolic compliance' in the sense that behaviour is included in target setting in compliance with the general trend across PharmaComp, but that the lack of weighting shows how Sales in practice still operates within 'the old' regime of pure business target focus.

The downplaying of behavioural targets in Sales may also have to do with the fact that Sales has a separate bonus system unrelated to PA:

Sales Manager: "There will probably be some that say that they do not have a close relation to their PA... Mostly PA is about the honour. Their entire bonus is related to sales. So when we reach final appraisal, they already have had their bonus paid out. So there is no reward for PA."

In Sales, managers and employees struggle to make PA meaningful because all the monetary reward relies on the parallel bonus system. When meaning is not to be found in the link to reward, meaning must be found somewhere else. This manager is indicating that meaning can be found by downplaying the importance of PA and by distancing oneself to it, so that the loss of meaning becomes less threatening. Such a story decreases the frustration among employees. They are not seriously challenged on their identities as sales employees, because this identity is more strongly related to the hardcore sales index realisation which the employees perceive the organisation gives first priority to, since this is what is rewarded monetarily. But of

course this kind of sensemaking at the same time devalues the significant resources spent on the PA process. If it is not important, why bother?

The task at hand in Sales is so much focused on realisation of sales targets that managers see it as irrelevant to include behavioural targets into PA. But by leaving behavioural targets out of PA while letting behavioural evaluation constitute 40% of the rating evaluation weight, they are in effect concealing the importance of behaviour in PA. The favouring of hardcore, measurable sales targets in PA is yet another example of how the management by measurement paradigm has a stronger influence in Sales than in the other workforces, and particularly compared to R&D. When a management by measurement regime is in operation, managers will focus on targets they can measure. Here, behaviour is the trickier element which is concealed to be able to maintain a meaningful story of objective measurability in PA.

Based on the above stories of behavioural target setting we can see how managers and employees are struggling with two opposing imperatives; the requirement for transparency against the requirement for measurability and comparability. Requirements that we know from the literature review also constitute problems in PA and which are inherently difficult to meet. In sensemaking terms we could call it the balance between plausibility and accuracy. So far in the stories we have looked at, it seems like plausibility beats accuracy, just like Weick (1995) prescribed it would in a sensemaking perspective.

Although we previously learned from Austin (1996) that almost all measurable targets are in fact only indirect representations of the real value that the organisation wants, we also learned that some things are more difficult to measure than others. For the Sales manager above, behaviour definitely is one of them. As the approach to incorporation of behavioural targets is obviously still in development across PharmaComp, there are variations across workforces and across performance cycles to how the balance of plausibility against accuracy will look like and managers' sensemaking is therefore continuously challenged by changes in the intra-organisational environment.

Meaning through adaption

It could be raised as an argument that both in Sales, R&D and Staffs the use of behavioural targets has only very recently come to the fore of the agenda, and that

these workforces are therefore still not mature in their approach to handling behavioural target setting. In the one year period I studied the 2011 PA cycle, behavioural targets were not yet implemented but they became so for the next period (2012). Could the dilemmas and uncertainty about the manager's role and objectives not just be caused by the fact that this was still very new to the managers and employees? One way to investigate this is to look at an example from Production, where behavioural targets have been in use since 2008. One Production manager explains her experience with the use of behavioural targets:

We have had these targets for 3 years now and it is my experience that if you just go with the standard text... it doesn't become tangible and there is very little in it for the employees.

The experience that this manager has had with behavioural targets is that if you just cascade behavioural targets top-down, it will alienate rather than motivate employees. This experience supports the reflections we previously saw when looking at the challenge of commitment in PA with multiple references to the problem of PA being a top-down process.

Also in Production, the incorporation of behavioural targets has changed the role of the manager in relation to setting PA targets from focusing only on business targets, to the incorporation of broader, but standard, behavioural targets. The above production manager did not experience a simple cascading of behavioural targets as being particularly meaningful. Therefore she installed localisation activities to allow meaning to be created around the behavioural targets at the local level. This is not in contrast to the official Production guideline that actually promotes local moderation of the behavioural targets. So, after all it looks like the experiences from three years of working with behavioural targets have taught Production to be flexible around local adaption of standard behavioural targets, although Production has the same transparency issue as the other workforces with a mismatch between behavioural target weight in PA target setting, and the weight behavioural evaluation has in final appraisal rating, judged from the Production rating matrix.

Discussion

In Figure 7-2 below, the four meaning creation stories of including behaviour in target setting are depicted in a diagram, based on the contrasting transparency and outcome imperatives. By weighting outcome high in the shape of behaviour

improvement, the stories of adaption, taken for granted and relocation place themselves in the right side quadrants. The adaption story from Production is the model that most transparently balances the objective of improving and measuring behaviour with the ability to allow local adaption.

By weighting business targets higher than behaviour in the communication to employees, but use concealed behavioural targets as a strong moderator when later appraising employees, the story of concealment in Sales places itself in the lower left quadrant. None of the stories place themselves in the upper left quadrant.

However, the recent development in the Staffs area I investigated seemed to move towards more focus on standard, top-down cascaded behavioural targets that are not adaptable in the local units. If all employees in an area have exactly the same standard behavioural targets, the transparency would be at an optimum. However, we have also seen from multiple quotes how managers and employees struggle to find standard behavioural targets meaningful and therefore the transparency would likely be increased at the cost of the outcome perspective in terms of improved behavioural performance. The expected result would likely be employee alienation from the behavioural targets, although I have not been able to investigate this in the current study because the standard behavioural targets were installed after the data establishment of the study.

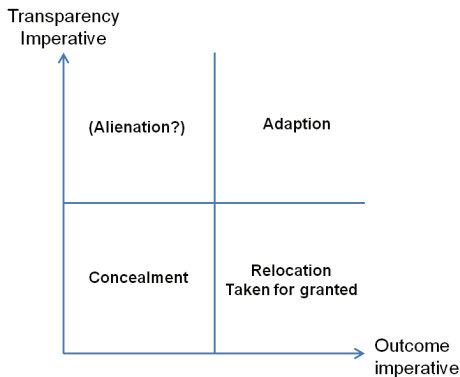


Figure 7-2: Behavioural targets approach dependent on imperative

A more dynamic way to analyse the stories would be through the meaning creation feedback model where we include management priorities, workforce filter, meaning

creation through establishing different stories and the dynamic effects of feedback from new meaning creation to articulation of other problems.

If we compare the inclusion and adaption of behavioural targets in PA in Production with the story from Staffs where behaviour was relocated to be dealt with in the IDP, it becomes easy to see the importance of environmental enactment on manager and employee meaning creation and the effect on the different feedback loops. The two scenarios are illustrated as two different paths through the meaning creation feedback model in Figure 7-3 below, both taking outset in the measurability problem of behaviour.

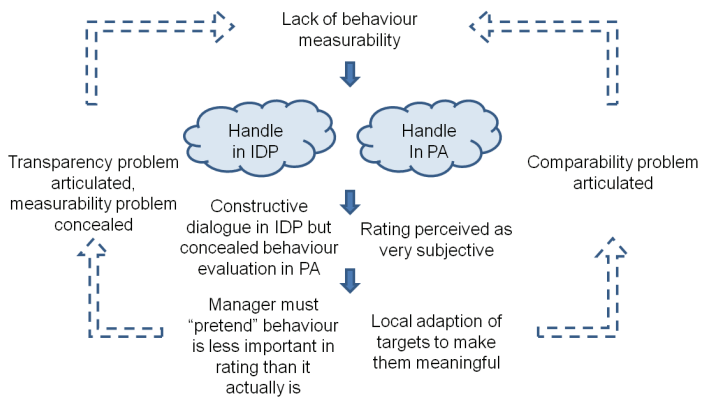


Figure 7-3: Relocation and adaption - handling behaviour in target setting

Although the two approaches are very different, they are both meaningful on their own terms. They also both kick off feedback processes that articulate other problems in PA. We saw how the relocation story in Staffs articulated a transparency problem in PA by not incorporating behavioural targets in PA, although behaviour was included in the performance evaluation. Similarly, we also saw how local adaption of standard behavioural targets was making targets more meaningful locally. However, this is at the cost of articulation the comparability problem in PA, because local variations in interpretation of the standard targets will make evaluations less comparable than if no local adaption had taken place.

The four workforces differ in terms of local adaption of approach to including behavioural targets, their environmental enactment, stories of individual meaning creation and feedback process outcome. None of these important differences would

be visible in a narrower, classical analysis of measurability problems in PA where the environmental factor and dynamic feedback processes would not have been included.

Let us finalise the analysis of managers' sensemaking in target setting with a quote from an HR informant who has been very active assisting management in PA improvement initiatives within a Staffs area:

I can make un-ambitious targets and over-perform. And then somebody else can have incredibly ambitious targets and who doesn't quite reach them. It puts too much pressure on the target setting so if that's not done right, the whole system fails.

If targets were always taken literal and rating always was a direct consequence of delivery on targets and only this, then a thorough target calibration process would be required or the whole system would fail. However, as we have seen and will see in more detail in the coming chapters about Midyear and Final Appraisal, there is much more to feedback and evaluation than what is documented and signed off at the target setting meeting in the beginning of the year.

8 Local Adaption and Meaning Creation in Midyear Review

8.1 Introduction

A normal perception of a manager's responsibility is to support and manage her employees through a range of different activities including the provision of feedback on a regular basis. However, in the context of the PA scheme, there is also a very tangible, once-a-year activity; namely to conduct a 'midyear review' with each employee to follow up on and provide feedback on the employees' progress on PA target realisation.

Four important challenges were identified in the midyear review phase of the PA process in PharmaComp. The first challenge concerns conveying meaningful feedback, particularly where feedback is negative or where there is a differences in the perception of manager and employee about what performance events are important or what conclusions can be drawn about the performance of the individual from those events.

The second challenge in midyear review is about the tangible articulation of feedback. I investigated whether managers used preliminary ratings at midyear review meetings and found that none of the managers I interviewed did this. Also, none of the employees I interviewed requested preliminary ratings. On the one hand this was not surprising to me since in my years as manager in PharmaComp I have only very few times been requested by employees to provide a preliminary rating. On the other hand, it made me curious to understand how feedback was then provided in alternative ways so that managers and employees felt that a clear communication of performance status had been conveyed. As it turned out, the managers and employees I interviewed had important but different views on the art of articulating preliminary performance feedback in a way that was maintaining a meaningful relation between manager, employee and PA without referring to an overall rating.

The third challenge in midyear review concerns situations where managers or employees experience that the others, themselves or both parties, are careless in the PA process. During the interviews, stories of carelessness were often tangibly connected to lack of preparation for PA meetings. Particularly where there was a

perceived difference in level of preparedness between manager and employee, this was seen as a challenge to meaning schemes.

The fourth challenge in midyear review was situations of disruption in manager-employee dyads. Four of the employees I interviewed experienced a manager change during the 2011 PA cycle. Thus, another manager was taking over the responsibility of evaluation although she had not been involved in setting targets, agreeing on priorities or changes during the first half of the year and generally had a less intimate knowledge of the employee and her situation.

8.2 The Challenge of Conveying Meaningful Feedback

The challenge of conveying meaningful feedback concerns the actions managers and employees conduct, and the stories they retrospectively create, of how conveyance of feedback can take place in ways meaningful to the parties. Below, two different stories serve to illustrate patterns of local adaption across workforces. I have labelled these the story of conveyance through distancing and the story of conveyance through enlargement.

Conveyance through distancing

Most people can accept praise without making an argument. But what if employees are reluctant to receive 'negative' feedback? Is it possible for managers to establish scenes where such feedback can still be meaningfully conveyed within the framework of PA? I discussed this with a Staffs manager who felt that there had been some disappointments in an employee's performance through the first half of the year and who wanted to provide feedback regarding that:

That is more difficult for her [the employee]. You can see that from her reactions. She asks; 'Is that what we are going to talk about?' And then I say that yes that is what we are going to talk about. Because it is a part of evaluating what has been going on... She gets irritated because she finds it less important.

Providing feedback is a two-way activity. If the employee is not ready to receive feedback that contains critique, chances are that the feedback experience will not be good for either party. The manager experienced that the employee reacted negatively to the feedback at the meeting. And the employee also experienced the meeting atmosphere as disappointing:

Staffs Employee: "In my kind of job, you work a lot on your own. You do not see your manager much. So when you finally have a chance to talk to her, it is a bit tiring that it has to be such a less cheerful experience."

To the employee, the problem with the meeting was that the manager's focus was on things that may well have been documented in targets in PA but that the employee found to be of less importance. According to the employee, too much time was spent at the meeting on discussing small things that went wrong instead of all the bigger things that were successful:

She thinks I have been a bit late on some of my responsibilities. Because I have been so very busy... Of course I could have been clearer in my communication with her about it, but when you are putting a lot of pressure into a tight schedule, and you still make it within deadline, but you are just not delivering materials those 3-4 days before the meeting that some prefer, then she doesn't think it is good enough... then she has a tendency to go into micro management and that doesn't work well with me.

What the employee experiences here is what Coens and Jenkins (2002) called the problem of manager centric feedback in PA.⁵⁷ To Coens and Jenkins, it is a paradox that modern organisations verbally promote empowerment but at the same time install PA schemes that are systematically undermining empowerment by putting responsibility for feedback provision solely on the manager. In contrast, Coens and Jenkins (2002: 22) proposed that "as healthy adults, people need to be responsible for their own feedback, performance and development".

The manager in the quote above is focusing on something that is important to her but which is clearly of much less importance to the employee. The manager and employee have different perceptions of what should rightly be prioritised for discussion at the meeting, but it is the manager who takes the lead and provides feedback on what she finds important. Receiving feedback on something perceived as less important is challenging the employees' meaning scheme. The feedback is perceived as less relevant and the employee's attitude becomes negative. The employee extracts cues from the situation and establishes a story around how the manager has the wrong focus and is micromanaging the employee instead of letting go of minor details and focusing on the big picture and successful achievements,

⁵⁷ Coens and Jenkins' (2002) argument that feedback is too often manager-centric is followed by a claim that employees should instead take responsibility for collecting their own feedback from many sources. They should not just have feedback from the manager, who only represents one perspective and who is often not the best informed. According to Coens and Jenkins (2002) feedback from people without a hierarchical interrelationship is often much more effective in installing improved performance because the inequality of hierarchical position can make it very difficult to have an equal dialogue, as the present example also illustrates.

Next, the challenge to the employee's meaning scheme in turn becomes a challenge to the manager's meaning scheme. The manager experiences that the employee gets irritated and is reluctant to take in the feedback. Sensemaking is social and ongoing and we see how this is leading to new sensemaking processes with the manager. The manager continues:

I told her that if she did not find it important herself she should still recognise that it is important to others... Next time, someone higher in the hierarchy could be involved and I know for sure that they will definitely have an opinion about these things... And that way, it will also become important to her.

From the categorisation of the task at hand in Staffs, we know that Staffs was in a middle position between the two contrasting workforces, R&D and Sales. The manager attempts to stay loyal to the PA scheme and provide honest feedback, but the mixed fit to the management by measurement paradigm of PA makes it hard to simply deliver objective, undisputable feedback. The medium complexity of tasks, the low newness of tasks, the high degree of specialised knowledge and the high interdependence are all factors that challenge the ability to maintain a strict management by measurement position.

So how does the Staffs manager handle the employee's negative reactions to the feedback? She handles it by offering new meaning of the feedback through alternative sensegiving. The manager is not insisting to persuade the employee that it is the perception of the employee that is wrong. Instead, she tries to fertilise the ground for alternative extraction of cues by generalising the problem to something beyond the relation between the manager and employee. The immediate manager-centric feedback was not meaningful to the employee, and the feedback session was therefore also challenging in the perception of the manager. By generalising the issue to be something that may or may not be literally true, but that will affect the employee negatively anyhow because "*others*" (more powerful people) will consider it as true, the manager retrospectively offers new meaning by distancing herself from the critique of the employee. Instead, she turns the story of critique into a story of a helping hand. The message of this new story is; 'I am not saying this to punish you. I am saying this to help you to avoid getting punished by others later'.

We have now seen how an employee met the challenge to meaning creation by first rejecting the sensegiving attempt of the manager by being reluctant to accept the feedback as relevant. And second, how this reluctance in itself challenged meaning

creation of the manager so that the manager through sensemaking created an alternative story that was more meaningful to the employee. By changing the story the manager was rationalising her choice of performance events and her extraction of cues from those events differently than in her first story, thereby allowing the employee to find the choice more meaningful. The feedback was in a way still manager centric, but it was not unaffected by the employee's perception and reaction. Through the manager's translation of employee perception, a new and commonly more meaningful story was created to substitute the first story, which was only meaningful to the manager.

The sensegiving offer from the manager did work in the sense that the employee concluded that in the future she would make sure that those 3-4 days' notice would be adhered to. At the later final appraisal interview both the employee and the manager concluded that the employee did in fact stick to this promise in the following half year where I followed them. So the manager feedback did bring the desired change in employee behaviour.

Conveyance through enlargement

Let us now look at an example of how an R&D manager experienced that the feedback dialogue needs to be restarted because PA is experienced to be thwarting the employee's behaviour:

It [a negative effect of PA] can be seen when people are more engaged in fulfilling their PA than in thinking about what is the overall purpose of it all... then they have to understand that they are connected to some higher objectives. So they need to understand their goals in the context of that rather than narrowly what is in their own PA.

The R&D manager is taking a holistic approach to providing feedback by not only focusing on performance related to PA targets, but performance in a broader sense. She is indirectly quite clearly referring to elements of a management by delegation paradigm in that she pulls in the need to be able to understand the higher objectives and thus be able to independently translate this into a company-relevant view on prioritisation of her tasks, rather than clinging to a strict adherence to what is documented in PA.

In fact, the manager experiences it as detrimental to good performance if employees focus narrowly on their tangible PA targets rather than on the less tangible "higher

objectives". This enlargement of feedback and evaluation criteria makes it possible for the manager to maintain meaning from the feedback interaction because it allows her to convey the feedback she thinks is relevant in relation to the overall purpose of daily work, no matter if the documented PA targets are met or not. Again, we see how in R&D, the management by measurement paradigm of PA is twisted towards the management by delegation paradigm to which we saw a better fit for R&D, based on the task at hand in this workforce.

If the employee is strictly focusing on the literal formulation of targets in her PA and the manager is focusing on the broader contribution and organisational value, it is not difficult to see how frustrations will be the likely outcome for both from the feedback dialogue. And vice versa; when employees are focusing on delivering business value in a broader sense and managers are sticking to what is perceived as details of PA targets, it is equally frustrating to the employee.

In the first case the employee will perceive the feedback as unfair, because it is not relating to the targets she thought they had agreed on. And the manager will perceive the employee's focus on the targets as narrow-minded and counterproductive in a broader context. In the second case it is the other way around. Establishing a meaningful scene for feedback is just as much about what to give feedback on as it is about how to give feedback. Not in an absolute sense, but in a sense that the involved manager and employee perceive as meaningful.

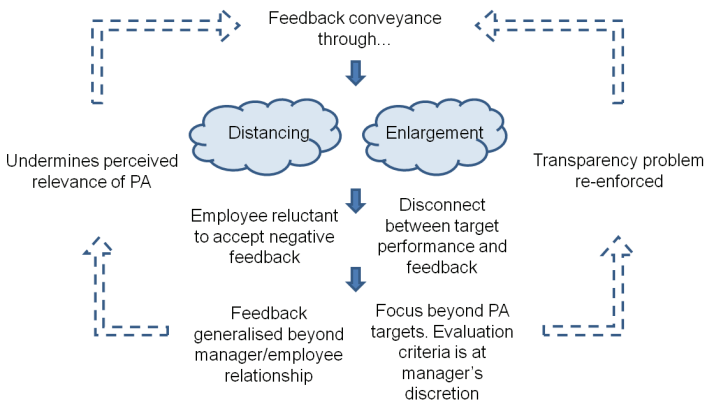


Figure 8-1: Distancing and enlargement - two ways to convey feedback

In Figure 8-1 above, the distancing and enlargement approaches for conveying feedback in Staffs and R&D are depicted in the meaning creation feedback model. The Staffs manager used distancing as a meaning giving mechanism to transform the feedback firstly perceived as negative by the employee into a story of a helpful hand. By referring to higher level stakeholders' expected perception, the manager could stick to the literal targets but distance the critique to be something beyond the relationship between the manager and employee. The downside of this approach is that it distances the feedback to the employee and thereby adds to undermining the employee's perceived relevance of PA.

In contrast to the sensegiving through distancing of the Staffs manager, the R&D manager followed an enlargement approach where PA feedback was just partially based on tangible targets in PA. In fact, the manager found it detrimental to the purpose of performing well if one had too much focus on targets in PA. This enlargement of the basis of PA feedback gives the manager the opportunity to include other sources and measures when conveying feedback. However, the feedback and evaluation criteria will consequentially be based on the managers' discretion, and the protection from arbitrary feedback that theoretically is provided by a strict focus on performance targets is undermined. In both cases, the transparency problem in PA is articulated, because performance feedback is only partially connected to PA targets.

8.3 The Challenge of Using Rating in Feedback

If PA schemes are expected to be controlling employees' motivation and performance through their link to extrinsic reward, one should think that employees would be very interested in using the midyear evaluation to get a status on where they are in terms of overall rating. It would even be logical to expect employees to try to negotiate a clear answer out of the manager on the question, 'what do I have to do from here to get a 4?'

Similarly, if ratings are perceived as important levers for employee motivation and performance, one would intuitively expect it to be meaningful to managers to send a clear signal to employees about how their performance should be improved for them to obtain a certain rating in accordance with their ambitions and potential. In fact, if

one is generally supportive of rating employees as a means to improve organisational performance, it would seem logical for managers to talk as much as possible about the rating with the employee, to keep the employee focused and motivated. Thus, if one believes that an employee rating is a strong motivator, why not use the midyear session to talk about where the employee stands in terms of rating and what it would require from the employee for the rest of the year to get a higher rating?

However, contrary to what could logically be expected, this was not the case in any of the manager-employee dyads I studied. Nowhere did managers and employees discuss a preliminary rating at the midyear meeting. Managers did not take initiative to discuss overall rating, and employees did not request it. This raises the question of why managers and employees working within a PA scheme that has as its main output numeric ratings, find it more meaningful to discuss performance feedback without referring to an overall rating. Meaning creation processes around articulation of performance feedback without referring to a performance rating became an interesting object for studying the dynamics of how problems in PA challenge meaning schemes of managers and employees. Thus, when it comes to handling the challenge of using rating in feedback, there was a striking similarity between the four workforces in my study in that none of them used preliminary ratings. However, just because the outcome is the same, it does not mean that the reasons for the outcome are the same across workforces.

Below, six different stories from the four workforces of why preliminary ratings are not provided by the manager or requested by the employee will be presented. Each of the stories relate to specific problems in PA previously dealt with in the literature review.

Each story also represent different cues about why a system zoomed in on producing a one-digit final rating seems to avoid any direct connect to the same rating at the midyear review. There could surely be more cues and stories, so these six are not the full picture. But they are interesting examples of how sensemaking processes in relation to feedback and midyear review can take departure in local adaption, different interpretations and meaning creation. As such, each story is a metaphor for actors' attempts to establish meaning when challenged by problems in PA. The action stays the same; preliminary ratings are avoided. But the rationalisation of the

action is different, because it refers to different problems in PA and different local adaption of PA.

The six different stories are: a) ratings do not have a positive impact on employee motivation, b) managers and employees experience a lack of control of rating outcomes that makes it meaningless to discuss preliminary ratings, c) ratings are generally irrelevant, d) preliminary ratings are not part of the process, e) managers are shy of conflict and want to keep the meeting in as easy an atmosphere as possible, i.e. avoiding tough discussions about ratings that may be disappointing to the employee, and f) the organisation sees the introduction of preliminary ratings as too costly in terms of management resources required.

Ratings are not positively affecting employee motivation

First, let us start by looking at the story about ratings having no positive effect on employee motivation. A manager in R&D explained at the midyear interview:

I think the employee has a right to know where they stand. But I don't use a rating, because I can't really use that number for anything. An employee for whom I think things are going really well, that can still be a 3. But a 3 will be perceived as negative. A 3 is not perceived as good, because when we have the employee satisfaction survey, then it needs to reach 4 to be ok. It is impossible to explain why a 3 is then suddenly fine when it is the employees who are rated. That simply doesn't work. So I don't use those numbers. Or as little as possible, I would say.

This manager is very clear in her translation of why she wants to avoid discussing rating in the context of midyear review. She knows from past experience that employees will simply not accept a story about how an ME rating is a good rating, although this is what is clearly stated in the corporate PA process design documentation and mirrored in the "ideal", "desired" or "expected" distributions that different areas have developed to guide their managers in the rating process. According to this reasoning, since the majority of employees will get an ME, a majority of the employees will also be disappointed by a preliminary rating, which will negatively affect employee motivation. The manager's argument is related to the problem of disappointing ratings (Poon, 2004) according to which the drop in motivation from disappointing ratings is so persistent that it can even be measured in a significant increase in intention to quit.

In reality, this manager is rejecting the value of appraisal altogether, and with that the management by measurement doctrine as it has been implemented in PA in

PharmaComp. It is hard to see how the above argument could be valid for midyear review only, and not also for the final appraisal rating. This manager simply believes that the organisation is better off without discussing ratings, because ratings deteriorate employee motivation. We should not be surprised to see this kind of reasoning in R&D because from our analysis of the task at hand and its affiliation to management regimes, we found that R&D was the workforce closest related to a management by delegation regime and the furthest away from a management by measurement regime.

But why do employees perceive an ME as a bad rating? According to this manager's story, employees notice that when managers are "evaluated" by their employees in the yearly employee satisfaction survey, anything below an average of 4 on a 5-point scale is considered a potential issue. There is a perceived incongruence in the use of the 5-point rating scale between manager and employee evaluation contexts that make employees suspicious about management's and HR's claim that an ME is a good rating when it concerns employees, but a bad rating when it concerns managers.

I also asked one employee why she did not ask for a preliminary rating at midyear, and she replied:

The motivation comes from something else... from being part of the projects. And from seeing that the other people are burning to answer these questions we work on. About how the drugs work and if we can do it better.

The indifference to the rating on part of this employee is directly linked to motivation. Motivation comes from something else than PA ratings. This explains why the employee does not request a preliminary rating; it would not be meaningful because it is not important for the employee's motivation. When the employee and manager share the feeling that ratings are not important for building motivation, and maybe even counterproductive to intrinsic motivation, it would be illogical to introduce the rating concept any more than absolutely required, i.e. at final appraisal where it is mandatory and unavoidable.

Ratings are beyond the managers' control

The second story is about how managers and employees experience a lack of control over rating outcomes that makes it meaningless to discuss preliminary

ratings. I asked an employee if she after the midyear review meeting had a feeling of where she was with regards to rating:

Production employee: "I don't have any feeling about that yet. It is also very dependent on how the entire area is rated. I have a feeling that won't be too high this year... it doesn't have anything to do with my project, but I can foresee that it will somehow hit me anyways."

Similarly, I also asked a manager if she would be forced to identify people to be rated AE:

Staffs Manager: "We have been forced to do that before. I don't know if we will be that again."

The quotes indicate that this manager and employee experience the final rating to be somewhat out of their hands. This relates to the lack of transparency problem in PA (DeNisi and Pritchard, 2006) in that it is not possible for the manager to communicate a clear answer to what performance is required to reach a certain rating. The rating will in the end depend on many other things than the perceived performance of the individual, e.g. the performance of others and the performance of the organisation and business area.

This cue extraction makes it meaningful to managers and employees to leave rating discussions out of the feedback dialogue. The lack of control of the rating at final appraisal would make a discussion between manager and employee about a preliminary rating meaningless. Not only would it be meaningless, it would also be risky for managers to set the expectations of the employees to a specific rating, because the rating distribution policies and the calibration process has the effect that managers will not be able to make any guarantees. However, a manager did in fact set expectations, not to the overall rating, but to the rating of single targets:

Staffs Manager: "I said that for the task on which she has not performed well, we have to agree that we do something about that now. It will get nasty if she sits with the same issue at final appraisal with her new manager." Interviewer: "So it was a 'pull yourself together' feedback?" Manager: "Yes. It was a warning that on this target it will not get above a 2 if there is nothing changed from now on. "

It is less risky to set expectations to a single target, where the manager has a lot more freedom to rate as she pleases at final appraisal, than to the overall rating, where the manager is bound by quotas, senior management challenge and calibration. Thus, it becomes meaningful to have a more tangible discussion about rating at the individual target level at midyear, than a discussion of a preliminary

overall rating, although it is the latter that is important in relation to reward, promotion, talent identification etc.

This way this manager finds a middle ground well in line with the middle position Production and Staffs were taking when their tasks at hand was mapped to the management by measurement vs. management by delegation regimes. PA is essentially a management by measurement regime, but there are so many other factors than the manager's evaluation of the individual's contribution that go into the final appraisal rating, so that it would be too risky to communicate a preliminary overall rating. As a middle way, preliminary ratings on single targets are less risky, because the manager here has much more discretion also at final appraisal.

A variant of the control story above is when managers leave rating discussions out of the midyear feedback dialogue with employees because managers feel that the dialogue will be too committing if it becomes tangible, i.e. specify exactly what actions will be required to reach a certain rating. I asked a Staffs manager if the feedback at midyear did not somehow bind the manager at final appraisal. The manager replied:

Of course if you have 10 targets and I have said for all of them that [the employee] has done really well then it is hard to get away with a 3 in the end... But for the medium group of 3s, where most will land, I do not raise them to the skies in my feedback. I use words like 'solid', and 'satisfactory'. I try to make my words match what I believe will be the end result when the year has been finalised.

Making 'promises' of certain ratings given certain performance activities, will minimise the degrees of freedom a manager will later have in the rating process and it will make it difficult for the manager to include new tasks and inputs that come to the agenda in the second half of the performance year. The Staffs manager above prefers to give indications through the provision of hints of articulation. This gives the employees an idea of where they stand, without compromising the manager's discretion in changing targets, priorities and ratings later. To this Staffs manager terminology is important. Providing a rating level indication to employees through cautious use of terminology while providing feedback is a meaningful way to balance the task of providing feedback on performance at Midyear against sustaining the ability to manoeuvre at Year End.

Ratings are irrelevant

For an employee who is used to being rated as a high performer, a downward change in rating from 4 to 3 can have significant impact on the employees' reward and career opportunities in years to come. So what happens when an employee who is used to getting high ratings in the 4s and 5s suddenly hits the wall and has trouble to meet business and behavioural expectations? Will the manager then be explicit about the employee's risk of losing the high rating? Will it not be very important to both manager and employee that a significant change in performance is required to maintain the high rating and that this is explicitly communicated by referring to the risk of a more mediocre rating? A Staffs manager with a usually high performing employee explained about such a situation in relation to the midyear meeting:

I am sure she had a different perception of this particular meeting. But I did not directly say that now she was more a 3 than a 4.

We see that it is the use of terminology and the detailed feedback on each target that does the job in the view of the manager. The manager is changing terminology and thereby tries to operate as a sensegiver to establish a congruence of cue extraction between herself and the employee from the feedback process, so that this can be meaningful to both. It is not necessary to destroy meaning by any unnecessary confrontation that would be the result of threatening the employee with a lower final rating. The employee will understand the implicit message that there is room for improvement and that final rating may be affected negatively if that room is not filled.

I also asked the formerly high performing Staffs employee if she felt that she was on the right path to getting a 4 like last year:

I guess it is shakier this time." Interviewer: "Was that explicitly stated?" Employee: "No". Interviewer: "Did you request that?" Employee: "No." Interviewer: "Why not?" Employee: "Because I sense that my customers are happy with what I do. That is the important thing to me.

The story of the employee illustrates how the reluctance to use preliminary ratings is not just due to managers' unwillingness, but also due to employees' inability to find value in preliminary ratings. This is relating to the problem of PA being manager centric (Coens and Jenkins, 2002). A preliminary manager rating is not valuable to this employee who is more interested in the feedback of the customer, which in this case is referred to as being positive. In fact, the employee later indicated, that

satisfying the customer and satisfying PA targets were not always parallel requirements:

PA measures projects on whether they reach the deadline and cost agreed. That is an ok starting point, but it is also a quite static view. You have to realise that the world is changing. The needs of the customer change throughout a project.

If there is a deviation between what is stated as targets in PA and what the employee experiences as the real requirements of the customer, the employee is in a dilemma; it is not possible to satisfy both. This employee prefers to satisfy the customer and therefore a preliminary rating from the manager is not interesting. In fact, a preliminary rating from the manager could be expected to be counterproductive because it would challenge the meaningful story the employee has created of why she is behaving the way she does, by questioning her prioritisation of customer requirements. We could argue that to protect the validity of this story, the employee defines preliminary ratings as being irrelevant.

Ratings are conflict generators

A fourth story that makes it meaningful to people to leave rating discussions out of the feedback dialogue is that managers and employees are shy of conflicts and simply want to keep the feedback meeting in as easy an atmosphere as possible.⁵⁸ The rating puts the feedback on edge which can sometimes become an emotionally demanding situation for both manager and employee. By avoiding this, both parties will have a nicer meeting where confrontation can be avoided. Of course, this would be a better explanation for why low-end performance is not discussed in terms of a preliminary rating. It is harder to see why high-end performers would not be pleased by the positive message that they were on the road to a 4 or a 5. But then again, it would be inconsistent and unpractical to maintain a process of giving preliminary ratings to high performers and not to the rest.

We know from the literature that it is a well established theory supported by multiple empirical studies that people, in general and in PA, will try to avoid conflicts (Bacal, 1999; Armstrong, 2008; Yariv, 2006). Tough feedback meetings are simply not nice to participate in and are often experienced as less meaningful than more comfortable

⁵⁸ See Yariv (2006) for an elaboration on why managers are reluctant to give negative feedback.

meetings. Most people, including managers, will prefer meetings where the conflict level is low. A Staffs manager told me:

It is of course much nicer if it is just praise. But life isn't like that, right. If you want to have people developing, there has to be some bumps on the road.

As this manager explains, having nice meetings is comfortable but that is not in itself the purpose of the PA process. If you want to move the employees you have to take some of the more difficult talks also. Taking the difficult conversation is part of this manager's role. But of course, one thing is to recognise that there is a need to also give the tough feedback, another thing is actually doing it. A Staffs manager who had a tougher than usual meeting with one of her employees explains the feeling this left her with:

I think she [the employee] felt the meeting was a bit... you know. And so did I. If it is a good meeting then both parties have a good meeting, right. But she hadn't reflected on these things before the meeting. It becomes more strained when you haven't reflected at all... I think she thought she should have lots of praise and then suddenly I sit there and give her critique.

This meeting was not a nice experience. The result is a loss of meaning for the employee, but also for the manager, who feels she gets neither acceptance nor constructive attention to her feedback. The manager blames the employee's lack of reflection for the uncomfortable situation but the result is still that the manager felt the meeting was unsatisfactory. The manager's frustration from the employee's lack of receptiveness to the feedback is a good example of how manager and employee can have different focus when they "chop moments" (Weick, 1995), or pick events, out of the on-going sequence of possible performance events and how this inconsistency leads to conflicting stories when they retrospectively describe 'what really happened' during the performance period up for review. The manager is confronted with a loss of meaning because the employee is not accepting the validity of the manager's story, and the manager establish meaning by concluding that the employee's lack of acceptance is caused by lack of reflection and self-insight.

Preliminary ratings are not part of the process

In Sales, where there was the best fit between the task at hand and the management by measurement doctrine, a Sales manager explained to me why she did not provide preliminary ratings at midyear:

As you can see here [in the PA form] there is no requirement to give a rating at midyear, so therefore I don't do that.

In other words, here the argument for not providing a preliminary rating at Midyear is purely practical; preliminary ratings are not part of the process design, so therefore they are not provided.

At the same time, the uncertainty about rating status is smaller in Sales than in the other workforces, because sales index realisation has such a significant weight in the composition of the overall rating, and because sales realisation is continuously monitored and rewarded financially with bonus on a quarterly basis. Managers and employees are updated on the sales figures monthly, so there is not much reason to discuss this again separately at the PA Midyear. However, when it comes to final appraisal, it is still the numbers that count:

Sales Manager: "I told her that at Midyear I am not focusing so much on the numbers but more on the personal drive and skill set. Then we will have to see if that is enough to drive it [sales] home the last few months [of the year]."

Preliminary ratings require unreasonably many resources

Lastly, yet an alternative story of why preliminary ratings are not provided at midyear review is about resource constraints. This story is less related to sensemaking of managers and employees and more to what we could call organisational sensemaking, or meaning creation at the organisational level. In a more general discussion of learnings from the PA implementation and rollout, one HR Staff explained to me that she sometimes doubted if the PA process with its link to rewards had become too resource demanding considering the value it brings to the organisation:

I ask myself if it is worth it. All the energy put into rating, calibration, runback and so forth.

The cost in terms of management resources spent on activities in relation to rating provision is significant. The informant's claim is relating to a general problem in PA described by Coens and Jenkins (2002) that PA requires significant resources, particularly from management. When this HR Staff, who had been involved in establishing the process in the first place, is concerned about the amount of resources spent in processes required to provide final ratings, then she probably would not be less concerned if a similar resource effort should be laid into provision

of preliminary ratings. Just imagine the cost of conducting a full calibration exercise twice a year instead of the current one calibration at final appraisal. No wonder that HR and Management would be concerned that this would require too many management resources.

On the other hand, proposing preliminary ratings without a calibration process would be dangerous, for exactly the same reasons that calibration is required at final appraisal. It would not make much sense to allow managers to communicate preliminary ratings without any calibration and central control, and then, in the end of the year, conduct calibration and control for the final rating. If employees experienced that preliminary ratings were generally overruled in the later final appraisal, they would have a hard time finding preliminary ratings meaningful. In fact, preliminary ratings in such a scenario would further undermine transparency rather than improve it.

No matter the reason for not using preliminary ratings at midyear, it can seem self-contradictory for the organisation to be a strong advocate for the positive effect of rating employees but at the same time not harvesting the full potential of it by using the employees' alleged appetite for higher ratings as a lever at midyear evaluation. But as we already have seen in many examples, meaning creation is not about telling stories about rational behaviour. Rather, it is about rationalising behaviour in hindsight.

Instead of calling it 'the challenge using rating' in midyear feedback, we could also have called it "*Listen carefully to what I (do not) say!*" Because the act of articulating performance feedback at midyear appears to be an art of providing performance feedback with great caution through a terminology of indications and hints and avoidance of referring to where the employee stands in terms of overall rating. The fact that none of the employees or managers I interviewed had discussed a preliminary rating at midyear is yet another paradox in PA; if ratings have consequences for employees' careers and reward and they are at the same time seen as an instrument that will have a positive effect on employees' motivation and

performance, it would be logical to bring up the rating also at midyear review, both for managers and employees. Nevertheless, I did not see this happen.⁵⁹

Discussion

Through six examples I have demonstrated how the same challenge to PA, the (lack of) use of ratings at midyear review, can be handled differently by different actors and according to different logics, in different contexts. Although providing different explanations to the same phenomenon, each of the five stories was meaningful on its own terms within their own contexts. As Weick (2009:141) puts it:

Sensemaking is not about truth and getting it right. Instead, it is about continued redrafting of an emerging story so that it becomes more comprehensive, incorporates more of the observed data, and is more resilient in the face of criticism.

In Table 8-1 below, the six stories are mapped to their relation to the two management regimes; management by measurement and management by delegation.

Workforce	Stories	Relation to management regime
R&D	<ul style="list-style-type: none"> Ratings are not positively affecting employee motivation and are therefore left out 	Principal argument. Relates to management by delegation logic.
Staffs & Production	<ul style="list-style-type: none"> Ratings are beyond the managers' control Ratings are irrelevant Ratings are conflict generators 	Both practical and principal arguments. Middle ground.
Sales	<ul style="list-style-type: none"> Preliminary ratings are not part of the process 	Practical argument. Stays within management by measurement logic.
Senior management & HR	<ul style="list-style-type: none"> Preliminary ratings require unreasonably many resource 	Practical argument. Stays within management by measurement logic.

Table 8-1: Mapping of six stories and their relation to management regime

Again, we see that the story from R&D, where the strongest fit to the management by delegation was found, contained a principal argument against using preliminary ratings, and thus twists PA away from a management by measurement philosophy. In contrast, Sales, where there was a better fit to management by measurement, has

⁵⁹ This is in line with my own experience. In my years as a people manager in PharmaComp I have only very few times experienced that employees have requested a tangible preliminary rating at midyear. I have never taken initiative to provide such rating and I have never requested the same from my own manager.

a purely practical argument against the use of preliminary ratings, and thus stays within the management by measurement regime. Unsurprisingly, the story from HR, the owners of the PA process, is also a purely practical argument against preliminary ratings.

Again, the picture is more mixed in Production and Staffs, with both practical but also some principal arguments against preliminary ratings represented in the three stories from these workforces. Also this is less surprising, since Production and Staffs were categorised in a middle position between the two management regimes.

8.4 The Challenge of Carelessness

One challenge to meaning schemes in midyear review that came up multiple times during my interviews with managers was the perception that employees were sometimes unprepared for, or careless in, midyear review. If a manager perceives PA to be important and the midyear review to be a key performance event with significance for the career and reward of the employee, it will expectedly be a challenge to meaning schemes if the manager perceives the employee as careless. To one Staffs manager, perceived employee carelessness was caused by minimum preparatory efforts of the employee before the midyear review meeting:

...that's another thing; her reflections on the targets. She only writes one line: On plan. When I get it, I write that maybe it is on plan, but you missed this and that.

Writing "On plan" to everything does not indicate that the employee has had any meaningful reflection on her own performance. Minimum or even complete lack of employee preparation before PA meetings was something that several of the managers I interviewed experienced. On the question if it is not annoying that employees do not prioritise spending an hour on being well prepared for the midyear meeting, a Staffs manager explained:

You could say so. If I was not prepared, she would hopefully judge me on that. It is also a precondition for my preparation. It takes longer time for me to prepare when she has not documented anything. When people are thorough in their documentation it is a big help to my thought process. Of course, I check what they write, but that is something else. When it only says 'Okay', then I have to do a lot more footwork myself.

According to the PA process as it has been defined by Corporate HR, there is no doubt that it is formally the responsibility of the employee to take the first step and

document a self-evaluation on each target at midyear. If the employee just writes 'on plan' for all targets, she can hardly claim to have taken that task seriously. To the manager quoted above, it is clearly annoying when employees have not done their part of the meeting preparation. It makes the manager's own preparation more difficult and time consuming. Managers then have to document the feedback from scratch instead of just review and comment on the self-evaluation of the employee. This takes time and requires that managers have either documented feedback on an on-going basis, or that managers can remember relevant examples for all targets and employees, which can be quite challenging for a manager with 10+ direct reports, as many managers have.

Considering the unsatisfactory outcome for the manager of the lack of employee preparation, the reader may be surprised to learn that managers in the cases I investigated typically accepted the lack of employee preparation in the sense that they did the extra work required to compensate for employee carelessness. If employee carelessness is challenging managers' meaning schemes, how can it be meaningful to managers to acquiesce and do the work that the employee should rightly have done? Why is the manager not just sending back the PA to the employee with the message; *'Sorry, but this is not acceptable. The meeting is postponed until you have lived up to your responsibility and seriously prepared a self-evaluation'*? Why is it meaningful for the manager to conduct the midyear review although the employee has obviously failed her responsibility to contribute constructively to the meeting by preparing her self-evaluation thoroughly?

In a sensemaking perspective the answer to these questions is that managers' acquiescence is meaningful because they extract cues from their compensating actions that form the basis of stories that establish meaning. These are stories about what the midyear review is 'really about'. Let us look at three examples from different workforces, of such stories that are meaningful to the manager. Again, as we have previously seen, the stories are based on similar events (perceived employee carelessness) but on different local adaption in PA and different extraction of cues from the events. The stories relate to different problems in PA and meaning is established along quite different paths.

Acquiescence due to governance fulfilment

The first story is about governance fulfilment in Staffs:

Staffs employee: "Somehow you could say that the PA process is in itself about governance fulfilling. The managers are all measured on whether they have fulfilled the process with their employees."

This employee directs our attention to the fact that managers are responsible for the timely fulfilment of deadlines in the PA process within their area. It is part of managers' core administrative responsibility to make sure that all employees are being managed in accordance with the PA process and this fulfilment is measured through the share of finalised PA process steps. For many managers, the PA process fulfilment within their area is even formulated as a PA target in itself, which is what the employee refers to in the quote above. However, for managers who have many employees, realising the PA deadlines is not trivial:

Staffs manager: "Deadlines [in PA] are quite squeezed."

It is a requirement for managers to make sure that the PA process is kept within the official deadlines. Since it is very common to have the meetings close to deadline, there simply is not enough calendar time for the manager to send back the self-evaluation for an extra iteration. It may be irritating to the manager to have the meeting under the conditions caused by lack of employee preparations, but it would be less meaningful to make an example and be tough on the employee at the cost of missing the manager's own target of 100% process compliance. In this interpretation, it is more meaningful to the manager to do the necessary extra work and thus fulfil her own process compliance targets. This relates to at least two problems in PA handled in the literature; Deadrick and Gardner (1999) claimed that focus in PA is often more on formality (process fulfilment) than on utility (performance improvement). And Piercy (1994) claimed that PA schemes were more focused on bureaucracy and administrative neatness than on handling real strategic issues:

Staffs manager: "It is not like people think that the PA process is really great, you know. It is something they do in the last minute. I had asked them to provide their input two days before the meeting. Maybe one third of them did that. The rest did it the night before... It is a kind of indifference to the process... They probably don't see that there is something in it for them that make it sensible to prioritise it."

If governance fulfilment is seen as the main purpose of the midyear review, spending energy on preparing detailed discussions of particular performance events is less

meaningful to employees, simply because it is not required for governance fulfilment. The mixed fit between the task at hand in Staffs and the management by measurement regime of PA also in this case is related to a mixed pattern of local adaptation. Employees are careless because they do not see much benefit from spending a lot of time on the process. They are doing the bare minimum preparation required for the process to continue, and let it be up to the manager to do what is necessary to be able to reach the process compliance target of reaching deadline.

Acquiescence to avoid conflict

The second story is about securing organisational performance in R&D. This story also concerns how some managers have mixed experiences with enforcing employee engagement in the PA process. However, in this example, the employee rejects improving her carelessly prepared self assessment although asked to do so by the manager. This case is a variant of the above example of governance fulfilment, but it is a more extreme case. Here, the manager is actually trying to enforce a quality process, meets resistance, and acquiesces.

If a manager decides to go for a tough approach in such a case it would be problematic because it would initiate an unproductive conflict with the employee without adding important value. It is more meaningful to the manager to avoid the conflict. It is more meaningful to accept a low quality, time consuming PA process which has the advantage of being compliant with deadlines and low on conflict. This type of cue extraction must be expected to be particularly likely if the manager had had similar experiences in the past, where confrontation had not been fruitful. An R&D manager explains about one particular midyear session in the 2011 cycle:

[There is] particularly one where I think that is simply because she does not accept the system. [She writes] as few comments as possible. And that makes it really difficult. I even sent it back and said that she had to put some more words on it because otherwise it would be useless. But nothing happened. Then I think it is difficult.

The particular employee mentioned in the quote is doing a good job and the manager is happy to have her in the department. It is also a very experienced employee. But the employee has literally abandoned PA. It appears as if the employee could not care less and she simply refrains from any active participation in PA. This is a clear challenge to the manager's role within PA and to the manager's maintenance of

meaning from being part of the PA process. How does a manager make anything meaningful out of that situation?

The manager's establishment of meaning is dependent on the cue extraction from the situation, on her perception of her own role in relation to the employee and the PA process and on her enactment of environmental resources. The manager could react by giving the employee a warning or even take tougher measures. And she could probably through that force the employee to write just a little more in PA so that it would reach the lower limit of what would be acceptable.

But consider that this employee is actually doing a good job. She just does not see the point in the PA process. What would the manager get out of embarking on the tougher approach? Would she get a more motivated employee? Not likely. Would she get better performance? Not likely. The manager modestly tries to get the employee to do a bit more preparation but the attempt is unsuccessful. This puts the manager in a difficult situation. The next step is either acquiescence or conflict escalation. In this case the manager acquiesces. She acquiesces and retrospectively establishes a story that shows how acquiescence is sensible if the objective for the manager is to get the best performance out of the employee. In a sensemaking perspective, it becomes meaningful to the manager to handle a difficult situation by, so to speak, not handling it.

In the above example, the manager took an extra step by sending the PA back to the employee, whereas in the previous example from Staffs, the manager did not send the PA back to the employee for re-work, and thus avoided the risk of conflict escalation. Both managers end up by avoiding conflict escalation by acquiescing, although at different steps on the conflict escalation staircase. At first, the two managers extract different cues. The Staffs manager acts acquiescently from start and retrospectively creates a meaningful story about how doing the work for the employee is more sensible than other alternatives.

The R&D manager at first escalates the conflict by sending back the PA for rework. This constitutes a meaningful story about how employee preparation is a requirement of the process and therefore needs to be enforced. However, the employee reaction is to ignore the message from the manager and thereby she undermines the meaning of the manager's story. To establish new meaning, the manager has to

retrospectively create a new meaningful story about the action that has already taken place, namely that the manager has now acquiesced and de facto accepted that the employee will not participate actively. This new story is about how acquiescence is meaningful because the employee is a good performer who just happens to have abandoned PA, and that the result of further conflict escalation would be detrimental to the manager's purpose of securing organisational performance.

This story relates to a very fundamental problem in PA as it was depicted by Kohn (1999); that it is still unproven that PA schemes increase organisational performance. In fact, Kohn goes further and claim that PA schemes have an outright deteriorating effect on organisational performance. The manager in the example senses that a performance decrease may be the outcome of insistence on employee engagement in the process. And because organisational performance is the key priority to this manager, it is meaningful for the manager to acquiesce and she accepts what is in reality a partial abandonment of the PA scheme for this single employee.

Further, we could argue that it is particularly in R&D we should expect to find such examples of employees' outright rejection of PA. The task at hand in R&D has the worst fit to the management by measurement regime that PA is affiliated with. And R&D has the highest concentration of highly skilled expert employees, often equalling or surpassing the organisational prestige of their managers. Particularly in R&D, we should therefore expect to find employees who not only see PA as meaningless, but who also have the organisational power base to challenge their managers' authority to enforce PA upon them, and to sometimes win that conflict, as was the case in the above example.

Acquiescence as creative process redesign

The third manager story is about creative process redesign in Sales. One of the Sales managers I interviewed had taken the ultimate consequence of employees' reluctance to prepare for midyear review. She had simply taken systematic and full responsibility for what should, according to the PA guidelines, be the responsibility of the employee:

I have prepared the PA before the meeting. None in my team are good at driving their own PA. I have almost stopped asking them to forward their input before the meeting because they don't get it done. I don't know if I have spoiled them... They have become so used to me finding all the numbers and writing it down.

This third story is related to the first two stories above and to the same problems in PA. However, here we are no longer talking about a single employee who gets special treatment for the sake of maintaining organisational performance. Here, the manager has in effect changed the PA process as it is laid out by corporate HR, by leaving out the process step that concerned employee input. Instead, the manager herself systematically does the meeting preparation work on behalf of the employees. The manager does this simply because she fears that if she did not, the employees would not get it done. We have already seen previously how in Sales PA is not connected to reward for Sales reps and how the instalment of targets were almost entirely top-down controlled. This may help to explain why Sales employees, at least in this department, are particularly reluctant to prepare for PA meetings. But from a governance perspective it would still be unacceptable to the manager to have a whole department missing its PA deadlines, so it is more meaningful to the manager to do the work by herself.

This example may at first seem surprising when we consider that Sales is the workforce where we found that the task at hand had the best fit to a management by measurement regime. If the fit is good, why does the manager not prioritise enforcing the corporate process by demanding employee engagement in self-evaluation at Midyear? In a previous example from the target setting process we saw how a Sales employee was completely alienated from her PA because she felt targets were distributed by management without her involvement. A Sales employee elaborates on this:

I cannot see the need for this [PA]... I guess it is something management needs to have to feel they are in control. Although I am not sure what they are using it for. I haven't seen the light in this.

In Sales, the definition of targets is strictly top-down and the measurement of target realisation is quantified and based on numbers drawn from another system. And further, bonus is allocated purely in accordance with sales realisation, where numbers are beyond discussion. In such a regime, there is little incentive for employees to invest time in adding comments in PA as part of a self-appraisal. What difference should that make?

However, for the manager, PA process compliance is still required. This explains why the manager in this illustrative, although maybe also extreme, example, is systematically doing the work for the employees in her department.

Discussion

In the three examples above we have dealt with the concept of preparing for midyear review. However, ideally the preparation for midyear review should not only be about spending an hour on providing input right before the meeting takes place. Preparing effectively for midyear review would require a more continuous effort. Thus, one particularly important element in preparing for the midyear review meeting would be to take notes of important performance events as they unfold. The reason why this is necessary has to do with the recency problem in PA (Coens and Jenkins, 2002). According to Coens and Jenkins this bias could also be called the 'what have you done for me lately' syndrome, because it concerns how feedback, unless very well prepared, will primarily concern what has happened most recently because that is what is most present in the memories of actors. Coens and Jenkins see this as one of the key problems within PA in general, because PA schemes install a calendar controlled paradigm on important feedback from managers to employees. There may be other occasions for feedback, but according to Coens and Jenkins, PA schemes make a differentiation in importance of normal (not committing) feedback and PA (committing) feedback that devalues the former and emphasises the latter.

To support on-going note taking - and to be able to compare managers' and employees' selection of performance events in their note taking, I offered a 'diary' template at the first interview with all managers and employees.⁶⁰ However, only two employees and one manager used the diary. And the manager only did it through the first half of the interview cycle, i.e. until midyear review. To the question whether the notes was a help to the manager during the preparation for the midyear review meeting with the employee, this Staffs manager replied:

Yes, I could remember the examples much better because I could run through them right before the meeting... I think she [the employee] maybe felt we got into too much detail with some of the things she considers to be minor. Or maybe because those are some of the things she would prefer we did not take up.

⁶⁰ See manager and employee diary templates in the appendix

The ability to draw on tangible examples in some detail was heightened by the manager's diary notes. However, the employee did not necessarily appreciate this higher level of detail. And this manager did not continue with the diary in the second half of the cycle. Like the other managers I interviewed, she gave as reason that she had been too busy with other things and that daily note taking simply is not high enough on the agenda to get it done on a regular basis. Or, to phrase it differently; the manager's diary helped the manager to remember performance events she would otherwise not have remembered. But still, it was less meaningful to continue doing this because the examples were not valued by the employee or they were in other ways not important enough to the process to justify the required manager effort of taking on-going notes.

Apparently, and in contrast to what one might have assumed, the PA feedback and midyear review does not require detailed documentation of performance events to be meaningful. In fact, as the above example shows, it is sometimes easier to create meaning from the process without those examples, maybe because maintaining the feedback dialogue on a less tangible level opens opportunities for softer interpretations and less conflict. In a sensemaking perspective this is not surprising, because we would expect plausibility to be of higher importance than accuracy anyways. But in a traditional PA optimisation perspective it is an important observation because it moderates the importance of many of the measurement and bias problems in PA. If accuracy is not necessary, lack of accuracy is less of a problem. And when accuracy is not important, the value of preparing thoroughly for PA meetings by taking detailed notes on an on-going basis becomes less meaningful.

Of course, this again raises the question of what the whole purpose of the feedback part of the process is. Is it to improve organisational performance or is it to provide tough, honest manager feedback to individuals based on tangible examples? That these two purposes are not always congruent and maybe sometimes even conflicting, should be clear from the above examples. Through sensemaking, managers create stories that make their translation of the situation and their past action meaningful. And in fact, as examples showed, it is sometimes easier to create meaning from the PA feedback process without detailed examples.

But if note taking was not a necessary ingredient for managers to create meaningful stories of what is going on at midyear review, why were also employees reluctant to use the diary? Would it not be to the advantage of the employee that all the good performance through the past half year was captured in the midyear review? One Staffs employee explained:

I have to admit that I didn't even look at [the diary]. I didn't have a particular need for just another tool. And it is also due to time pressure. I have had more than enough to do. So a new tool which is not urgent doesn't have high priority against daily work tasks. And after 2-3 weeks, you forget all about it.

In this quote we see how the employee reasons much along the same lines as the manager above, who also thought that preparing for feedback through note taking is a low priority task that will eventually get squeezed out by other, higher prioritised tasks. Another employee had also forgotten about the diary:

R&D employee: "I completely forgot about it. We are so busy in our daily work so we don't really think much about PA"

Being busy is one thing, but it is the second part of this employee's statement that is the more interesting in terms of understanding why taking notes aimed to support the PA process is one of those things that is hard to get done. If PA in general is not something this employee has much relation to in daily work, then the PA scheme has become detached from the daily work life of the employee and has as such lost importance for the employee. Things that are not important are easy to forget.

A third employee who did not use the diary was even more direct about this interpretation:

Staffs employee: "I don't find it important, not at all. I work the same way no matter if there is such a document [PA] or not. But it is important to my manager. That I was told... there have been occasions where I could have used the diary; where my manager has praised me. But I didn't do it. You are always in the middle of something. You run from meeting to meeting. I have been busy."

Again, we see reference to being too busy to prioritise conducting on-going documentation of performance events as preparation for PA meetings. But the employee also offers an explanation to why preparing for PA meetings get lower priority than other tasks in a busy schedule; because the employee simply does not find the process itself particularly important. Here, we are back to the compliance fulfilment story we also saw above; if you find that PA is not doing any good to you, it becomes meaningful to reduce the resource efforts you put into the process to the

absolute minimum, and that is to show up at meetings and sign off meeting conclusions.

Above we have seen how managers I interviewed had experiences with employee carelessness in preparation of PA meetings. Although we cannot draw any statistical conclusions based on a qualitative study like the present, it would not be an unreasonable hypothesis that the managers I interviewed are not unique in this respect, although future studies are required to verify the generality of this. In the cases analysed above we saw how some employees were disengaged in PA because they did not find PA important or because they were too busy to spend time on preparing PA meetings. We also saw that when faced with employee carelessness, the managers I interviewed typically acquiesced and accepted a lower quality process and sometimes they also accepted to take over some of the PA process responsibilities that the employees should ideally carry.

That employees can be careless and reluctant to spend even minimal efforts on preparing for PA meetings is an important observation because employee alienation and disengagement is threatening the quality of the process. As PA schemes get more connected to other HR processes like reward, promotion, talent identification, dismissal etc., it becomes a paradox that employees feel disengaged and alienated from the process to the degree that they do not find it worthwhile to be well prepared for PA meetings. This study is only a first step in understanding the nature of this paradox and it would be greatly beneficial to see future studies deepen our understanding of this phenomenon.

8.5 The Challenge of Disruption

From the above analysis, it should be clear that although sensemaking processes are essentially social and happen in a complex interplay between actors (here, particularly managers and employees), individuals do play a crucial part. If the dyadic relation between a manager and an employee is of key importance to the meaning creation dynamics in PA, we should be interested in looking at what happens if one of these individuals is suddenly substituted by another in the middle of a PA cycle. This is what I call the challenge of disruption.

As I explained in the methodology chapter, four employees experienced a manager change during the course of the study. For all, the change happened shortly before or after the midyear review. This relational disruption in the middle of the process is a challenge to continuity in mutual agreements and expectations set at the target setting stage. With a new manager, targets and performance may be translated differently. This generates changing conditions for meaning creation for the employee and the manager. Across the workforces, two distinctly different patterns of meaning creation were observed; one was about active change management where the manager operates as a mediator of PA information to her successor. The other was about downplaying the change through taking a manager role as a PA accountant where no mediation is required beyond the PA targets themselves.

Disruption and the manager as mediator

I asked a Staffs manager if she knew that she was on her way to a new position when she had the midyear meeting with the employee:

Yes, and it was a bit funny. At that time I knew I was on my way. So I put a bit more effort into it than I normally would have done at midyear. Simply to document it all so that my successor will have something regarding these people when they reach year-end.

The manager is on her way to a new position. Therefore, she will not be the one responsible for making a final rating of the employee when year-end is reached. Hence, the manager cannot count on her own memory about performance events and agreements made at target setting, because she will be somewhere else at that time. For hand-over purposes it becomes meaningful to make a more thorough documentation of status at midyear than the manager would normally do. This way, the manager proactively relates to the problem in PA raised by Raymark et al. (1999) and others, that PA evaluations are characterised by incomplete information. This is a problem, which would be exacerbated by a disruption in management relations.

On the other hand, the Staffs employee did not perceive the midyear meeting preparation to be any different than usual:

I am trying to be fairly detailed in my own comments anyways. So I don't think there's any need to write more.

I asked a Production manager also on her way to a new position and who had a similar experience, why the upcoming organisational change meant that this midyear was more thorough than normal. Could it not be that thorough always?

Yes it could. But traditionally, it is more about a status on the targets... So this time it was a bit more thorough and I thought that was really good.

There are at least two interesting pieces of information in this quote. First, the Production manager tells that a normal midyear is "more about the status on the targets". We could call this the manager role of *accountant* in the feedback process, because the manager in that scenario is controlling the status on targets similar to the exercise accountants conduct when they audit the books of a company. But in this example, because it was the last session between the manager and the employee, and because the documented outcome should serve as input for the new manager, it was different. The discussion was better prepared and the themes discussed were broader than the usual strict target follow-up at Midyear. We could call this the manager role of *mediator*, because the manager here is concerned not only with following up on status on targets, but also on mediating detailed information to her successor through the PA system so that the transition to the new manager can be fair and smooth.

The second interesting piece of information the above quote gives us is that the changed context and conduct of the midyear meeting is actually perceived to be a good experience by the manager. The more thorough preparation and documentation is perceived as being meaningful because it facilitates a better founded discussion of the performance of the employee and of the transition to the new manager.

Disruption and the manager as accountant

In Sales, the manager perception of the effect of an organisational change on the PA midyear meeting was different than the above. On the question whether the manager had a discussion with the new manager where they went through all the PAs of the employees affected by the change, the Sales manager replied:

"I have not had midyear with X. Her new manager had that... You know, I did make some handover. But it simply became overwhelming for us if we should go double-up on midyear participation." Interviewer: "Is that possible because your targets are very 'hard' in Sales?" Sales Manager: "Yes, exactly."

This story is different from the previous examples, where the old manager was conducting the midyear review. In this case, the manager change happened in July, but it was the new manager who conducted the midyear review with the employee in August. Because of workload, the old manager could not participate in the meeting. The workload argument is relating to the problem that PA requires significant resources, particularly from management (Coens and Jenkins, 2002). In this case, the problem is handled by letting the new manager take the midyear meetings by herself although she will not be able to base the feedback on any personal knowledge of the employees' targets or performance.

The manager perceives it as a reasonable prioritisation because targets are so clear-cut in Sales. The logic in this argument is that if you have a Sales target to meet index 100, and any manager can pull your numbers out of an IT system, any manager could also provide the feedback on whether or not you are on target. The Sales employee appeared to share this view:

The targets are fixed, you know. They come from above - that's how it is. So we just relate to that as it is... I had put in my comments but basically it is just about looking up the numbers.

We may say that the situation in Sales is an extreme case of the accountant role of a manager mentioned above. The advantage of favouring quantifiable targets is that a disruption in manager continuity becomes less of a problem. Just like a company can change its accountant without changing the valuation of its assets, the new manager can look up the numbers and provide feedback to the employee at midyear.

However, does the manager role identity as an accountant not leave all the surrounding discussions about what went good or bad, the tangible feedback on behaviour in performance events, the discussion of influence of external factors etc. out of the equation so that midyear risk becoming a strict target status follow up without much opportunity to reflect on reasons and thus opportunities to improve? It could be in some cases, but in this case that conclusion may be too stretched. At least, that risk is moderated by the fact that the new manager in this case was also an experienced manager within Sales who shared an office with the old manager and who knew her new employees quite well already, although not from a manager-employee relationship.

Discussion

In Figure 8-2 below, the two different approaches are depicted in the meaning creation feedback model. The model illustrates how the manager as an accountant is a consequence of downplaying the disruptive element by enacting the perception that PA targets are self-explanatory so that further elaboration is not required, to be able to pass on the PA responsibility to a new manager. In this case, the disruption does not constitute a challenge to meaning schemes. However, with this approach there will be a loss of richness in the feedback the new manager will be able provide because she can only base it on the scarce information held in the PA targets.

In the manager as a mediator example we see in the model the enacted assumption is that there is more to PA than what is stated in the targets. Therefore the disruption challenges meaning schemes because there is a fear that the manager change will bring a loss of this undocumented richness. The manager mitigates this risk through mediating additional information, but by doing this she at the same time undermines the perception of PA as being objective and transparent in the sense that there is a direct link between target realisation and performance feedback.

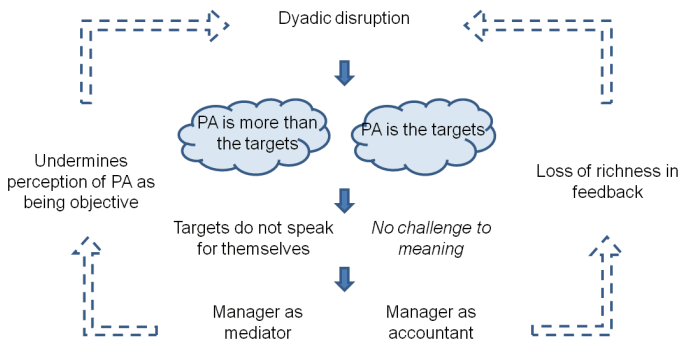


Figure 8-2: Mediator or accountant - different handling of dyadic disruption

The example from Sales with the manager as an accountant fits well with the categorisation of the task at hand in Sales as being the workforce with the best fit to the management by measurement scheme. When repetitiveness is high and complexity, specialisation and task newness is low, there is a good foundation for establishing a measuring regime to account for individual performance. When PA is only about the targets, and performance on the targets can be directly pulled from an

IT system without needing any translation by the manager, the manager becomes an accountant who 'audits the books' of the employee. In such cases management disruption is not an issue.

However, when performance is more than what is in the PA targets, the manager requires knowledge of this 'more' to be able to provide feedback. In case of management disruption, it will become an issue if this 'more' is not handed over to the receiving manager. We have already seen how in Staffs and Production there was less of a fit to the management by measurement regime. We have also seen how un-quantifiable elements play a larger role in these workforces than in Sales, when it comes to evaluation of performance. Therefore, based on the differences in the task at hand and the fit to the two contrasting management regimes, it is no surprise that managers in Staffs and Production are more prone to experience management disruption as a challenge that they need to actively handle by taking the role as mediators of knowledge to the receiving manager.

Let us finalise the analysis of the meaning scheme challenge of disruption in midyear review by returning to the example of a Staffs manager who made an extra effort of documenting the midyear review in PA, to support a better handover to the new manager. Interestingly, when it came to the actual handover, this was done more informally and without looking in PA:

Staffs Manager: "We have discussed the employees regarding their tasks, personality and possible performance issues. But we did not look into PA."

The more thorough work on preparing and documenting midyear review was in this case not activated in the actual handover process. PA was not even opened as part of the handover. Of course, the more thorough comments will be available to the new manager when she later will be taking the employee through the final appraisal, so the effort is not wasted. But the details of the status on each PA target are not necessary to facilitate the manager handover, which is a more informal discussion of "tasks, personality and possible performance issues". Here, we again sense the apparent disconnect between the detailed targets defined in PA and the overall manager evaluation of employee performance status. The latter is not the result of the sum of the former. It is possible to go directly to the whole without first going through its parts. Or, to formulate it in sensemaking terms by rephrasing a previous reference to Garfinkel's (1967) study about how jurors work; it appears as if

managers first establish their overall impression of an employee before they retrospectively make that impression rationally meaningful by selectively referring to the elements (PA targets and other factors not described in PA) that support this story.

Murphy (2008) suggested that it is a problem in PA systems (and in PA research) that they often ignore or downplay within-person variability in performance. In other words, ratings are static measures for dynamic criteria. But what about variability in actors? This is not a theme which has received a lot of attention in the research. Most contributions deal with PA as if actors were the same over time although today's job market in many industries is characterised by high external and internal mobility and turnover. In many organisations it is common practice that managers change position or that units are reorganised every 2-4 years. In such organisations, manager-employee PA relationship disruption will be more the rule than the exception. And since we have already seen examples of how managers and employees are not documenting everything of significance in the PA system, it becomes important to understand what happens to all the tacit knowledge and the relationship effects, when an employee gets a new manager. No matter if it is in the middle of a PA cycle or at the beginning of a new cycle. I would argue that this is a problem in PA that would be worth investing further studies into.

9 Local Adaption and Meaning Creation in Final Appraisal

9.1 Introduction

The manager's key responsibility in the final appraisal process is to provide and communicate an appraisal rating on the employee's performance on each PA target and to provide and communicate an overall final appraisal rating.

In the literature review I found that the final evaluation is the part of the PA process that has attracted most attention in the literature on PA. It is not clear if this is so because the rating phase is the most problematic part of the process, or if the overweight of problems identified in this part of the process is caused by other factors. Nevertheless, the literature review uncovered a multitude of challenges to the rating phase that would all be elements in the local adaption and meaning creation in this phase of the process.

Below, the analysis will be divided into a number of themes that I believe exemplify local adaption and meaning creation processes in final appraisal across workforces particularly well. First, I will analyse challenges to evaluating business performance. Second, I will analyse challenges to evaluating and including behavioural performance in the provision of a final appraisal rating. Third, I will look at different challenges to managers and employees in different rating distribution policy settings. Fourth, I will analyse managers' and employees' challenges in the rating calibration process.

When it comes to providing a final rating, the PA process in PharmaComp is very top-down organised. In fact, in three of the four areas I investigated, the final ratings were provided exclusively by management and several weeks or even months before the employees were required to provide their self-evaluation as input to the appraisal feedback meeting with the manager. In these cases, the self evaluation therefore could not have any effect on the final rating decision.

In the analysis below there will be more focus on managers' meaning scheme challenges in final appraisal than on the meaning scheme challenges to employees. The reason for this is that so much of the process around the provision of a rating is a manager-centric and quite exclusive process. Hence, to understand the actual

appraisal process and its challenges to meaning creation in practice means looking at meaning scheme challenges of managers', simply because until the final appraisal meeting the employee is normally only sporadically involved in the process.

9.2 The Challenge of Evaluating Business Performance

The evaluation of business performance is at the heart of any PA scheme. After all, one of the main purposes of installing PA in the first place is to improve business performance of the whole (the organisation) by improving the business performance of its parts (the employees). Below, three different perspectives on challenges in business performance evaluation are dealt with. First, the textbook story of causality between target/evaluation criteria and final appraisal rating is discussed. Second, the antithesis to this view, the non-causal relationship between target/evaluation criteria and final appraisal rating is discussed based on examples from R&D and Staffs. Third, an example from Sales is analysed to see how retrospective change of the rules of the evaluation game can be a meaningful response when established rules challenge meaning schemes.

A story of causality

If the investigation of managers' and employees' reflection on PA can be compared to peeling an onion, then the outer layer of the story of providing a final rating in my investigation was often the story of causality. One important element in this pattern is the initial reflections of managers' on their own evaluation of employees' realisation of targets set at the beginning of each cycle. For example, I asked a manager in Staffs, what rating an employee could expect to get, if ten goals had been established and the employee had realised them all, and she firmly replied "From ME to O."

However, the literature has established that it is very difficult to set targets that are not either too ambitious or too easy (DeNisi and Pritchard, 2006). And in the previous analysis we have seen how the managers I followed in PharmaComp struggled to set the right targets and found it very difficult to calibrate targets even across their own employees, let alone across larger populations. Thus, we have good reason to assume that employees will have varying levels of ambition built into their targets. I therefore asked the same manager what would happen if an employee's targets were

very unambitious - could she then get an AE rating although meeting all targets? The manager replied in short: "No."

This manager is very clear in expressing a sensible story around final appraisal in PA; if you have realised your targets, you will at least be ME. And we can infer that if the targets are unambitious, then you can be ME with a very modest performance. In that case it would be the responsibility of the manager to make sure the targets are more ambitious next time.

Another manager in Staffs also emphasised the precedence of business performance in the constitution of a rating:

You don't get EE for a large effort. You get it for your deliverables. Luckily, there is often correspondence between these two, but not always. Some have to spend more time on delivering the same.

This story of causality between targets, individual target evaluation and overall final appraisal does seem very intuitive. I have above quoted Staffs managers, but I could as well have quoted managers from the three other workforces. At first, most managers I interviewed reiterated the story of causality when accounting for how final appraisal ratings were provided. At first, there was not any clear pattern in relation to differences in the task at hand in the workforces, and the initial response when asking managers about the relation between business targets and evaluation.

Either, this could be seen as a sign of managers providing an 'auto-reply' based on their understanding of PA as a management by measurement tool. Or it could be taken at face value and translated as if the causal story is in fact the 'real' story. PA evaluations are based on strict bottom-up evaluations of each single performance target, so that the overall evaluation is the weighted average of its parts:

Sales employee: "Yes, that is how it ends up. Though, some things weigh more, it is all accounted for in percentages. There is not much human judgement in this. It is hardcore business."

However, we have already previously seen how the behaviour evaluation was promoted in some workforces to moderate business target evaluation in the constitution of final ratings. In the case of introducing a rating matrix in Production and Sales, behaviour was a strong moderator of the overall evaluation of employees so that a business performance of ME could be translated into an overall

performance of AE if behaviour is unacceptable, even though this was not reflected in corresponding behavioural targets.

We also know that in modern organisations employees are often dependent on others so that they can make brilliant individual contributions and still not reach their targets, due to others' failure, external forces or sheer coincidence. Or the opposite way around; they could be lazy and incompetent but still achieve their targets because others compensated for their inferior contributions. So do such external factors, although undocumented in PA, not mean anything when business results are evaluated by managers?

The story of causality creates meaning around the rating process by maintaining a strong, causal link between defined targets, realised individual outcome ("*deliverables*") and rating. You may work as hard as you want, but if you do not realise the targets you will not get an ME. It is a simple story about causality, transparency and clarity of targets and evaluation criteria.

Sensemaking is about creating such simplified stories that make actions meaningful in retrospect. So we should not be surprised to find such simple stories based on causality and rationality. Instead, we should challenge managers' reflections and ask them how the story of a strong link between business performance and rating matches the strong moderating effect of behaviour that we know is being implemented in Sales and Production. And we should also ask if interdependence on the actions of other actors or to external factors may not after all affect the evaluation of an employee. Thereby we may provoke the managers to reflect and establish revised stories that reinstall meaning by incorporating such factors into new retrospective meaning creation of past events. From that we should go on challenging the managers and employees on their reflections on other possible sources of moderation, e.g. bias, calibration results etc. This is what the qualitative interview approach allowed me to do in PharmaComp, and by this approach alternative stories were introduced by interviewees.

We will look closer at the most important of these elements below and see how actors, when the causality story is challenged by their experiences in PA, create new meaningful stories.

A story of non-causality

Let us look at another example of a manager reflecting on the link between targets, performance and evaluation. I asked a Staffs manager if she knew the expression "you get what you measure", and if she thought that holds true for how the employee's incentive to innovate and perform outside the boundaries of her PA targets would be affected by the focus on documented targets installed by the PA process. The Staffs manager answered:

I think that is true. But I also think it is about making people feel that they are not just evaluated on their targets but also on what they do outside their targets.

Notice that this manager is the same who in one of the quotes above subscribed to a story about a strong link between targets, evaluation and rating. Now, challenged by the possible negative effect of such a rigid target/rating causality regime on innovation and proactivity beyond the targets, the manager moderates the story. Now, it is important that employees understand that they are not just evaluated on their targets but also on what they do outside their targets.

I also asked employees about how they saw the connection between innovation and PA:

R&D Employee: "I am thinking about it as two opposites. PA - that is where you are static and where you put everything into neat boxes. And then there really is no room for all the ideas you get the day after... And then again, you are not so motivated by the PA targets, because they have to be realistic. It is just a fact that it is easier to be innovative when you have very ambitious targets and have to think along unrealistic paths."

So, there is awareness that a too strict causality based relationship between target realisation and final appraisal rating will have negative side effects and maybe even be a barrier to innovation. Although the story of causality is intuitively rational and nicely congruent with the management by measurement doctrine behind PA, reflections on downsides of PA challenges meaning schemes of what is required to secure high performance and employee motivation, and led to moderated stories that were required to create new meaning from this dilemma.

This thesis/anti-thesis composition through reflection nicely demonstrates Weick's point about sensemaking being an on-going process where actors continue to establish and maintain stories based on what is most meaningful to themselves. If I was looking for the objective truth I could have asked which is true; is there a strict

causal relationship between business target realisation and rating, or is there not. But in a sensemaking perspective that question may not be relevant for anything but to provoke new reflection and possibly new stories. The point is, that there is no way we in a sensemaking perspective can say which of the two answers is most true. In fact, both answers may be true in the sense that sometimes there is a strong causality, but in other cases under other circumstances the manager moderates the causality. It depends.

Changing the rules of the game

Considering what we have seen so far, we would have expected to see Sales, being the workforce with the best fit to a management by measurement regime, as the area where the story of causality would be most dominating. Sales is also the workforce where we have so far found the most mathematical approach to target setting and performance measurement, and the workforce where we have seen most reluctance to incorporate non-business related targets in PA.

Nevertheless, prior to interviewing for final appraisal in Sales, I had been looking at the PA documentation of a Sales employee and noted that most of her targets were evaluated at ME. Still, the final appraisal rating was EE. This made me curious how such a discrepancy could appear in a mathematically based management by measurement system. I asked the Sales manager how this could be:

We have performed extremely well on sales of [product y] in the affiliate compared to anchor budget. But the sales targets to sales representatives have been very ambitious. She reached index 100,5. Normally, that would correspond to an ME. But as an affiliate, we have performed to an O on [product y], so we cannot have that all the sales representatives are NME or AE. Therefore, in cooperation with my manager, we have had to make a graduation so that she got an EE.

In 2010, this Sales manager's department was in a peculiar situation. Product Y was a new product that was introduced with considerable success and hit the budget targets much faster than anticipated when the budget was finalised in Q3 2009. However, sales representative's targets are set quarterly based on the sales in past quarters and on forecasts. Since the sale of product Y in Q4 2009 had shown impressive growth rates, sales representatives got very ambitious targets for Q1 2010, which they could not meet. Even though this was somewhat compensated in the following quarter, the yearly average ended at the ME level.

The reader will recall from the analysis of sensemaking in the target setting process that in Sales there is a strict definition of rating from the realised index. A 100,5 index corresponds to an ME. It is a simple, causal, objective and transparent interrelation where neither the manager nor the employee need discuss the rating because it has been clearly defined in advance what it takes to get a given rating.

However, in this case the model did not work. The manager and the organisation had to moderate the target in hindsight so that a 100,5 index realisation was not equal to an ME but to an EE. Following the principles established and communicated at target setting would be a serious challenge to the meaning scheme of managers and employees because the organisation, which is measured on its realisation of the budget, was evaluated to an O for its sale of product Y. How can then the people who actually realised the sale then only be ME or even lower? How can you create meaning from a story where the whole (the organisation) is successful while its parts (the sales representatives) are not, if you believe in a PA system that presupposes that organisational performance is the sum of individual's performance?

Management recognised the issue and enacted the change in the environment (the surprisingly high adoption rate of product Y) and retrospectively redefined the link between target and evaluation so that a 100,5 now corresponded to an EE. The link between performance and evaluation has been improved although at the cost of a poorer link between target and evaluation. Thereby a new and more meaningful story is established; that the success of the organisation is still linked to the success of the individual. But one could still reflect on why it is that the organisation performs to an O, while one of the best performing sales representatives only performs to an EE. A more meaningful story may have been established, but it still contains potential challenges to sensemaking.

While sales targets constituted 75% of target weight in PA in Sales, the remainder 25% was constituted by other elements. The most important of these are field days (the number of days a sales representative 'produces') and callplan compliance (the degree to which a sales representative realise targeted meetings with pre-defined customers). A Sales manager explained how these targets were evaluated at year-end:

Field days and callplan compliance targets are fixed. They are set at year start and they are managed through the sales system and KPI reports. It is all approved by

European HQ at year start. They are fixed no matter what. No matter if you get ill or anything. Then you have an explanation for why you didn't achieve it, but the targets remain the same.

In our previous analysis of managers' sensemaking in the target setting process we saw how the other three workforces seemed to have varying practices for how to adjust targets through the performance year as things unfolded and assumptions were challenged by the development of things. In contrast to this, we can see from the last two quotes how Sales practices two quite different logics to target evaluation; on the one hand sales targets that are hard numbers with pre-defined numeric success criteria which routinely get adjusted quarterly and where evaluation principally is automated based on index realisation. And on the other hand process targets that get set at year start and cannot be changed no matter how the environment changes or whatever happens through the year, but where again, target realisation is automatically calculated based on documented customer visits.

In Sales, the retrospective change of the rules of the game (changing the ratio between sales index and performance evaluation) fills at least parts of the meaning challenging gap between the O rating at the organisational level and the much more modest ME rating of a high performing sales representative that would have been the case without the change. By changing the rules of the game in hindsight, a more meaningful story can be developed, but essentially at the cost of a key foundation in the PA management by measurement regime; the strict causal relation between predefined target criteria and post-performance evaluation. In that sense the Sales case is an interesting example of how the story of causality in Sales is turned into its opposite - a story of non-causality - by a deliberate management action to change the rules of the game in order to be able to maintain meaning. This is illustrated in Figure 9-1 below.

9. Meaning Creation in Final Appraisal

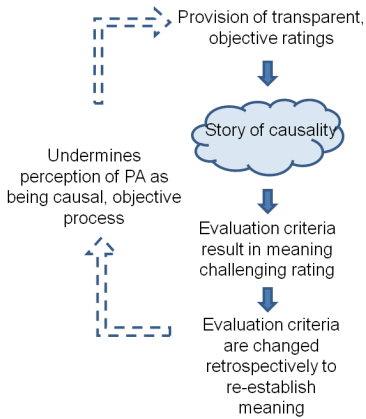


Figure 9-1: 'Changing the rules...' as meaning creation feedback process

Figure 9-1 illustrates how the provision of transparent and objective ratings is the ideal in the management by measurement regime of PA. This is supported by a story of causality between business targets, individual performance and final appraisal ratings. However, in this case the result is challenging meaning schemes of managers and employees because dynamic changes in the environment have made the causal outcome unacceptable. New meaning is established by changing the rules of the game. However, thereby the provision of transparent and objective ratings is challenged, because there is no longer a clear link between targets, performance and rating. Although the task at hand in Sales has the best fit to the management by measurement doctrine in PA, also in Sales the outcome of this logic can be so challenging to meaning schemes that the logic must be changed to be able to maintain a meaningful outcome.

9.3 The Challenge of Evaluating Behaviour

Above we have started our analysis of local adaption and meaning creation in final appraisal rating by focusing on how managers and employees established different stories to maintain meaning of the influence of business performance on rating provision. We also saw that one of the challenges to meaning schemes was the existence of moderators to the precedence of business performance. One particular moderator that all managers persistently mentioned as important was the impact of

behaviour. Below, I will look closer at meaning creation processes particularly regarding the inclusion of behavioural evaluation as a component in rating employees for final appraisal.

Inclusion or exclusion - two opposing translations in Sales

Let us start by looking at the situation in Sales at some length. As we have already seen, Sales is the workforce best fitted to a quantifiable, business target realisation focus. In a previous chapter, we saw how meaning creation through concealment characterised the handling of behaviour in the target setting process in Sales. So, for the performance year I investigated, behavioural targets were not used in Sales. Nevertheless interesting reflections and patterns around inclusion of behaviour in the final appraisal rating surfaced during the last interview round.

Above I analysed meaning creation in business performance evaluation in an example in Sales where targets on a particular key product had been set unrealistically high and where none of the interviewed sales representatives had realised sales above the predefined ME level. For one sales representative the average index realisation on her key target product had been 99, corresponding to an AE rating. Although sales constitute 75% of the weight in PA, the employee had nevertheless received an overall rating of EE. With departure in this apparent mismatch I asked the manager how the employee could get an EE when she did not even realise her most important sales targets. The manager explained that this was because the employee had been rated very high on behaviour. Then I asked if Sales management had more generally used the behaviour component to compensate for the fact that they had set business targets too high, and the Sales manager replied:

Yes, that's what we did in my opinion. But also, sometimes it is exactly where the targets are too ambitious that you really have to be professional and structured, rather than where the targets are reached more easily.

The manager retrospectively establishes a story to create meaning out of why, despite the fact that the business performance according to the predefined criteria only was to an AE, the employee ends up with an EE. The two key points of the story are: a) that Sales management had set unrealistically high targets so that it was impossible for sales representatives to over-perform and therefore they must be compensated through prioritising the weight of business performance lower, and prioritising the weight of behaviour higher, than normally; and b) that it is exactly

when the targets are very tough that employees have the opportunity to show their full behavioural potential. Although a more accurate relation between target evaluation and final appraisal would have led to a lower rating, in this case accuracy is trumped by plausibility so that a meaningful translation of what happened can be maintained.

The second argument about getting the better behaviour out of employees when targets are stretched could be seen as an argument for always setting unrealistically high targets. However, we know from the literature review that multiple studies would seriously question such a hypothesis. For example, within an expectancy theory framework, DeNisi and Pritchard (2006) argued that perceived inability of an individual to fulfil a required task, whether it is due to lack of capabilities or resources, will directly affect the employee's perception of the link between action and outcome and thus the employee's motivation and performance. So setting unrealistically high targets would according to these theories have a directly negative effect on the employee's motivation and performance. But, as we have now seen many times, meaning creation is not a rational game; it is based on what makes most meaning when actors look back at past actions of themselves and important others, and retrospectively establish plausible stories about what happened, and why it happened.

The above example made me curious to find an example where the opposite had happened; i.e. that an employee had over-performed on business targets but was rated lower because targets had been too lenient, based on a behaviour oriented argumentation. A Sales manager explained:

That can easily happen. I have one who got an AE. She actually over-performed [on business targets], but that doesn't correspond at all to the behaviour she has shown.

Again, this sales manager uses the moderating effect of behaviour in a situation where an employee, despite an unacceptable effort, has been able to over-achieve on her business targets. Giving the employee the rating she should have had according to the predefined business success criteria would not be meaningful to the manager, because it would not correspond to the behaviour the employee had shown. It would have been accurate, in the sense that there would be a clear causal link between target evaluation criteria and actual criteria, but it would not have been plausible. To maintain meaning, the manager set performance criteria aside and

gave the AE rating because it more meaningfully represents the manager's translation of what really went on, despite business outcome.

We have in a previous chapter seen how Production since 2009 has worked with centrally defined, cascaded behavioural targets. But we also saw that it was not until the 2012 process that this became common in the other workforces. Thus, in the 2011 cycle I followed, it was in all other workforces than Production mostly up to the individual management teams, or even the individual manager, to define what they understood to be good behaviour. A Sales manager defined what constitutes good behaviour this way:

For example, this weekend Y was at a conference, and Monday morning there was a mail from her with a full summary of it... That is really good work. You have to remember that. It is objective and tangible.

Through this example, the manager illustrates one possible way to show good behavioural performance outside what was agreed in the PA targets. The employee did not have a target in her PA that she should make sure to participate in the conference and deliver a summarising memo immediately after conference. Probably, there is no difference in added business value of making the memo Sunday night rather than Monday morning. It is not about realising PA targets or adding tangible business value. It is much more about attitude and demonstrating extraordinary engagement. By delivering the memo so that it is ready Monday morning the employee demonstrates that she is engaged and ready to pay sacrifices (sales representatives do not get overtime payment) for the benefit of sharing her experiences with her colleagues and the organisation.

Employees also have a clear perception of how behaviour is more about the 'how' than the 'what' of performing, even without having tangible behavioural targets in PA as reference:

Sales Employee: "I think it is about the way you realise your targets. This year, I had realised one of my targets to an O, but I didn't get it because I had done it in another way than they had preferred... If you looked at the result, it was an O, but my manager went in and evaluated differently."

When senior management, as in Production in 2009 and in Sales from 2012, formulate a rating matrix that clearly establishes procedures for how to balance behaviour and business performance, and if this matrix is broadly communicated among managers, we would expect that it would become more plausible, more

transparent, more legitimate, easier to communicate and thus more meaningful to the managers, to incorporate a heavier weight to behavioural performance when establishing stories about how they provide final ratings. We would expect managers to enact this change to create new and more meaningful stories where behaviour takes a more prominent place in the process of establishing a final rating.

However, contrary to this expectation, one Sales manager I interviewed kept underweighing the importance of behaviour even after the introduction of a rating matrix had clearly stated the principles for weighting the two dimensions:

It has to be something very, very significant to change the rating of the targets. It is the business realisation that counts. So for the rating, it is not behaviour. In a way that would also be a bit unfair because people evaluate [behaviour] so differently. Then it is suddenly my manager who has to evaluate if my behaviour is good. And if I had another manager she would have evaluated it differently. I am in favour of making targets measurable. Behaviour is important to talk about, but the targets are much more important. It is the measurability that counts in our rating matrix. It takes a lot to change a business target performance [by behaviour]. And that is also how it ought to be.

We see in this quote how the manager, despite the rating matrix which indicate a significant moderating effect of behaviour on final rating, still stick to what we could call the old story of behaviour being of minor importance compared to business performance. Sensemaking is not a fully top-down controllable process. Senior management can introduce a rating matrix but if cascaded PA targets are still almost exclusively about business outcome and a manager values transparency and objectivity highly, and at the same time feels that behavioural evaluation is subjective and highly dependent on the individual manager and even essentially unfair, then it takes more than communicating a rating matrix to make the manager see the alternative story as more meaningful. There may be a change in the environment that allows a different enactment of rating actions, but if the manager still extracts the same cues from the process as she did before the change, her story will remain unchanged and meaning creation will not be challenged to a degree requiring new stories.

Discreet inclusion of behavioural evaluation in R&D and Staffs

In R&D there was no rating matrix and no centrally determined and cascaded behavioural targets in the 2011 cycle. Nevertheless, as part of the target setting process some managers and employees may have agreed on 'softer' targets

regarding team contribution and other elements which link to attitude or broader behavioural performance expectations. However, with the lack of a central guidance on the balance between business performance and behavioural performance, also in R&D it is to a large extent up to the local level, or even to the individual manager, to establish a meaningful balance dependent on the manager's perception, what cues the manager extracts and how the manager enacts the environment. An R&D manager explained her view on the importance of behavioural evaluations in the provision of ratings:

You could say that if only you realise your targets, in theory it doesn't matter how you do it. But you can do it by being dedicated and by showing lots of energy or you can do the minimum effort required. You know, the final rating is an evaluation of your performance and everything else. Your attitude and things like that.

This manager is establishing a two-sided story where the target realisation on the one hand has prevalence so that "*in theory it doesn't matter how you do it*". But this is just theory – another thing is practice; how you do it does actually play a role. The final rating is not just an evaluation of your business performance, but in fact also of "everything else". Although behaviour played either no role at all or an inferior role in PA targets in the area in R&D I investigated, behaviour did play a role when this R&D manager provided final ratings of her employees.

A Staffs manager explained how her own experience in her role as employee collided with her experience in her role as manager:

I got a 3 this year and was specifically told that I complain too much. That is not something that is described in my targets. So that is a very explicit example of the subjectivity of my own PA. I think it is unfair - or, it depends. It depends on whether you can recognise it yourself... Personally, I found it unpleasant. But on the other hand, I do the same myself.

Through the literature review we learned from Coens and Jenkins (2002) that in PA schemes feedback is manager-centric and from Poon (2004) that disappointing ratings have negative effects on employee motivation. Further, we know from Langan-Fox et al. (1998) that there are frequent issues with the link between actual effort and evaluation and we know from Justice Theory studies like Thurston and McNall (2010) that there similarly are frequent issues with perceived fairness in evaluation contexts. The manager feels she is unfairly punished for being critical, but in the final sentence in the quote, the manager at the same time recognises that she does the same thing herself to her employees.

What we see here is a clash of role identity perspectives; to the manager in her role as employee it is “unpleasant” to get tough feedback on behaviour because it is perceived as subjective and without any reference in the actual PA targets. At first, the manager thinks it is unfair. But right after, she moderates this to being dependent on whether the employee can actually recognise the feedback. In other words, the feedback is fair if the employee acknowledges the feedback, but unfair if the employee disagrees to the behavioural feedback.

A special case of behavioural evaluation would be in situations where the manager has changed during the course of a PA cycle. We have previously seen how all workforces except Production were in 2011 reluctant to let behavioural targets play any significant part in PA targets. We have also seen how some managers are in different ways struggling to find a meaningful way to include behavioural performance in the evaluation of the employees, and how their senior management has also been reluctant to follow the example of Production and establish clearer and stricter governance on the same. So what happens if the manager changes? If business performance targets are at least documented in PA, there is something to relate to for the new manager. But if behaviour is not described in the PA targets and the manager enters late in the performance cycle so that she does not have much impression of what behaviour the employee has shown through the year, how is a meaningful story of behavioural evaluation then created? A newly appointed Staffs manager explained:

This year it has primarily been on their results, because I didn't know their behaviour so well. I have only known these people for four months. I don't know their history and it is not until now that I have a feeling of who pulls the most and who takes most initiative.

In this situation, behaviour is simply taken out of the equation. We have seen above how behaviour was seen by some managers as an important element in the final evaluation, but this manager who does not yet know her employees well enough to provide a fair evaluation of their behaviour, refrains from including behavioural performance in the evaluation. Alternative stories could have been created; the manager could have let the previous manager (who was still with the organisation) determine the behavioural evaluation or she could have made a behavioural evaluation based on her limited impression from less than a half year's exposure to

the employees. But in this example no behavioural evaluation was made and the manager relied only on the business targets in PA.

Putting behavioural evaluation to the front of the stage

In Production, where the concept of a rating matrix has existed the longest, a manager had no problem with attributing significant weight to behaviour in the provision of ratings:

It means very, very much. You may have done fine on all your targets but if you have gotten there by stepping on everybody else, then you don't get a good rating. Behaviour can pull you up and it can pull you down.

This manager creates a more explicit story of the importance of behaviour in the rating process than some of the more reluctant stories we have seen in Sales, R&D and Staffs. Here, there is clearly a place for behavioural evaluation. It is an integrated part of the manager's role to evaluate behavioural performance. Here, we do not see any concerns about subjectivity and challenges to fairness – evaluating behaviour is important because it “*means very, very much*” in the establishment of a final rating. The manager has internalised the inclusion of behavioural evaluation into PA and a meaningful story can be established by enacting the environmental factors available in Production; i.e. senior management's explicit emphasis on the importance of behavioural performance through the establishment of centrally defined behavioural targets, through establishment of an evaluation matrix, through the provision of particular definitions and guides to behavioural evaluation and through communication of the same. Since this has been a process running over the last three years in Production, organisational adaptation and individual internalisation is more advanced than in the other workforces.

Discussion

Table 9-1 below shows a summary of the stories about behavioural targets and appraisal established across the four workforces, compared to the workforce relation to the management by measurement regime of PA.

9. Meaning Creation in Final Appraisal

	Sales	R&D	Production	Staffs
Fit to management by measurement	High	Low	Mixed	Mixed
Behavioural targets used	No	No	Yes	No
Target setting story	Concealment	Taken for granted	Adaption	Relocation
Behavioural appraisal story	Inclusion or exclusion	Discreet inclusion	Behaviour at the front of the stage	Discreet inclusion

Table 9-1: Summary of stories about behaviour in PA, by workforce

First, Table 9-1 shows that Sales was the workforce with the best fit to the management by measurement regime, while R&D had the worst fit. Production and Staffs had more mixed relations to the management by measurement regime. In 2011, it was only Production that worked systematically with the inclusion of standard behavioural targets and with clear communication about the importance of this in the later evaluation. In target setting, four different stories were identified, so that in Sales it was a story about concealed behavioural evaluation. In R&D, it was a story about behaviour being taken for granted (and therefore superfluous in targets). In Staffs, it was a story about relocation of the behavioural dialogue to the individual development plan. And in Production it was a story of local adaption of the centrally cascaded standard behavioural targets.

In the appraisal phase, again four different stories were identified. In Sales we saw two contrasting examples of manager stories; one where behaviour evaluation was excluded because business performance was perceived as the only relevant element in performance evaluation. The other was an adherence to the concealed rating matrix according to which behaviour was a significant moderator to business performance. In R&D and Staffs we found a story of discreet inclusion of behavioural evaluation, although behavioural targets were not available to justify this. In Production, behavioural evaluation was openly communicated and prioritised as an important element in employee appraisal.

Although the picture is a bit blurred, a few important coherences are worth emphasising. First, different local adaption is clearly manifested in the workforces when it comes to the question of how to relate to behavioural targets and evaluations.

Second, neither Sales nor R&D, who had the best and the worst fit to management by measurement, were systematically establishing behavioural targets and evaluation. This may at first seem strange, but in fact both stories make sense. For Sales, it makes good sense to be sceptical about the more subjective behavioural evaluation component because it is challenging the objectiveness of the traditional focus on realised sales indexes that is more in line with the management by measurement regime. For R&D, it makes sense to be sceptical towards measuring behaviour, because the idea about measuring is as such problematic due to the low fit to the management by measurement regime.

Third, Production and Staffs that both rely on more mixed positions in relation to the management by measurement regime, two different approaches to behavioural targets and evaluation have been selected. This difference in approach is not possible to account for solely on the basis of the data in Table 9-1. To account for this, we would have to include more elements into the analysis. For example the differences in workforce culture and history, where we would find that Production has had a more standardising and systematic approach to PA than the other workforces, with more particular policies valid at workforce level. In that way, it is not surprising that also in respect to behavioural targets, Production is the workforce with the more systematic approach.

9.4 The Challenge of Rating Distribution Policy

One important design decision organisations need to make when embarking on a PA scheme is whether the rating scale is an absolute or a relative scale. On an absolute scale, each employee will be evaluated against defined criteria on an absolute scale, ideally unambiguously defined in the targets documented in PA. Thus, the evaluation is independent of the performance of others. In theory, everyone in an organisation could be outstanding within the framework of an absolute scale.

In contrast, a relative rating scale means that each employee's performance is evaluated not just against the targets set for the employee, but also against the performance of others. The purpose of having a relative rating scheme will often be to be able to differentiate talent in an organisation. Further, where there is a link between rating and reward (bonus, promotion, salary increase etc.), a relative rating

scheme also aims to facilitate a meritocratic distribution of ratings and rewards where scarce reward resources can be distributed in a way so that the best performers get most.

Rating regime is in the literature often dealt with, with departure in the question of perceived fairness. The question of fairness relates to whether managers and employees perceive the rating process and outcome as fair. According to Justice Theory there are three basic dimensions of justice perception; distributive justice, procedural justice and interactional justice (Thurston and McNall, 2010. Narcisse and Harcourt, 2008. Roch et al., 2007. Poon, 2004). In the following I will reuse these categories but substitute 'justice' with 'fairness' which was the term mostly used by the interviewees and which I believe in this context has interchangeable meaning with the concept of justice. The question of fairness is central to the general legitimacy of the appraisal scheme but also to the effect ratings have on employees "affective and behavioural responses" to the scheme (Thurston and McNall, 2010).

One particular case of a relative distribution is the forced distribution, where it is pre-specified that certain proportions of employees have to be placed within each rating category (Grote, 2005). So not only are employees evaluated against each other, there are also only certain shares of employees that can be rated in the top categories. And similarly, there is a requirement to rate a certain number of employees in the lowest categories. In Grote's (2005: 137) words, the purpose of installing a forced distribution is to force managers to take the evaluation more serious so that real talent differentiation is achieved, because

...as we have seen, too many managers would rather fudge the facts and inflate ratings than face having the hard conversations that truthful performance assessments often require.

So, in a sense the organisational need for instalment of forced distribution is a general and basic mistrust to managers' ability to fulfil their role in a PA scheme without being forced to live up to a specific predefined rating distribution.

As we have previously seen, in PharmaComp there is no corporate policy of forced distribution in PA. However, we have also seen that multiple workforces are working with similar concepts in the shape of "ideal distribution", "expected distribution" and similar phrasings. Below, I will start by looking at stories from R&D, Production and Staffs about meaning creation through translations of relative distribution. Secondly, I

will look at stories of meaning creation through translations of absolute distribution in Sales.

Ratings as relative distribution

In PharmaComp, the transition from a stage of multiple local PA practices towards a more homogenised scheme where stricter enforcement of distributions are becoming commonplace, has been a development taking place over more than a five year long period. I asked an R&D manager who had been working long with the company, if she also provided employee ratings before the corporate PA system was launched:

No we didn't. It was something new. It was something we had to discuss first, if it was a good idea. We ended up by thinking it was a good idea - at least until we got to where we have to reach a specific average. Then, if someone over performs you have to identify another underperforming to reach the target. That is meaningless.

This manager started out by being supportive of using ratings as an instrument for provision of feedback to employees. To her, it was when distribution policies were introduced the system became "*meaningless*". As we have already seen previously, it is clear that Production, R&D and Staffs enforce a relative scale. Further, in the areas I investigated in all these three workforces, there was different ways and degrees of rigidity for how to install the realisation of a specific distribution. Most rigidly, we saw that in the R&D units I examined, the units were instructed to achieve a specific one decimal average rating and a specific numeric distribution within each of the five rating steps. In Production, we saw how a "recommended" or "expected" distribution was communicated and how managers could deviate from this, but would be challenged when doing so. Similarly, in one of the Staffs areas I was investigating we saw how an "ideal distribution" was used from which deviation could also be allowed at the department level, but again not without being challenged by senior management.

I asked a Staffs manager, who was particularly sceptical towards the way relative ratings were provided in the organisation, why she thought her organisation put so much emphasis on governing the process of relative evaluation and rating distribution:

Staffs Manager: "I think it is because they fear that managers would overrate their employees. Because it is easier to rate people high than to rate them low."
Interviewer: Do they do it to force managers to take the conflicts because managers would otherwise be reluctant to do that?" Staffs Manager: "Yes". Interviewer: "Do you agree to that hypothesis?" Staffs Manager: "No, not at all."

The manager is of the perception that senior management believes that managers would, without some sort of coercive governance, rate their employees too high due to managers' tendency towards conflict avoidance. We already heard from Yariv (2006) that managers are reluctant to give negative feedback and from Coens and Jenkins (2002) that managers are generally lenient in evaluations because their preference is to avoid conflict, so it is not an unreasonable assumption that senior management has similar concerns. In fact, this anticipated manager lenience is one of the key reasons listed by a "forced ranking" advocate like Grote (2005) for why senior management should introduce such schemes in the first place; to force management to take tough but necessary decisions that they would otherwise be reluctant to take. This is what Coens and Jenkins (2002) have termed the introduction of "resource scarcity" of top ratings. Since the whole rationale behind the forced distribution is to ensure differentiation of employees, senior management will not be able to accept a general lack of differentiation due to managers' leniency. Therefore they, according to the story of the above quoted manager, install and enforce rating distribution policies.

Roch et al. (2007) found that absolute rating schemes are generally perceived as fairer than relative rating schemes. So, one challenge to sensemaking of managers and employees working within a relative distribution paradigm will be to handle possible perceived unfairness. In Staffs, one of the managers I interviewed referred primarily to procedural fairness when reflecting on whether she thought employees had been treated fairly:

No, not completely. It is not always transparent to them either; why they get one rating and not another. Some of it gets very subjective... I think we try to do it as objective as we can. But fair and totally objective and transparent - no, that's not how it is. Not at all.

Management do their best to be objective, but it would not be meaningful for this manager to claim that the process is completely fair to the employees. This is both due to the inherent subjectivity of ratings which we have dealt with many times above. But it is also because the process is not transparent to the employees; lack of transparency seems in itself to be an element in lack of procedural fairness, because lack of transparency makes it impossible for employees to find out whether the outcome of the process (the rating) is provided in a fair way.

One element of fairness concerns whether people feel they are rated in accordance with the same standards across the organisation. One Staffs manager had a particular problem with the way this was practiced in her organisation:

I don't think the distribution is relative when you look at VP level and above. So that is really a dilution of the whole concept.

As we saw in a previous chapter, the distribution of ratings is direct proportionally connected to the hierarchical level one is working at, so that the higher one is in the hierarchy, the higher is the average rating. The practice of rating distribution that this manager refers to is one where there are different standards; Vice Presidents (VPs) and higher ranking levels are rated according to a standard where relativeness is non-existent - otherwise they could not all get 4s and 5s. In contrast, employees are rated relatively and a certain share of 3s and lower ratings are required and enforced by senior management.

Another critique of relative distributions is that they install a more competitive and less collaborative mindset in employees because the failure of others has an indirect positive impact on the evaluation of oneself (Coens and Jenkins, 2002). I asked an HR staff who had been heavily involved in improving the PA process in an area if she had experienced any of that:

We talked a lot about absolute versus relative. What is it really? We are somewhere in the middle. We don't really clarify whether it is absolute or relative. And I think that does make a difference. The SVP showed our ideal distribution to our employees – what we had had and where we were going. And people asked: “does that mean in my department where we are five, one of them has to get a 2?” And [the SVP] said no; this is a distribution over a population of 200+ employees. So there could be three of you who all got a 4 in your department.

In this area, the concept of an ideal distribution was introduced. Faced with this, managers were curious to understand how this impacted the requirements for the distribution in each of their own departments. An ideal distribution that has to be realised within a given area is by definition a relative rating scale because you have to compare people to make sure it is the right ones who get into each category. But as we see in the quote, senior management made it clear that the ideal distribution was meant to cover the entire organisation. It did not necessarily apply at the individual department level where there would be room for deviation from the ideal distribution. However, for the ideal distribution to be realised, any department deviating from the ideal distribution would have to be balanced by a counterweighing

deviation in other units within the area. This way, the relativeness was moved from the individual level of employees to also include the group level of units, which then again would affect the ratings at the individual level in the affected units. This iterative complexity of rating distribution in any unit deviating from the ideal distribution, or affected by other unit's deviation from the same, is probably what makes the HR staff think that the organisation is somewhere in between relative and absolute distribution frameworks.

To enforce the realisation of the ideal distribution and to make sure managers make broad use of the full rating scale, this Staffs area in one year put pressure on managers for them to identify low performing employees who should be rated AE. An HR staff involved in the process explained how the message had been conveyed to managers:

The message was; look at your people and see who the worst one is. He gets the 2. So, you know, [the managers] didn't like that. But on the other side, they found those 2s. I coached some of the managers before giving 2s – for some of them it was hard because it was the first time.

This enforcement of identification of low performers is related to what Grote (2005) calls "forced ranking", i.e. the idea that independent of, though often in parallel with, an appraisal rating distribution scheme, manager's should be able to list their employees on a ranking list in order of performance, so that they will be able to identify who is best and who is worst. The enforcement of identifying 2s was initiated by senior management who thought it would be unlikely that managers would have volunteered to move down their own people from the ratings they had originally proposed if senior management had not interfered. With the intensified focus on enforcing an ideal distribution and identifying low performers, there was some concern among middle managers:

HR Staff: "One manager said 'this means to me that I can say to my employees: You know I wanted to give you an Exceeds but I can't because this is the distribution we are working at.' [The Senior management representative] stood up and said: 'If you do that you are failing me as a manager.' So he got an opportunity to really express himself very clearly as a leader."

It was very important to the senior management representative in this area to make it clear to middle management that they had a clear responsibility to take ownership of the ratings they provided within the ideal distribution scheme in the area. It would not be tolerated if managers blamed the distribution system for any ratings that may be

disappointing to the employees. The distribution was the playing field and the managers had to take personal responsibility for how they rated their own employees within the boundaries of that playing field even when the system forced them to provide lower ratings than they would have done without the system. Blaming the distribution scheme would be an easy way for managers to move any conflict with employees over ratings from the personal manager-employee relationship to the less personal employee-system level, and maybe even installing manager and employee to be on the same side against that system. However, it would also undermine the employees' trust in the system and would thus undermine the system itself.

The concept of an ideal distribution rather than a forced distribution was intended to install a less rigid system leaving some room for managers' discretion. Nevertheless, a Staffs manager explained how she practically had understood her superiors' communication about the ideal rating distribution:

Last year it was really bad. We were hit by this message that there could only be two in the area that could get a 5 and there should be two 2s in each department. They wanted at least 10% 2s. It is simply not true that 10% are 2s... Then you line up your employees and the ones at the bottom are defined as 2s. Personally, I don't like that. It is not particularly pleasant if those employees have actually reached their targets, but they lose when compared to the next in the line.

Again, we see that this manager feels that the degrees of freedom in terms of possible deviation from the ideal distribution are limited by tangible requirements to avoid giving too many top ratings and to enforce identification of low performers. The manager implicitly refers to the method of forced ranking in her description of how she practically went about identifying the required 2s. So, although forced ranking is not formally introduced in PharmaComp, in this managers' department it for all practical purposes is. The manager does not like to have to communicate 2s to employees who "have actually reached their targets". But the manager loyally does it and she also stays loyal to senior management's requirement of taking personal responsibility and not blaming the system.

To the employees in Production, Staffs and R&D, the communication about an absolute or relative use of the rating scale is not always perceived as very clear:

R&D employee: I have understood that we get rated on an absolute scale. But in reality that is not the case, because in the end we are compared with each other. They want to cover the entire rating scale."

When it comes to what consequences employees see from the perceived relativeness of the ratings in terms of the incentive to cooperate, the same R&D employee explained:

That the rating is relative works fine for the team players. But for those who are not, I am sure they think about it. [It means] you don't help each other as much, if that could influence your own rating. The relative element in it, I find that peculiar. Where I was working before [joining PharmaComp] they looked at your performance in relation to your own targets. Not in relation to how your colleagues had performed.

A Staffs employee also had only a vague idea about the precise evaluation process, but imagined there was a relative distribution based on a mechanistic logic suited to the purpose of distributing rewards:

I imagine it is a dictate from above. That 'this is how the [bonus] pool looks like' and then management has got to make the ratings fit into that. Somehow that is a hopeless way of doing it. It would fit on an average unit. But when you have a unit filled with smart people, it is hopeless... People who would get an Exceeds elsewhere, get a Meets here, since it has to fit a normal distribution. That is a bit demotivating.

Senior management's initiative to improve the practice of PA in the Staffs area was running over a two-year period and involved many other elements than the introduction of an ideal distribution and enforcement of identification of low performers. It also included training sessions for managers and a closer look at the relation between performance rating and other rewards:

HR Staff: "It got better in the years that I was there. We were doing it with bonuses too. It definitely was getting closer... We worked on this for two years while I was there and I think we made a lot of progress. It was huge. From the work group, the management team, the managers, training sessions, and sending articles out to the managers."

The area was getting closer to an ideal distribution, as we have also seen was the case in other workforces where stronger enforcement of rating distribution has been implemented over recent years. But to get there, the middle managers had to be brought to the same page, and it is clear that in this area it took a lot of management resources to get there. How did managers and employees experience this intensified focus on fulfilling the ideal distribution? A Staffs manager explained:

It is communicated that we do not have a forced distribution. But if you ask me, that's exactly what we have. If we have too many EEs we get challenged. And some [employees] will be moved down when it get reviewed higher in the system.

This manager sees a misfit between the communicated flexibility of the system and the reality she experiences during the rating process. It is explicitly communicated

that there is not a forced distribution but in reality the manager experiences that positive deviations from the ideal distributions are challenged to a degree so that it in reality is experienced as a forced distribution. Thus, this manager perceives that there is a disconnect between what senior management communicates and what the manager perceives is practiced. Such a disconnect is a challenge to the meaning scheme for the manager who must maintain a meaningful story of what she is actually practising when providing ratings. As we would expect from a sensemaking perspective which is oriented towards interpreting actions, the extraction of cues is in this case clearly much more oriented towards actions (what do I see senior management practice?) than towards symbolic articulation (what do I hear senior management say?). Therefore, the manager's story is that no matter what is communicated, the organisation practices a forced distribution approach to PA rating and therefore the manager's role and responsibility is to keep as close as possible to the ideal distribution when evaluating her employees.

Ratings as absolute distribution

In contrast to the three other workforces, Sales is less focused on realising a specific rating distribution. I asked a Sales manager if she had to reach a certain distribution when rating her employees:

No, we don't have that. You could say that there are quite a lot who get an ME. And because we update the sales targets [quarterly] and sales targets constitute 75% of their PA, it means that it is pretty difficult to get much over 100%.

In the part of Sales I investigated, I found that rating calibration was less formal and with less emphasis on reaching a certain rating distribution. Senior management in Sales was apparently less concerned with rating distribution and did not enforce policies to install stricter governance of the same. We have already seen how business targets in Sales were very quantitative and how rating criteria were predefined⁶¹ and based on 1:1 meetings between a manager and her superior rather than on larger, cross-unit calibration meetings like in the other workforces.

A likely reason for this difference in approach to rating scale would be that Sales is not basing its bonus allocation on PA ratings, but on a parallel bonus system. Thus,

⁶¹ However, we have also seen how behaviour and changing circumstances in some cases could moderate or even alter the predefined criteria.

in Sales there will not be any need for stronger differentiation of ratings to be able to produce the required differentiation for bonus allocation purposes.

I asked a Sales employee if she knew whether her rating was made according to a relative system so that only a certain percentage could be rated EE or above:

No, I don't think so... I believe my manager is good at evaluating us individually. She knows us well.

In this area of Sales, it appears that not only is rating absolute rather than relative, the manager is also fairly good at making the rating match the perception of the quoted employee, i.e. there is a perceived correspondence between expected outcome (relation between targets and performance) and actual outcome (overall rating). It is exactly this relation that is at risk in a relative rating scenario.

One of the Sales managers I interviewed was also quite explicit about the different effects of distributive and procedural fairness in rating provision in her department:

I think they perceive [their evaluation] as fair. I think they are happy about the comments I write. Particularly in a year like this where they didn't think the targets they got were fair.

According to the analysis of this manager, her employees generally perceive the distribution of ratings as fair. However, the manager is aware that 2011 was a year where the sales representatives in her department had very challenging targets, as we have previously seen, to a degree that was unfair to the employees because it made it close to impossible to realise the targets. The unfairness of the targets became even more of a procedural issue to fairness because, as we have seen, the management team was evaluated against the much lower budgeted sales in the anchor budget and therefore received high ratings and bonus:

Sales Manager: "This year we had the challenge that the management and back-office team went home on index 300 and got a large bonus because we were measured against anchor budget. While the sales team didn't get bonus - that's not fair."

In this manager's perception procedural fairness is clearly challenged by the fact that employees are evaluated against other targets than the management team is. In a sensemaking perspective, we can formulate it like this: It is difficult to establish meaningful stories from a situation where procedural fairness perception is undermined by significant and systematic variations in targets so that some (management) can significantly overperform while others (sales representatives) at

the same time underperform. However, here the problem was the targets, not the ratings. The rating itself is absolute and perceived both by managers and employees as fair when held up against the documented targets.

However, in some cases Sales employees were less impressed with the fairness of the system. These were what we could call "extreme" cases where external factors heavily influenced target realisation but evaluation did not take any account of that. In those cases, distributive fairness perceptions appear to sometimes be negatively affected by the strong causality of an absolute rating regime. For example, one high performing Sales employee had really struggled to meet a target that had been severely out of touch with what was in fact achievable:

I think it is difficult to make a fair system... Whether it is fair, I cannot find out. It is so aggravating to achieve 99% on something you have worked on like crazy. That feels unfair.

This employee understands why she gets a low rating although she has performed best in class among her peers on a very stretched target. In a target-rating causality perspective the rating may be rationally fair, but considering that the targets lack realism and relation to the relative performance of peers, it "feels unfair". However, in an absolute rating regime, relative performance is not considered.

Another employee in Sales had spent the first two months in a year helping a new colleague by handing over customers in an area previous managed by the employee. Although this was done in full agreement with the employee's manager, the effect was that her callplan targets for number of customer visits in her own area could not be met. As previously explained, Sales callplan targets are defined at European level, and there is no process for accommodating for changed priorities in the shape of updated callplan targets.

Sales Employee: "I just realised that this means that when we reach 1 March I will have missed 2 months in my own callplan... You know, I could have pulled out and let [the new colleague] take care of herself to be able to concentrate on my own callplan. But you don't do that. Not to a new colleague and not to the business. At some stage it may start to sound like bad excuses, but it is how reality is. These systems try to capture reality but that is just so very difficult sometimes."

In this example the assistance to a new colleague is delivered at the cost of the ability to meet own targets. Again, in a relative rating scenario there is the opportunity for management to say that the Sales representative's performance, relative to her colleagues, is better, because despite the lower delivery on callplan she also

managed to help her new colleague get established. However, in a strictly absolute rating regime, the relative effort compared to colleagues who did not do the extra handover work, is not taken into account.

Discussion

Below, the two different approaches to rating are depicted in Figure 9-2.

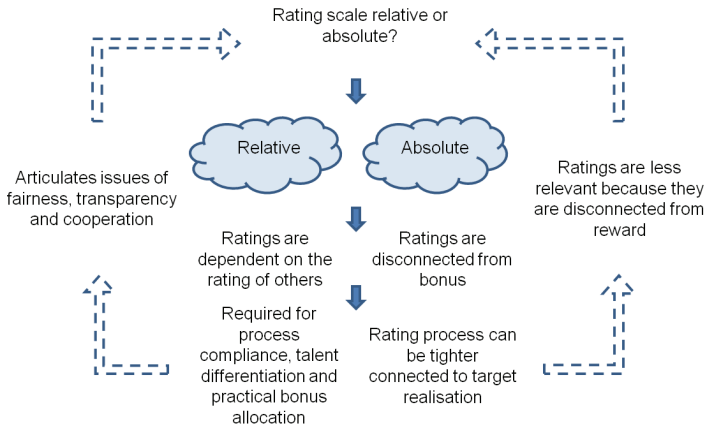


Figure 9-2: Relative or absolute - two different approaches to rating

The figure illustrates how in Sales, the practical use of an absolute rating scale disconnects the individual rating from the performance of the ratee's colleagues. This is possible partly because the bonus allocation is disconnected from the PA rating process so that there would not be any consequences to bonus allocation of an uneven distribution of PA ratings. In itself, the missing link to reward is a challenge to meaning schemes because it tends to make sales representatives indifferent to PA. On the other hand, it is possible to create meaning around the fact that a fairly direct link can be maintained between target realisation and PA rating. So, in that sense and in a fairness perspective, the absolute rating scale model has an advantage. In this case unfortunately at the cost of the employee's perception of the relevance of the PA scheme.

The three other workforces take the other route through the meaning creation feedback model. Here, the rating process is clearly relative so that ratings are not just dependent on one's performance, but also on the relative performance of others. This

challenges meaning schemes of managers and employees in terms of fairness perceptions and perceived effects on incentives to cooperation. In this case, meaning is created around process compliance (management says we must do it like this) and practical requirements for relative talent differentiation to be able to distribute bonus pools in a rationed manner. In the feedback phase, this path articulates a number of PA issues like lack of perceived fairness, lack of transparency in the shape of clear connect between target realisation and final appraisal, and issues with incentives to cooperate and help one's colleagues.

The difference in the use of rating regime across workforces is a very significant example of how a standard PA system is locally adapted. Two very different regimes, an absolute and a relative rating process, are used in the same organisation and within a single PA scheme. The use of the absolute rating scheme in Sales matches the fit between the task at hand in Sales and the management by measurement paradigm in this workforce. Being the workforce most inclined towards quantifiable, unquestionable hard facts as the basis of targets and target evaluations, Sales will be able to maintain the clearest link between target, performance and rating within an absolute scheme, because here ratings will not be 'polluted' by the moderating effect of the performance of peers. On the other hand, in the three other workforces where the fit to management by measurement was less evident, there is a weaker case for an absolute rating regime. Further, and very importantly, the missing link to bonus allocation in Sales make it practically possible to maintain an absolute rating regime, while the opposite is the case in the three other workforces, where differentiation of employees is required to be able to distribute reward.

9.5 The Challenge of Calibration

Calibration is the process of aligning rating practices across units within a given area of the organisation. In a previous chapter we saw how calibration in 2011 was one out of five prioritised "global minimum benchmarks" that all units in PharmaComp should follow.

The purpose of calibration in PharmaComp is at least twofold. First, calibration should support management teams in aligning their rating criteria so that all employees within the area to a reasonable degree are rated against comparable

criteria and so that management teams share input to the feedback and rating of each other's employees. This process will ideally both qualify the amount and quality of knowledge that is used in the provision of ratings and at the same time increase the protection of employees against arbitrariness caused by differences in managers' approaches to rating. Second, calibration is the process whereby senior management in the three workforces operating a relative distribution make sure to realise the preferred rating distribution across the organisation. This is done by iterating discussions and recalculations of distribution ratios and sometimes rating averages, until the desired distribution has finally been met.

Rating calibration is not a subject that has raised a lot of interest in PA research, at least not under the same heading. Grote (2005: 153) describes the practicalities of the calibration process as a way to handle rater reliability bias, which is a purpose similar to the first purpose mentioned above. Armstrong (2007: 113) has similar thoughts about balancing the "favouritism or prejudice" of managers through calibration, but he also mentions that a supplementary purpose is to challenge managers whose "distribution is significantly out of line", corresponding to PharmaComp's second purpose described above.

Calibration was the part of the PA process the employees new the least about. Most employees I interviewed, only had a very vague idea about what was going on in the calibration process. On the other hand calibration was also the single process step that the managers I interviewed felt most strongly about and wanted to talk most about. It was probably also one of the elements that gave way to most manager frustration and was challenging managers' meaning schemes the most. Therefore it is only natural, that managers' perspective will constitute most of the empirical examples in the below analysis.

However, before turning to the analysis of meaning creation in the calibration process, I will start by discussing the case of Sales which I see more as an example of non-calibration. In my view, it would be misleading to call the Sales process of adjusting ratings a process of calibration. In fact, the term 'adjustment' is more precise. Below, I will argue why I think this is so.

Non-calibration - rating adjustment within an absolute rating regime in Sales

Above, it was found that Sales was in certain ways distinct from the three other workforces. First, targets were cascaded from above without any employee influence. Second, the majority of the targets in Sales were of a quantitative nature, easily measurable through a sales accounting system. Third, for sales targets there was predefined index realisation/rating causality so that the employees could at least in theory themselves calculate the target rating without waiting for the manager appraisal. Fourth, there was no strict policing of rating distribution. If an area over-performed on numbers, they would also over-perform on ratings. No one saw it as a problem, at least in theory, that every employee got 4s and 5s, as long as the numbers justified it. There was no talk about a normal distribution. And fifth, there was no rating-bonus causality as in the other three workforces.

Also in calibration, Sales deviated from the other three workforces I investigated. In Sales, calibration was not organised as a management team meeting but rather as a 1:1 meeting between the manager and her superior. When ratings are merely calculated, there should be less need for peer input and iterative discussions in management teams. The calibration then is more a sanity check to make sure no errors are made in the rating process due to misunderstandings or extra ordinary events. In that sense, it is imprecise to even talk about a calibration process in Sales. At least, the two purposes of the calibration process in PharmaComp mentioned above, to support management teams in aligning their rating criteria and to make sure to realise the preferred rating distribution across the organisation, are not relevant in Sales. In fact, calibration is by definition not particularly meaningful in an absolute rating regime, because calibration in the first place is aimed at dealing with the relativity of ratings.

One of the Sales managers I interviewed had to change a rating upwards after an adjustment meeting with her superior manager. But since in Sales there is no focus on realising a certain distribution or average, in this example the change was concerning a specific individual's rating and based on a specific argument relating to the realisation of the absolute rating regime:

Sales Manager: "I had to move one from AE to ME-, because the sales targets were actually realised, so that weighed somewhat higher."

In this example, the initial rating by the manager was actually incorrect in the sense that the manager had not weighed the realisation of sales targets high enough. Other targets may have been unrealised and there may have been some issues with behaviour that made the manager suggest an AE rating. But through the adjustment process with her superior, it was made clear to the manager that the rating had to better reflect the precedence of realisation of sales targets. In this example, the change of rating is less of a challenge to the manager's meaning scheme because there is a clear argumentation related to the performance of the employee and the prioritisation of targets. As it seems, the manager has accepted that an ME- is the more correct rating and a new and meaningful story can be established around why this is so, which can later be conveyed to the employee.

The example illustrates how the purpose of the adjustment process is very different from the purpose of the calibration process in that the adjustment element is there to make sure to safeguard the absolute rating regime itself.

Before turning to the analysis of meaning creation in calibration in the three other workforces, let us take a look at a quote from an HR staff who was explaining her perception of how calibration works in PharmaComp:

Typically, ratings are too high. If you deviate [from the expected distribution] you need to submit explanations. If the unit has over-performed it may be all right to be in the high end. But if not, the feedback will be to go back and revisit the ratings.

In the quote we see how calibration is closely related to the concept of relative rating and predefined distribution categories. Each manager has the freedom to rate her employees within the boundaries of the expected distribution. However, the general tendency is that ratings are inflated because managers provide ratings in the high end of what is expected. In cases of visible unit over-performance relative to other units, a unit may be able to get through calibration with a higher average rating than other units. But if higher rating average cannot be justified from higher relative unit performance, the manager will through the calibration process be told to go back and revisit the ratings. In effect, the manager will be told to change ratings downwards, so that the distribution is in better alignment with the expected distribution.

Two different approaches to calibration can be separated out of the study of calibration within relative rating regimes in units in R&D, Production and Staffs. Although there are many similarities between the two approaches, there are some

distinct differences that can be analysed and understood along the lines of the workforce particularities. I have called the R&D approach "Meaning creation through controlled inclusion". The Staffs and Production approach I have labelled "Meaning creation through management negotiation".

Meaning creation through controlled inclusion in R&D

The case of calibration in R&D is somewhat ambiguous. In some ways, R&D is operating the most strict, quantitative, almost automated regime when conducting calibration. But at the same time, R&D stands out as the only workforce in my study where employee self-evaluation in effect had any real influence on calibration of employee ratings.

R&D Manager: "We have to land on a specific average; 3.8 or whatever it is set at. You could say that is a bit funny, but it depends on the overall performance of the organisation... For the entire organisation we line up the employees with salary and their rating with one decimal. And then the spreadsheet automatically calculates the bonus."

In the R&D area I investigated, employees are listed in a spreadsheet with a one-decimal rating for each, so that in practice a 50-point rating scale is utilised for calibration purposes, rather than the 5-point rating scale defined by the corporate PA process. However, it is the 5-point scale that is used for the later communication to the employees, after being rounded in accordance with specific quotas. Based on area performance each unit is then required to meet a one-decimal rating average defined by R&D senior management. Further, quotas are defined for the distributions of employees within each of the 5 main rating categories, so that a unit can have only X numbers of Outstanding, Y number of Exceeds and so forth. Both the overall rating average and the category distribution must be met, and this is the overall purpose and success criteria of the calibration process.

An R&D manager had some concerns about this approach:

I don't know about that calibration process. That you have to reach a certain average and a certain quota within each category. Sometimes it can be a bit restrictive... It is really a shame if you work in a high performing unit, right. There, the demands will just be so much higher.

We see how the calibration process in this manager's perception installs the units and employees within an area as competitors. As the overall area average is predefined, then any high performing unit or employee would have to get a

correspondingly higher rating at the cost of their neighbour units or employees. Otherwise the pre-defined average would be jeopardised.

The very quantified and controlled approach in R&D could be expected to have a negative impact on employee perceptions of the PA system. After all, being ranked in order of comparison to your colleagues with a one decimal precision would as a minimum raise some interest and concern into the validity of the ranking. If the ranking was visible to the employees, employee A would likely question whether the performance of employee B was really 0.1 better than employee A's and so forth. The competition element would become very visible with the consequential expected negative side effects such as challenges to fairness perception (Thurston and McNall, 2010; Roch et al., 2007), deteriorating cooperation (Kohn, 1999) and, in the last instance, increased intention to quit (Poon, 2004). However, and one might say luckily for management, the employees are generally not very well informed about what goes on in the calibration process.

I asked an R&D manager if the employees knew that as part of the calibration process they were all placed in one long spreadsheet and ranked in order from lowest to highest performer:

No. And they don't need to know that. They know it is a relative scale and that they are compared against each other. I also make an effort to explain that it is not just me - I get all these inputs and then we discuss it in the management team.

And further:

The employees are not informed about the decimal, they are told if it is a big or a small 3, for example. We only use it [the decimal] to calculate bonus.

The message to the employees is that calibration is about cross-checking the relative ratings across units based on inputs from the entire management team. According to this manager, employees do not need to know the more practical elements of calibration; e.g. that all employees in the area are in fact ranked in order of performance with a one-decimal precision.

An R&D manager told me they had previously utilised an approach for settling ratings before employees were involved in the process, similar to what we will below discuss when turning our attention to calibration in Staffs and Production. However, recently the area in R&D I investigated had included the incorporation of employee self evaluations into the calibration process:

R&D Manager: "It has varied from place to place. Previously, we did it differently. Back then, we gave a shadow rating. You collected input from project managers and so forth and then you provided your input already at Midyear. And then there was a pre-calibration. So that when you had the meeting with the employee, it had already been decided how she would be rated."

So previously, the ratings were decided much before the employees were requested to provide their self-evaluation. The main reason for having an early decision on ratings was the need for ratings in the allocation of bonus. So the administrative requirement for early ratings was prioritised against the possibility of including employee self-appraisals in the rating calibration process. However, after the change of procedure in this R&D area, the final ratings could now be affected by employee input because calibration was not finalised until after employees had submitted their self-appraisals. I asked an R&D manager if it was then possible for her to communicate the final appraisal rating at the meeting with the employee, since calibration would take place after employee self-appraisal had been discussed at the final appraisal meeting:

No, I can't do that. I give an evaluation of the individual targets and a range indication of what her rating will likely be... So at the meeting I said to her that she had done really well and would likely get an ME - probably a large ME... And then we are three managers who calibrate all the professionals in our areas in relation to their salary and performance. And in relation to the rating the organisation gets. That determines how many we can give ME, EE and O. For example, this year we should reach an average of 3.7, within some predefined [personnel type] clusters... For the academics that will be within the functional area so that there are enough people. So we then have to reach a certain distribution of how many can be in each rating category and a certain overall one-decimal average.

To accommodate real employee input, this area of R&D has changed the prior process so that employee self-evaluation and manager-employee appraisal meeting is conducted before final calibration. The upside of this is that employees actually can affect their own rating if they come up with significant new information in their self-appraisal. The downside is that the manager will not be able to communicate a final rating at the appraisal meeting and that final calibration has to be conducted relatively late. Still, it leaves managers with a difficult puzzle because restrictions on distribution and a predefined rating average with one-decimal precision significantly limits the ability of managers to make changes to their rating distribution. In practice, if one employee through her self-appraisal notes convinces her manager to get an EE instead of an ME, the manager accordingly has to identify an EE that should be moved to ME. Otherwise the rating average and the quota distribution will deviate

from the predefined targets and the manager will have a hard time getting acceptance of that.

Figure 9-3 below illustrates the 2011 calibration process in the area of R&D I investigated.

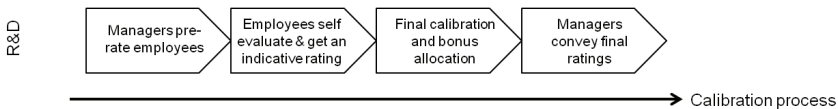


Figure 9-3: The calibration process in R&D

No matter if calibration is before or after the employees' self-appraisal and the final appraisal meeting between manager and employee, at some stage the manager has to convey the message to the employee. Since we have already seen how one purpose of the calibration process is to change ratings so that a certain rating distribution is achieved, there will be cases where managers have to convey a rating which is different than what they had suggested before calibration.

As we would expect, particularly the communication of disappointing ratings is a challenge to managers:

R&D Manager: "You could argue that you could write comments that described really well how good it was to get a 3. You could do that but the issue is that when people get that 3 they stop listening to what else is said. Then it doesn't matter any longer. That message is not taken in by the employee and that is really a shame."

Thanks to normal distribution governance and the calibration process, most employees will get a 3.⁶² We have previously seen how senior management and HR struggle to persuade employees that 3 is a good rating. Nevertheless, in the experience of the above R&D manager, from the moment the manager communicates a 3-rating, the employee stops listening to what the manager says because of sheer disappointment. Communicating the rating becomes a barrier to conveying the attached story that was supposed to provide meaning to the employee. From a sensemaking perspective this is as we would expect employees to act because people generally make sense of actions (the manager's provision of a 3) and start immediately to extract cues from that. If they perceive 3 to be a

⁶² A 3 corresponds to an ME on the five point rating scale.

disappointing rating, the managers story about 3 being actually good, is not meaningful. Therefore, the employees will create alternative stories that are more meaningful. For example, that the manager is biased and the rating unfair.

Particularly for low-end ratings, this phenomenon of unilateral focus on the rating can be problematic, according to one of the R&D managers I interviewed:

If we are in the low-end performance category, then I would like to put some words on why things are not satisfactory. I don't want to do that with a number. I think that is unworthy, you can't use it for anything. I would prefer to put words on it and give the employee some ideas to which direction the employee could move, or discuss what she wants, or discuss the reasons. Really, it could be about some expectations that weren't set from start. So if the employee says, 'but I thought that...' then that is what we should focus on. We shouldn't focus on a number.

The managers' role as someone providing sensegiving in cases of low ratings is a very demanding one. We know that measurement is generally very difficult (Austin, 1996), ratings are heavily biased (Stark and Poppler, 2009; Latham et al. (2008); Cardy et al. (1998); Lee and Son (1998); Varma et al. (2005); Cook (1995) and many more), PAs are about managers' rather than employees' perception of events (Coens and Jenkins, 2002) and that the dialogue between manager and employee is thwarted by the very existence of an appraisal scheme (Waite and Stites-Doe, 2000). So it is not surprising that managers have a hard time when they have to convey disappointing ratings. In cases of low performance, this R&D manager would prefer to do without the rating, because a rating in the form of a single number is an "unworthy" way of communicating something that may be quite complex - and may even be caused by misinterpretations of which the manager herself is partly responsible, e.g. lack of mutual expectation setting. So, in light of this, it is less surprising to hear one R&D manager's overall perception of the value of the rating of employees when I asked her if she thought that PharmaComp would be better off abandoning the use of rating altogether:

Actually, yes. And the direction we have been moving in my part of the organisation the past couple of years, with decimals in ratings - I can't stand that... There can be someone to whom I say they are doing really well and get EE. But also they could be even better. But some think 'I don't need to do any more than this'. And that was not the point of it all. The point was to say that the employee has a great potential and that maybe we should set even more ambitious targets [next year].

The controlled inclusion of employee influence in the R&D approach to calibration matches R&D's inclination towards the management by delegation regime. The lack of trust in an objective and quantifiable target-performance-evaluation causality in a

setting characterised with high degree of specialisation, changing tasks and high task complexity makes it important to be able to include employees' self-assessment into the calibration process. On the other hand, the very mathematical approach to calibration in the area of R&D I investigated, with a one digit ranking of all employees and strict quotas and averages that must be realised, is somewhat contradictory to what we would expect from a workforce with a limited fit to a management by measurement scheme. One possible explanation could be that this aspect of the calibration process is driven more from the practical need of differentiating for purposes of bonus allocation. Another supplementing explanation could be the fact that R&D in previous years has been criticised for being too lenient with PA ratings compared to other workforces. The stricter governance can in this light be seen as R&D senior management's attempt to install tighter control of the rating distributions to accommodate that critique.

Meaning creation through management negotiation in Staffs and Production

In the areas of Staffs and Production I investigated, the calibration process was taking place long before the employees were requested to make their self-evaluation. In effect, the rating calibration process was conducted without including employee self-ratings. The difference in the process steps sequence compared to R&D is illustrated in Figure 9-4 below.

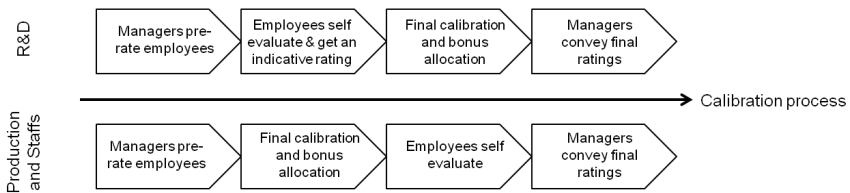


Figure 9-4: Two different calibration processes across workforces

I asked a manager in Production if the final rating was established so early that no matter what positive surprises the manager got when reading the employees' self-appraisals it could have no impact on the rating:

Yes. And you could say that isn't really fair. But the bonus pools need to be distributed and I think we did that already in November. So already then we had to look at how we overall rated our employees. And then we have the [PA] meetings in

9. Meaning Creation in Final Appraisal

January and February... So we have kind of cheated a little. I don't know if the employees know this but it has always been like that.

In this manager's area, calibration and final rating is taking place in November to allow bonus allocation. Still, in January and February, all employees are requested to conduct a self-evaluation which serves as input to one-to-one meetings between employees and their managers. According to the official process guideline, it is on the basis of this self-evaluation and the manager's own perception based on collection of all relevant input, that the manager should propose a final rating that should then go into the calibration process. But in reality, the process is turned upside down; calibration and rating decision comes first - self-evaluation and manager-employee dialogue comes after, when there is no longer any possibility of changing ratings.

It is tempting to compare the process of discussions between managers and employees' regarding self-appraisals, as a garbage can exercise. Decisions are already taken and the purpose of the manager-employee discussion is to establish stories that can make the decisions meaningful to the affected actors. Again, we see how sensemaking is a retrospective activity related to past actions and decisions, although in this case the employees are not necessarily aware of this.

As we saw, the manager quoted above mentioned that she was unaware whether the employees know about the twist of process timing and sequence. If it was important to management that employees knew this, it would be fairly easy to tell. However, it would of course be a risk that this would undermine the employees' engagement in the self-evaluation and manager-employee dialogue about rating. If employees were aware that their self-evaluation had absolutely no impact on their rating, why should they bother to engage into that activity? This risk was exactly what one of the managers I interviewed in Staffs referred to when I asked her if she was telling her employees how the calibration process worked:

No. Their motivation to go in and make a good self-assessment would be very small if they knew we had already finalised the ratings in December.

Employees I interviewed generally had very little knowledge about what was actually going on in the calibration process. However, at least one had an impression and was not impressed about the meaningfulness of the employee inclusion into the process through self-appraisal:

I believe it is a show trial. You have a system to justify something that is in reality decided in a completely different way.

If employee self-evaluation is not included into the calibration process in Staffs and Production, what then goes on in the process? I asked a Production manager if she had to rate her employees in accordance with a specific distribution and how this was connected to the calibration process:

Yes, give and take. And then they discuss it one step higher in the organisation where they decide if someone needs to be moved; if there are too many EEs or if there is too few in AE... I am not fully supportive of that process. I mean, I understand that we have to make sure that we do not give an EE in one unit for something that would have corresponded to an AE in another unit. But I don't like the fact that we are not free to give the rating we think is right and that we maybe end up by giving a rating that we don't think is right.

We see that this manager understands the alignment rationale behind the expected distribution and the calibration process but at the same time feels uncomfortable with the fact that senior management is deciding if she needs to move people from one category to another. Forced distribution scenarios where senior management through calibration across units make changes to the managers' own rating propositions based on "ideal distribution" schemas, is challenging this managers meaning creation because it ultimately forces her to provide ratings that she does not think is right. Conflicting cue extractions are at play; one cue is the governance fulfilment cue that has to do with the managers' responsibility to live up to procedures and instructions from senior management. Another cue is that the manager feels it is her responsibility to provide fair and individual ratings based on the employees' performance against the targets mutually signed in the beginning of the year, i.e. to take responsibility as direct line manager and provide ratings perceived as "*right*" by the manager herself.

Interestingly, the manager extracts both cues and they both have a direct relation to the manager's role identity since they support different ways to reproduce the manager role. How can the manager in a meaningful way maintain her role identity when dealing with these two contradicting demands? One way would be to go into the calibration process by taking the role identity of advocate for her employees:

Staffs Manager: "You become an advocate for your employees. And after a few years you learn to be well prepared for those meetings. The better argument you provide, the less your ratings will be challenged."

When this manager retrospectively reflects on her experience with calibration meetings she sees her role identity as being an advocate for her own employees. An advocate is characterised by being someone who puts a case on someone else's behalf (Oxford Dictionary Online). That is, the manager sees it as her objective to optimise the outcome of the calibration, not in terms of reaching the fairest possible distribution across the whole business area, but in terms of getting the best possible ratings for her own employees. And as any good advocate would subscribe to, it is important to be well prepared and build a strong case, because the stronger the case, the "*less your ratings will be challenged*". A successful advocate is one who is well prepared and who wins her cases.

However, making sense through maintaining an advocate role identity also has its challenges. A Staffs manager explained how she would prefer the management team had a common aim of establishing the most objectively correct distribution of ratings by working as a team, instead of being a group of individuals who promoted their own employees:

There is a tendency that VPs and CVPs⁶³ push this forced distribution so I feel there is a tendency that it [the calibration] becomes a trench war where we all sit and defend our own employees' performance, and VPs challenge us. I would prefer we could do it more together as a management team.

According to this manager, the practice of rating calibration has become a practice of individual optimisation rather than a practice of sharing and alignment. The management team does not work as a team when they calibrate; they compete against each other about favourable outcomes in the shape of large shares of good ratings, relative to their manager colleagues. In the analysis of one Staffs manager, this phenomenon is directly related to the introduction of distribution enforcement:

We go through each employee one by one. For some, we get challenged on an individual employee. And then you get challenged on the distribution. My manager calculates the distribution in the spreadsheet and then she says that 'this is not good enough, I will get pushed back higher in the system'. And then we have to get back and adjust.

In this quote we see how the success criterion for the senior manager at the calibration meeting is to reach the predefined distribution. Ratings are discussed for each individual, but it is all the time with an eye on how the distribution looks. An

⁶³ VP=Vice President. CVP = Corporate Vice President (one hierarchical level above VP)

artificial scarcity of higher ratings has been introduced and this installs competition among managers to bring back to their departments the biggest possible share of the scarce resource. The introduction of distribution enforcement is an important change to the environment of PA. These managers extract a competition oriented cue; my manager colleagues are my competitors in the race for higher ratings. They enact their environment by updating their role identity so that they, in relation to rating calibration, substitute the role of management team member with the role of employee advocate in a race for taking the largest possible share of the scarce resource 'top ratings'.

Ideally, to work well the calibration process requires that over-performing units are allowed to deviate positively from the predefined average and distribution, while lower performing units should then compensate. A Production manager explained her view on this when I asked her if she would be allowed to break the expected distribution in case her unit had over-performed:

I hope I would be able to. If one day half the department had performed at EE, I would like to think it would be possible to give them those ratings. But after having been part of the process, I doubt it.

In this quote we sense some disillusion caused by the experiences the manager has gained through being part of the calibration process for years. Action in the form of past experience is a more important source of cue extraction for sensemaking than is proclaimed policies and procedures. If previous experience shows that for many reasons it is hard to get through with a general average increase in a unit, despite the unit over-performing, then it becomes more meaningful to see the distribution and the rating average guideline as more than guidelines. Rather, it will be meaningful to see them as a forced distribution schema, for all practical purposes.

One manager was very direct in her translation of forced distribution schemes in connection to calibration:

Staffs Manager: "I am not so happy about that forced distribution which I think is the underlying theme. I would prefer you could be challenged on the ratings you give by someone who actually knows your people and their performance. Instead we have these calibration meetings where CVPs and VPs participate. They honestly don't know anything about the people. And they sit and have an opinion on whether the employees are 3 or 4, right. That is probably what I would be most ashamed about, if the employees knew that."

This manager elaborates on the element of being challenged on her ratings. In the current setup it is senior management who challenge the managers and they do it while aiming to realise a specific distribution of ratings rather than with outset in tangible knowledge about the performance of the people whose ratings are discussed. The manager perceives the better objective of the rating process as to provide the most correct rating to each individual in her department; hence any qualified input from other people who are familiar with the actual performance of the employees would be welcomed. Senior management however, in this manager's perception, has a different objective; to realise a specific rating distribution across all areas, even at the cost of compromising on the provision of correct ratings at the individual level. For this manager, this contradiction of objectives is clearly a challenge to the meaningfulness of the entire calibration process, even to the degree that the manager would be ashamed if the employees were aware of the real process.

In a previous chapter we saw how ratings are typically centred in the middle, with 89% of all ratings on ME and EE in 2010. A Staffs manager had told me that she only had 3s and 4s in her department in the latest rating cycle. I asked if her distribution was not very narrowly centred in the middle:

Yes it is. But it is a negotiation game, you know. I said that I didn't have any 2s but not any 5s either. Then it was okay. Well, I am not too happy about those directives.

Bacal (1999) called this preference for primarily using the middle ratings the "central tendency of rating bias". He primarily attributed this bias to managers' preference for avoiding conflicts over low ratings and envy among employees over high ratings. This manager offers an alternative explanation to the central tendency phenomenon; she refrains from proposing any 5s because then it is easier for her to avoid giving any 2s without compromising predefined boundaries of acceptable rating averages. Not only are managers advocates, they are negotiating advocates. A good negotiator is someone who excels in the give-and-take of a negotiation situation, i.e. someone who is good at identifying what she can give the counterpart (i.e. senior management) that is valuable to that counterpart (e.g. conformity to average rating requirements) and in that transaction get something herself that is more valuable to her (avoid having to provide any 2s).

We have now seen how the central tendency can be caused both by managers' reluctance to conflict and by managers' negotiation strategies. However, a manager in Staffs had a third alternative explanation to the phenomenon:

I only had 3s and 4s. We only had one 5 in the entire VP area. We have a lot of people who are highly competent, but we try to make it very hard. So that up there, you have to be really outstanding... We have many challenges to keep within the EE quota. And when you have a lot in that group, then the question is what it takes to differentiate what is required [to get a 5]. So then we have to all completely agree. There are many who do a great job within their area but who are not visible to the rest [of the management team]. Then it is difficult.

According to this manager, the entire management team have to agree for anyone to get a 5. And since the performance of most employees is only visible to a fraction of the managers in an area, this makes it very hard to pass this 'visibility' criterion. The manager mentions that the area has many high performing employees but also that the area has "*many challenges to keep within the EE quota*". From this we can deduct that management in the calibration process has to concentrate on limiting the number of people in the top rating categories to be able to satisfy senior management's expectations to average rating in the area. Focusing on identifying possible 3s in the 4-group is a different and somewhat contradicting exercise to identifying 5-candidates in the 4-group. And if any further 5s were identified, the logic of the distribution governance would dictate that the management team would have to identify even more 3-candidates among the 4-group, as if the task was not difficult enough already. Hence, it is meaningful to this manager and her colleagues to make it very hard to get a 5 - even to the degree where only one employee across the entire VP area got the highest rating.

On the other hand, if a manager does have employees that really are outstanding performers, one should think that managers who subscribe to the role of advocate in the calibration process would try to promote 5s. I asked a Staffs manager if she had any potential 5s in her area that she was expecting to put forward in the calibration process.

Yes, I think so. It depends on what I can get through with when facing that forced distribution. But if I could pick, I would have a couple of them. But realistically, you have to be lucky to be able to give one.

Clearly, there is a discrepancy between on the one side what this manager sees as meaningful in terms of relation between rating and performance, and on the other side what the manager expects will be the outcome of the calibration process.

A fourth factor that may affect the central tendency of ratings and the avoidance of using higher ratings, is when senior management 'punishes' an entire area for the performance of individuals within that area. In Production, one of the managers I interviewed experienced that her management team was required by senior management to get a lower rating average because a high priority area target had been missed. The manager explained what consequences this had:

One of the things that went wrong in my area this year is that we have been rated very hard as an area because of some very small things that have then affected the entire process... It affects the rating average for the management team and it affects the bonus pool for everybody.

The missed area target means that the management team will generally be rated lower and the bonus pool for the entire area will be smaller than would have been the case without the missed target. However, very few of the managers and employees were actually involved in the activities that led to the missed area target, but the entire area is nevertheless affected. This kind of effect is what Kohn (1999) calls the "collective punishment" inherent in any appraisal scheme operating with team goals. As this manager explains, the lower rating firstly accounts for the management team itself. So in theory, the management team could decide to have an unchanged distribution of ratings among the employees. But since the bonus pool for the area had also been reduced due to the missed target, this would put the management team in another dilemma; either they should give less EEs, or the bonus size corresponding to an EE would be diluted because the bonus pool is now smaller.

When managers at calibration meetings suggest ratings for their employees they do it with outset in the targets they mutually have signed, the inputs the manager has obtained from people working with the employee and by the managers' own impressions. Thus, the rating they suggest is meaningful to the manager based on all this information. When the outcome of the calibration process then is that the manager has to change one or more of the ratings she has suggested, this clearly has the potential to be a challenge to the meaningfulness of the rating for the manager. I asked a manager in Staffs if she had to change any ratings due to the 2011 calibration:

Yes, I had one. It was a bit of a surprise to me. I had given her EE, but I was told she should have an ME. It had been decided higher up... I thought that was really a shame. And I was frustrated that I was not even included in the discussion about why she had to be moved down.

The manager is surprised because the employee in her perception clearly is a high performer that deserves an EE. Therefore it is frustrating for the manager to see that not only is the rating changed by her superiors, it is done without any involvement of the manager herself. The rating becomes less meaningful to the manager which is a bad starting point for the manager's role as a sensegiver to the employee, when the rating is to be conveyed.

Although in most cases ratings changed through calibration will be of a downward nature, it is not always so. One of the managers I interviewed in Production had experienced the opposite:

When you have rated your employees yourself and you then get the message that one of them must be lifted to a higher rating, and you don't really feel that anyone of them would deserve that; then I have a hard time with the process... It is a good question why they force this on us managers. That's a question I can't answer right now.

This manager got the feedback from the calibration process that she should change a rating upwards in her unit. This message is reflecting a 'quota' approach to ratings, where the objective of realising certain predefined distributions and averages will be realised by moving around certain quotas of ratings within different categories until the puzzle is finished. As it appears from the quote the manager is free to decide whose rating is to be changed. But since she feels the ratings she already has provided were meaningful to her, she is not able to change a rating upwards without experiencing a loss of meaning. A Staffs manager supplements this quota realisation element of the calibration process:

Sometimes they change a specific person's rating. But typically, you have two 4s and get the message that you can only have one. It is mostly taking the rating down. We have no less than 3-4 iterations in a VP area.

We see here how the calibration process is a puzzle that requires multiple rounds of iterations before the desired distribution is realised. This is not surprising, since senior management in an area needs to collect all inputs from departments in each round of changes before they can calculate the overall distribution and average. And then the area itself is calibrated with other areas and so forth up through the system. And for each change up in the system, this has to be communicated back downwards in the organisation and implemented as rating changes. Not until the distribution and average is acceptable at all levels of the organisation can the iterations settle on an end result.

Calibration in production and Staffs has above been analysed under the same label; meaning creation through management negotiation. We have previously seen how both Production and Staffs had a mixed relation to the management by objectives regime of PA, compared to the more extreme contrasting cases of R&D and Sales. We have also seen how instrumental arguments about the need for early provisioning of rating for bonus allocation purposes has caused employee self-appraisal to be literally irrelevant in the rating and calibration process in these two workforces. Rather than including employees' self-appraisals in the evaluation process, managers take the role as negotiating advocates for their employees through calibration, so that their main objective becomes to optimise their calibration result in the shape of a lion's share of better ratings, as opposed to establishing a fair and transparent rating distribution across units in an area.

Figure 9-5 below illustrates the two different ways through the meaning creation feedback model, which I have described above.

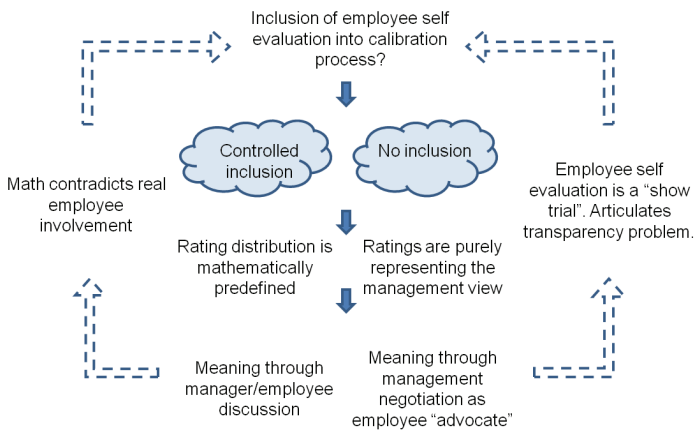


Figure 9-5: Inclusion or non-inclusion of self-evaluations in calibration

The left side of Figure 9-5 depicts the R&D approach, where employees' self-appraisal was included in the calibration process. In R&D we saw how the mathematical distribution of ratings through a one-decimal forced ranking schema with strict rating category quotas and predefined one-decimal rating average requirements was challenging the meaning schemes of managers. However, the inclusion of employee self-appraisals had a moderating effect around which new

meaning could be created. Nevertheless, the rigidity of the design made it very difficult for managers to make anything but minor adjustments from the self-appraisal dialogue, so that the tightness of distribution policies partly contradicted real employee influence.

The right side of Figure 9-5 depicts the process in Staffs and Production, where employee self-appraisal was conducted after the calibration had taken place. Here, employees' self-appraisals are in calibration terms meaningless to the managers because ratings purely represent the management view. Meaning for managers is created around how they perform the role as negotiating advocates for their employees and through that process optimise the outcome for their own unit, in the shape of a lion's share of the scarce resource: high-end ratings. Obviously, to the degree employees are aware, the exclusion of self-appraisals supports a perception of their involvement in PA as being a "show trial" where the outcome has been defined long before they have a chance to provide their view on things. However, to many employees the calibration process is less than transparent, which is in itself another problem in PA articulated by the Staffs and Production approach.

10 Conclusions and Implications

Before summarising the conclusions of the present study, let us revisit the research question one last time:

How is managers' and employees' meaning creation affected by problems in PA and what patterns of local adaption and meaning creation can be observed across organisational workforces?

I have investigated how managers and employees across four workforces in a large private organisation were challenged by problems in PA. I have also uncovered patterns of local adaption to the standard PA scheme and connected meaning creation patterns in the shape of certain stories that make the standard scheme become meaningful in different ways across the workforces.

Ten particular challenges across the three phases of PA were analysed. The ten challenges are listed in Table 10-1 below.

Target setting	Midyear review	Final appraisal
Establishing commitment through target setting	Conveying meaningful feedback	Evaluating business performance
Handling behaviour in target setting	Using rating in feedback	Evaluating behaviour
	Carelessness	Rating distribution policy
	Disruption	Calibration

Table 10-1: Ten challenges across the three phases of PA

For each challenge, different patterns of local adaption and stories were identified for how managers and employees go through a meaning creation feedback process while dealing with challenges in PA.

It has been argued that not only is there a difference between the workforce translations of how problems in PA should be handled, but also that such differences have a direct impact on how meaning schemes of managers and employees are challenged, how meaning is again created and what feedback processes to reinforcement of current problems in PA or articulation of other problems in PA that take place. Below I will summarise the conclusions drawn from the study, argue why these are important contributions to the literature and discuss implications of the findings for our understanding of problems in PA.

A) There is a paradox in PA

Through a literature review of the past 20 years' research into PA, I have demonstrated how PA in modern organisations is infused with lots of problems and that managers and employees have much reason to be frustrated with the process. However, at the same time PA is one of the most institutionalised features of running big corporations everywhere in the Western world. The PA institution is to a remarkable degree consistently reproduced by business scholars, consultants, HR professionals and senior managers across sectors, industries and organisations.

Over the past decades, more and more subtle problems in PA have been identified, isolated and analysed by scholars without any clear breakthrough in establishing viable solutions to the problems, or in establishing scientifically solid evidence that PA altogether does much to improve organisational performance (Kohn, 1999). Similarly, organisations in practice have continuously striven to optimise their PA processes by working to eliminate individual problems in PA. But so far the most visible effect of these efforts appears to have been that other problems, old or new, seem to surface as soon as the previous problems have been attended to.

There is a paradox between the volumes of problems dealt with by academics and practitioners on the one hand and on the other hand the high degree of institutionalisation of PA in modern organisations. The problems in PA challenge frames of meaning for managers and employees and provoke important meaning creation processes that affect their perception of the PA process. The paradox is the result of adhering to PA despite the lack of ability to resolve these problems.

B) The standard PA scheme is locally adapted across workforces, depending on the fit between the task at hand and two contrasting management regimes

I have been using Austin's (1996) separation between management by delegation versus management by measurement as a typology of differences in the task at hand in the four workforces. I have shown that, although PA is implemented as a one-size-fits-all scheme in the case organisation, comparative findings across workforces demonstrates the existence of important differences in local adaption and meaning creation across workforces.

Particularly Sales and R&D form contrasting cases in terms of characteristics of the task at hand, and characteristics of local adaption of the standard PA scheme. The

present study shows how a standard corporate PA scheme can be twisted in different directions in different workforces to such an extent that in R&D it can even be argued that the scheme in some ways has more similarities with management by delegation than with management by measurement.

We see this in R&D for example in letting go of tight control in target setting and by, as the only workforce, incorporating employee self-evaluations into the calibration process. Of the four workforces examined, R&D was found to have the best fit to a management by delegation regime. And in some cases R&D succeeds to twist what is in essence a management by measurement regime towards its opposite; a management by delegation regime, which managers and employees in R&D are better able to find meaningful in their local context.

Table 10-2 below summarises the patterns of local adaption and meaning creation across the four workforces and the ten challenges in PA. The table illustrates the significant differences in local adaption, i.e. variations in how the standard PA system is practiced across workforces. For example, Sales, with the best fit to management by measurement, distributes targets top-down, while R&D, with the worst fit to management by measurement, involves employees in target setting.

In Midyear, Sales managers are taking an accountant role in cases of manager disruption, while Staffs and Production managers take the role as mediators. The advantage of favouring quantity in Sales is that a disruption in manager continuity becomes less of a problem. Just like a company can change its accountant without changing the valuation of its assets, the new manager can look up the numbers and provide feedback to the employee at midyear.

In final appraisal, Sales operates an absolute rating scale without team calibration while the other workforces operate varying principles of relative rating scales with calibration. The use of the absolute rating scheme in Sales matches the fit between the task at hand in Sales and the management by measurement paradigm in this workforce. Being the workforce most inclined towards quantifiable, unquestionable hard facts as the basis of targets and target evaluations, Sales will be able to maintain the clearest link between target, performance and rating within an absolute scheme because here ratings will not be 'polluted' by the moderating effect of the performance of peers.

10. Conclusions and Implications

On the other hand, in the three other workforces where the fit to management by measurement was less evident, there is a weaker case for an absolute rating regime. In addition, and very importantly, the missing link to bonus allocation in Sales make it practically possible to maintain an absolute rating regime, while the opposite is the case in the three other workforces, where differentiation of employees is required to be able to distribute reward. Further, in Staffs and Production employee self-appraisals are not included in the calibration, while they are in R&D.

Workforce	Fit to management by measurement	Patterns of local adaption and meaning creation		
		Target setting	Midyear	Final appraisal
Sales	High	Targets are <i>distributed</i> Behavioural impact is <i>concealed</i>	No preliminary rating due to <i>not being required in standard process</i> Acquiescence to employee carelessness through <i>creative process redesign</i> Management disruption handled as <i>accountant</i>	Behaviour is <i>excluded</i> OR <i>included</i> Rating on <i>absolute scale</i> <i>Non-calibration</i>
Staffs	Mixed	Target commitment through <i>manipulation</i> Behaviour is <i>relocated</i> to IDP	Conveyance through <i>distancing</i> No preliminary rating due to a) <i>irrelevant</i> , b) <i>generate conflict</i> , c) <i>beyond managers' control</i> Acquiescence to employee carelessness due to <i>governance fulfilment</i> Management disruption handled as <i>mediator</i>	Behaviour is <i>discretely included</i> Rating on <i>relative scale</i> Employee self evaluations excluded, manager as <i>negotiating advocate</i> in calibration
Production	Mixed	Target commitment through <i>mixed</i> approach Behaviour is incorporated through <i>adaption</i> of standard targets	No preliminary rating due to a) <i>irrelevant</i> , b) <i>generate conflict</i> , c) <i>beyond managers' control</i> Management disruption handled as <i>mediator</i>	Behaviour is <i>at the front of the stage</i> Rating on <i>relative scale</i> Employee self evaluations excluded, manager as <i>negotiating advocate</i> in calibration
R&D	Low	Targets committed through <i>involvement</i> Behaviour <i>taken for granted</i>	Conveyance through <i>enlargement</i> No preliminary rating due to <i>negative effect on motivation</i> Acquiescence to employee carelessness to <i>avoid conflict</i>	Behaviour is <i>discretely included</i> Rating on <i>relative scale</i> <i>Controlled inclusion</i> of employee self evaluations

Table 10-2: Summary of patterns of local adaption and meaning creation

These and the other identified differences between the workforces are not trivial. They form fundamentally different approaches to how to establish targets, provide

feedback and evaluate and calibrate performance ratings. Uncovering such differences is important, because the overwhelmingly dominant approach in the literature on PA is to study large organisations as if they were single entities,⁶⁴ while in reality they consist of highly different sub-entities, each adapting standard systems differently in their local contexts, depending among other things upon the task at hand and the fit to the management by measurement regime.

By reaching a better understanding of this process of local adaption of a standard PA system, I have also added to the understanding of the paradox in PA. In theoretical terms, local adaption means that particular local practices are established, supported by local meaning creation through stories that make the notion of PA more meaningful in local contexts. Thus, while PA schemes are ubiquitous in modern organisations and embedded with multitudes of problems, local adaption of PA is an important way for practitioners to establish local meaningful practices and stories about PA that counter the widely experienced challenges in standardised PA schemes.

This is also an important finding in a broader theoretical and practice context because the present study of PA is just one example of how one through the study of meaning creation processes can analyse local adaption of standardised management systems. While the present study illustrates how differences in the task at hand can be an important explanatory factor to understand significant differences in local adaption of a standard corporate PA system, there is reason to believe that similar differences would surface if one studied other standard management systems the same way.⁶⁵

C) Meaning creation is an important key to understanding PA in practice

In contrast to what one might have assumed the results show that the PA midyear review does not require hard facts or detailed documentation of performance events to be meaningful. In fact, as examples have shown, it sometimes appears to be

⁶⁴ For example Armstrong (2005, 2006) and Lazear & Gibbs (2009). Sometimes the approach is founded on an even more aggregate level than the company level. For example Shibata (2002) contrasting PA practices in US and Japan, and Amba-Rao et al. (2000), who contrasts foreign and national firm PA practices in India.

⁶⁵ Still within HR, other examples of management systems being often dealt with at firm level, where inclusion of the local adaption perspective could be an interesting new approach, could be talent management (e.g. Capelli, 2008), people attraction and recruitment (e.g. Taylor & Collins, 2000), succession management (e.g. Fulmer & Conger, 2004) and the concept of strategic HR through installation of an HR partner concept (e.g. Ulrich 1997, 2005).

easier to create meaning from the PA process without detailed preparations. For the manager this was because maintaining the feedback dialogue on a less tangible level opened opportunities for balanced articulation, softer interpretations and less conflict. For some employees, it was because investing time and effort in a process that one feels alienated from and disengaged in, would not be meaningful.

In a sensemaking perspective the missing need for hard facts or documentation is not surprising, because in sensemaking plausibility is of higher importance than accuracy.⁶⁶ But in a traditional management by measurement PA perspective it is an important finding because it moderates the importance of many of the measurement and bias problems in PA. If accuracy is not necessary, lack of accuracy is less of a problem. And when accuracy is not important, the value of measuring and of preparing thoroughly for PA meetings by taking detailed notes on an on-going basis, becomes less meaningful.

PA schemes are not what they are designed to be - they are what actors make them become. Even though people work within the same company and under a global corporate PA scheme, different organisational layers, management teams and manager-employee dyads will exploit available options to add and subtract bits and pieces so that the PA scheme becomes more meaningful to themselves. By this, they will reproduce different environments that can again be enacted by actors for the creation of different interpretations of what is going on in PA.

This is important for our understanding of the problems in PA because it demonstrates that problems in PA are highly contextual.⁶⁷ What is a problem-derived challenge to actors' meaning creation in one context may be fully meaningful to other actors in other contexts. Without including meaning creation in the context of enactment of environmental frames, analysis of problems in PA risk having limited relevance for the people who are working under PA schemes. The bad news is that meaning creation introduces high contextualisation which at its extreme makes sociological theorisation difficult. However, the good news for our prospects for

⁶⁶ An alternative way to translate the weighting of plausibility rather than accuracy, is the notion of "face" (See e.g. Goffman, 1967). Critical feedback can be threatening to the involved parties. And very tangible, non-negotiable feedback can be very threatening. Non-tangible, less accurate feedback is easier to convey at a lower threat-level, because it opens up the possibility of adjusting the conveyance depending on the response of the counterpart, and also more generally opens up for different translations of what is being conveyed.

⁶⁷ As opposed to the more universal (firm-valid) problem focus of many scholars, e.g. Brown et al. (2010), Poon (2004), Tziner et al. (2001) or Wagner and Goffin (1997).

sociological theorising is that workforce affiliation and task at hand characteristics were found to have a strong explanatory power when looking for patterns in dealing with challenges in PA across an organisation. These structural factors affect local adaption and meaning creation, but are also themselves reproduced through the distinct translations of PA in the different workforces in an ongoing cause and effect interchange similar to what Giddens (1984) has termed a structuration process.

D) Problems, meaning scheme challenges and meaning creation in PA are interlinked to problem re-enforcement or new problem articulation in a constant feedback process

The study demonstrated that the dynamics of different PA problem interchange processes caused by workforce moderation and different enactments can be described and analysed within the framework of a dynamic meaning creation feedback model. For example, we saw how an ambition to align expectations and increase transparency have over time made senior management cascade more centrally defined targets across the organisation. But we also saw how this increasing share of centrally cascaded targets was seen by managers and employees as a symbol of reduced influence on targets leading to reduced commitment. In practice therefore, the top-down problem in PA should not be dealt with in isolation from the problem of transparency in PA (DeNisi and Pritchard, 2006) since, at least in this case, measures to handle the transparency problem (standard targets) have a negative impact on commitment through articulating the top-down problem. This phenomenon became most evident when we contrasted Sales and R&D, where differences in the task at hand produced the biggest difference in fit to management by measurement.

In Sales, the transparency problem has clearly been mitigated by a strong top-down cascading of targets fitting the task at hand in this workforce. In R&D, there is also target cascading, but it is moderated by a much looser interpretation of the negotiability of targets, fitting the task at hand in this workforce. In Sales, meaning is challenged by the lack of employee involvement in target definition. Sales employees created new meaning by alienating themselves from the process which as a feedback loop effect articulated the top-down problem in PA and lack of employee commitment to the targets. In contrast, in the R&D example, the moderation of the

effect of target cascading meant that meaning was less challenged in this case, new meaning creation not apparent and therefore the feedback loop from meaning re-creation was also much weaker. Hence, in the R&D example, the top-down problem in PA was not articulated to the same degree in target setting.

As part of the present study, a meaning creation feedback model was developed to analyse how actors' handling of one problem in PA can articulate another problem in PA in a complex, on-going and dialectical interrelation between perceived problems, actions and reactions of actors, sensemaking of actors and articulation of new problems in PA. Throughout the analysis of meaning creation challenges above, many other examples of these dynamics have been presented.

For example, for the punishment problem in PA, i.e. that any disappointment in a PA system relying on extrinsic rewards will be perceived as punitive by employees and thereby deteriorating for motivation, the analysis has shown that some managers are well aware of this risk and pursue alternative ways to improve behavioural performance in a more legitimate way. Relocating the dialogue about behavioural improvement from the PA frame where things are more serious, more committing and with tougher consequences, to the alternative frame of the individual development plan (IDP), is one meaningful alternative for some managers to get around the punishment problem in PA. But this articulates the transparency problem in PA, because relocating the behaviour dialogue away from PA does not change the fact that behaviour is still an important element in the evaluation and rating of employees.

We see how the relocation of the behaviour dialogue to the IDP frame leaves us with an opaque connection between targets and evaluation so that accuracy will be challenged in this relation. In a sensemaking perspective, this may still make good sense to managers because sensemaking was never about accuracy in the first place; it is much more about plausibility. However, we also saw how senior management across workforces seem to favour handling the transparency and control imperatives over the handling of other problems, since from 2012 all workforces operate with cascaded, standard behavioural targets in PA.

E) Challenges and problems in PA are connected as many-to-many relationships

A key element that kept coming up in the context of establishing commitment in target setting was the top-down problem in PA (Coens and Jenkins, 2002) because lack of involvement generally was experienced by employees and most managers as deteriorating commitment. However, while investigating meaning creation of employees and managers as they confronted challenges to target commitment, we also saw how they made reference to other problems in PA, for example the problem of PA being bureaucratic and control-focused (Piercy, 1994) and the problem that targets only indirectly represent real value contribution (Austin, 1996).

Similarly, we have seen how actors directly or indirectly make reference to multiple problems in PA when they reflect on the other challenges in PA. For example, we saw how the measurability problem in PA (Austin, 1996), the transparency problem in PA (DeNisi and Pritchard, 2006), the punishment problem in PA (Kohn, 1999), the problem of PA being designed in a way that only fits handling the 5% lowest performers (Coens and Jenkins, 2002), the top-down problem in PA (Coens and Jenkins, 2002) and the comparability problem in PA (Coens and Jenkins, 2002), were all brought into play in the analysis of the challenge of including behaviour in target setting. We have seen that when actors face practice challenges it is often not related to one isolated, particular problem in PA. Rather, individuals appear to perceive a mix of problems in PA as intermingled into a more general challenge to their meaning schemes. Hence, to understand problems in PA is not only to understand the cause and effect of each single problem, it is also to understand how multiple problems interact to constitute broader challenges to individuals.

The relation between a practice challenge in PA and problems in PA is a many-to-many relation where challenges can relate to multiple problems and where the same problems can be related to multiple challenges. Practice challenges are generally based on broader categories than single tangible problems in PA; either on an agglomerated perception of problems in PA or on a more diffuse perception of what is the root cause of challenges.

In Table 10-3 below the relation between challenges in target setting and problems in PA is summarised to exemplify the many to many relationship between challenges in

PA and problems in PA. A similar picture could be drawn for the challenges dealt with regarding midyear review and final appraisal. It is important to emphasise that the problems in PA that actors relate to depend on the particular situation in which they find themselves, including the environmental frames they operate within, the actions of other actors and their own particular sensemaking of these elements. The list of relations illustrated in Table 10-3 is therefore not a final list valid for any manager and employee in any PA context. Other actors will in other contexts likely relate to different combinations of problems in PA.

<i>The challenge of...</i>	<i>...related to problems in PA</i>
...establishing commitment through target setting	<ul style="list-style-type: none"> - The top-down problem in PA (Coens and Jenkins, 2002) - The problem of PA being bureaucratic and control focused (Piercy, 1994) - The problem that targets only indirectly represent real value contribution (Austin, 1996)
...handling behaviour in target setting	<ul style="list-style-type: none"> - The lack of measurability problem in PA (Austin, 1996) - The lack of transparency problem in PA (DeNisi and Pritchard, 2006) - The punishment problem in PA (Kohn, 1999) - The problem of PA being designed in a way that only fits the 5% lowest performers (Coens and Jenkins, 2002) - The top-down problem in PA (Coens and Jenkins, 2002) - The comparability problem in PA (Coens and Jenkins, 2002)

Table 10-3: Relations between challenges and problems in PA in target setting

This finding is important because much PA research and many practice PA improvement initiatives aim to handle one or a few problems in PA as if they were of universal importance and as if they could be treated in isolation from other problems in PA. This common approach can be seen as being part of a more general trend in social science where research tend to become narrower and shallower while the problems in practice are rarely narrow by nature (Van de Ven, 2007). My study indicates that the relation between problems in PA and challenges to PA practice as perceived by actors is much more complex than the common approach would indicate. By continuously treating problems in PA as isolated entities we also reduce the effect of improvement efforts.⁶⁸

⁶⁸ This conclusion relates to Grint's (2005) discussion of three different types of problems; tame, wicked and crisis. In this terminology, we could say that PA by decision makers is typically being framed as a tame problem requiring rational analysis and appropriate management process installation to work properly. In opposition, what I have argued is that PA could be framed as a wicked problem, embedded with dilemmas and uncertainty, requiring leadership to ask questions and involve employees rather than dictate solutions.

F) Meaning creation is also role dependent

We have seen how in some cases managers and their direct employees extract different cues from what is going on in PA. For example, we have seen how an employee in Sales was emphasising the positive elements in having behaviour included in PA targets because it would direct the attention to elements of her performance that she found to be important supplements to the narrow sales index realisation. To be meaningful to the employee such targets should be more about human values than about business outcome, and they should include how she worked as a colleague and how she worked with the customers. In contrast, the manager saw it as an advantage to have as few targets as possible and did not find it meaningful to include behavioural targets. Both because these would have to be un-weighted since all weights were already used by "hard" outcome targets on sales and callplan, and because this would just add to the number of targets and thereby make it more difficult for managers and employees to have clear priorities and to stay focused on what is most important.

In other words, the same events and actions can be interpreted in different ways dependent on the role of the observer. Again, if we continue to analyse problems in PA without reflecting on how the characteristics of our empirical object fundamentally affects the outcome of our analysis through differences in meaning creation patterns, we will at best only capture a fraction of what is really going on in PA. At worst, we will completely misinterpret our data. Hence, the fact that managers and employees can elaborate different stories about the same performance events is an important finding because the whole purpose of PA schemes is to create differentiation of employees on an objective, comparable and transparent foundation. But when managers and employees perceive challenges differently, interpret problems in PA differently and enact different and sometimes contrasting stories about what goes on in PA, this is fundamentally questioning the objectivity, comparability and the transparency of the process.

To sum up, the present study has demonstrated that if we want to understand how problems in PA affect people in organisations, we need to pay attention to the complex interrelation between local adaption of standard PA schemes, meaning processes and problems in PA. We need to supplement the current focus on rational analysis of individual problems, with a more non-rational focus on meaning creation

processes and frames of enactment. Particularly, the study has shown how patterns of local adaption and meaning creation can be analysed and understood as dependent on differences in the task at hand within different workforces.

Further, the study has demonstrated that articulation or reinforcement of perceived new or existing problems in PA is affected by the way actors in the first place react to challenges in PA. Without an understanding of these feedback dynamics, attempts to handle problems in PA will only be partial and feed an on-going cyclical exchange of problems in PA, rather than bring hope for more lasting improvements, or alternatively, decision to abandon rating of employees.

By continuing to focus on understanding and solving single problems in PA as if they could be isolated from their contexts and dealt with independently, we likely will, indirectly, exaggerate other problems in PA because we have neglected the complex interrelatedness of problems in PA. This also forms an explanation as to why, despite decades of research and practice experience with PA schemes, both scholars and practitioners still face the same problems in PA as they did twenty years ago.

11 Qualifications and Guidance for Future Research

A general reservation against the findings in this study could be directed towards the study design. How can I claim any generalisable conclusions from a mainly qualitatively based study of a few selected managers and employees in the Scandinavian branch of a single organisation? The short answer to this is that I have not claimed any kind of generalisation of specific patterns across workforces to be valid for larger populations, neither outside nor inside PharmaComp. I am fully aware that other managers and employees in other organisations or units could and certainly would have provided different angles on local adaption of PA and on how problems in PA are challenging schemes. In fact, that is exactly one of my main points; that the single dimensional way problems in PA are too often dealt with in research is missing the important dynamics of reality by operating at an aggregate level across entire populations or entire organisations, as if local practices, cultures and meaning creation patterns were not important. I claim they are and that much more attention should be directed towards the dynamics of those processes if we are aiming for increased understanding of how problems in PA affect people and organisational performance. Therefore I would encourage scholars to dig more into the patterns of local adaption of standard PA systems and interrelations between problems, meaning creation and feedback dynamics, either by supplementing this study's focus on workforces, or by focusing on patterns of meaning creation across other intra-organisational demarcation lines like e.g. geography, national culture or centre-periphery distance. Or, by looking at inter-organisational or even inter-industry level particularities that affect patterns of local adaption and meaning creation in PA.

Furthermore, for each of the main conclusions in this study, more specific further reservations can be made. Some possible reservations are dealt with below.

First, I have claimed there is a paradox in PA between the ubiquitous and institutionalised presence of PA in large organisations and the vast, seemingly unsolvable number of problems identified in PA. Critics may question whether that is not the case for any popular management technology; that it will have caught the interest of scholars who will by definition always have a problem focus. So that after some decades of being in use, any standard management technology would have a similarly high number of critical studies attached to it.

However, such critique would miss two important points. First, the possible existence of similar paradoxes within other standard management technologies does not affect the existence of a paradox in PA. Rather, it should make us become interested in whether standard management technologies in a broader sense possess built-in paradoxes because they ignore the need for local adaption and meaning creation and therefore provoke problems for people working within the schemes. Second, the existence of a significant number of scholarly investigated problems in PA is not unrelated to the practice of PA in organisations. The problems are not just academic constructs; they are real and tangible challenges to managers and employees in their everyday life. People operate every day within PA schemes that are embedded with multitudes of practice problems connected among other things to the rational economic assumptions and extrinsic control focus of the schemes, and people are, put simply, not happy about it. Still, organisations insist on maintaining the schemes. To me, that is an interesting paradox.

Second, I have claimed that local adaption and meaning creation differs across workforces depending on the fit between the task at hand and two contrasting management regimes; management by measurement and management by delegation. Critics may question how I envision that managers in practice should be able to use this knowledge to deal with problems in PA? If meaning creation processes are so strongly affected by the way they are situated under specific conditions, how can any general management advice be put forward that can be understood and implemented in practice?

To this critique I would respond that I do not see much value in general PA cookbooks of which there are so many, with single dimensional, recipe-like responses to single problems in PA, when analysis has just demonstrated how the task at hand in different workforces heavily influences meaning creation processes, management and employee responses and problem re-enforcement or articulation. A recipe advice like "if you face problem A, then do B which will make C happen and you will be good" may be simple and communicable. But in reality there is a good chance that D, E, F or even all of them, is just as likely outcomes of doing B, as C is, depending on the nature of the enactment frames and meaning creation processes at play in the situation. The effects of doing B are dependent on where and how affected people are situated in terms of frames for enactment. Therefore, I would

again encourage researchers to dig much more into how local adaption and patterns of meaning creation can be investigated at a sociological level, because this will in turn provide a much more qualified understanding of what management measures will be relevant to activate to handle challenges to meaning creation in PA in ways more promising than what we have so far seen from the long list of cookbooks on how to install PA effectively in organisations.

One place to start could be to examine other cases to further refine and understand the connection between the categorisation of the task at hand against particularities of local adaption and meaning creation in PA. This could involve introduction of additional explanatory factors into the model. For example, in my study Production and Staffs both relied on mixed positions in relation to the management by measurement regime, but still two different approaches to behavioural targets and evaluation have been selected. This difference in approach is not possible to account for solely on the basis of the categorisation of the task at hand that I conducted. To account for this difference, we would therefore have to include more elements into the analysis. This could for example be differences in workforce culture and history, which may also impact local adaption and meaning creation. In my case we would find that Production in PharmaComp has for years had a more standardised and systematic approach to PA than the other workforces, with more particular policies valid at workforce level.

Third, I claimed that meaning creation is an important key to understanding PA in practice. Critics may point to the task at hand, not meaning creation, as the key to understand differences in workforce adaption of a standard PA scheme. However, I see them as mutually dependent. Differences in the task at hand is a framing condition, whereas meaning creation is the process through which local adaption, embedded in this framing condition, is formed, in the shape of local processes, procedures and stories about what is really going on in PA.

Fourth, I claimed that problems, meaning scheme challenges and meaning creation in PA are interlinked to problem re-enforcement or problem articulation in a constant feedback process. Critics may find that feedback processes to management initiatives are only to be expected - after all, the purpose of management initiatives is to get an organisational reaction. Seen in that light, critics may find that the more important study object would be to establish a framework allowing managers to

predict exactly what feedback process, or reaction, management should expect from a specific type of intervention in PA.

I would agree that such a framework would be valuable to practice, although hardly possible to fully realise. I also believe that my study can be seen as a first step in the direction of getting closer to such understanding. By introducing the importance of local adaption based on a categorisation of the task at hand, this study provides a more complex, but in my opinion also a more realistic, view on how practices of standard PA schemes vary inside organisations. By subscribing to this, managers and scholars will stand a better chance of predicting how certain management initiatives within PA will have different effects within different areas of the organisation, and it will be possible to direct problems in PA in more appropriate ways, although the dilemmas will still be present. However, although I see my study as a first step in this direction, I also recognise that a lot of work is still needed to clarify the circumstances under which management and scholars will become able to predict the effects of specific management initiatives in PA.

Further research into the dynamic patterns of interrelatedness of problems in PA, differences in the task at hand, meaning scheme challenges within particular enactment frames and feedback loops to re-enforcement or articulation of problems in PA, certainly looks to me to be a more promising path to explore new learnings about how PA affects people in organisations than more studies that add to the long list of investigations into single bias or other problems in PA, as if they could be dealt with in isolation and at a general level.

Fifth, I have claimed that my concept of PA challenges and problems in PA are connected as many-to-many relationships. To this, critics may insert that this is because my concept of PA challenge is in itself only a meta-concept used for overlapping categorisation of problems in PA. If this is true, it is no surprise that there is a many to many relationship between meaning scheme challenges and problems in PA.

To this my answer would be that while it is true that my concept of PA challenges can be seen as a meta-level categorisation of problems in PA, this critique misses the point I have made in my analysis. The point is that managers and employees, within the frames of enactment possibilities defined by the task at hand through their

workforce affiliation, perceive challenges in PA not at a single problem level, but at a meta-level. By studying single problems in PA in isolation researchers have ignored what I claim to be true; that problems in PA are not isolated objects that can be dealt with in singularity - they are interrelated and any solution to a problem must take account of that. Hence, I would see interesting perspectives in more research into the interconnectedness of problems in PA.

Sixth, I claimed that meaning creation is also role dependent. Managers and employees, although observing and discussing the same performance events, do not necessarily share meaningful stories about what actually took place. Sensemaking is a hindsight activity allowing post-action rationalisation. We cannot assume that managers and employees share the same stories about what happened and how the individual performed in any given situation. Critics may say that this is neither surprising nor new, because in any evaluation there is bias, and research into PA bias is well developed over the past 20 years.

To this my response would be that I acknowledge all the work done to analyse bias and I also recognise that I have not contributed much to the body of bias research in the present study. Again, this was not the point I have been trying to make. In drawing the conclusion about role dependency I was not thinking about evaluation bias, but about meaning creation of the PA process as such. In one example we saw how a manager worked to increase transparency through minimising the number of targets, while her employee perceived it to be challenging to her meaning creation that more targets were not included. And in another example we saw how a manager saw the PA process as becoming more top-down controlled while her employee saw the process as open to individual involvement. More research into how role perceptions affect meaning creation and attitudes in PA could be beneficial to deepen our understanding of these relationships. And not only manager-employee dyad dynamics could be of interest here. While I have focused on the perspective of managers and employees, also the role impact on perspectives of top management, of HR professionals and even of consultants specialised in providing organisations expert advice on how to implement and improve their PA processes could be interesting to understand at more detail than is available today.

In addition, while looking at role dependency in PA, the question of disruption in management relations would also be an important subject to pursue further. In many

organisations it is common practice that managers change position, or that units are reorganised, sometimes every 2-4 years. In such organisations, manager-employee PA relationship disruption will be more the rule than the exception. And since this study has shown multiple examples of how managers and employees are not documenting everything of significance in the PA system, it becomes important to understand what happens to all the tacit knowledge and the relationship effects when an employee gets a new manager. No matter if it is in the middle of a PA cycle, or at the beginning of a new cycle. I would argue that this is yet another problem in PA that would be worth investing further studies into.

12 Appendices

12.1 Appendix 1: Abstract in English

A thorough literature review shows how it is well established in the literature that performance appraisal (PA) in modern organisations is infused with many problems. However, at the same time PA is one of the most institutionalised features of running large corporations everywhere in the Western world. Problems in PA challenges managers and employees and provoke important local adaption and meaning creation processes across workforces in order to allow a meaningful fit between the practice of PA and the task at hand in each workforce.

Through 34 qualitative interviews with manager-employee dyads, the study design included a comparative study of local adaption and meaning creation in PA across four different workforces over a full one-year PA-cycle in a case organisation where a one-size-fits-all PA scheme had been implemented. Important differences in environmental dynamics and in manager and employee enactment patterns were found. Understanding these patterns is not only important for our understanding of how PA schemes affect individuals' and organisations' performance. It is also more fundamentally important to our understanding of the dynamics of the problems in PA which have been uncovered, analysed and attempted to be solved over the last 20 years' scholarly research into PA, without any real breakthrough.

Austin's (1996) separation between management by delegation versus management by measurement was used as a typology of differences in the task at hand in Sales, R&D, Production and Staffs in the case organisation. Although PA as such is clearly designed to operate within the management by measurement paradigm, comparative findings across workforces show that important differences are present. Particularly, Sales and R&D form contrasting cases in terms of characteristics of the task at hand and characteristics of PA adaption and meaning creation. Thus, the study shows how a standard corporate PA scheme can be twisted in different directions in different workforces to such an extent that in R&D it can even be argued that the scheme in some ways has more similarities with management by delegation than with management by measurement. The study discusses the implications of this and guidance to future research is provided.

12.2 Appendix 2: Abstract in Danish

Et omfattende litteraturstudie viser, at det er veletableret i litteraturen at præstationsevaluering (PA) i moderne organisationer er præget af et stort antal forskelligartede problemer. Samtidig er PA et af de mest institutionaliserede aspekter af hvordan store organisationer drives i den vestlige verden. Problemerne i PA udfordrer ledere og medarbejdere og driver vigtige processer i forhold til lokal adaption og meningsskabelse på tværs af funktionelle områder, for at muliggøre en meningsfuld sammenhæng imellem PA praksis og arbejdsindholdet i forskellige funktioner.

Ved hjælp af 34 kvalitative interviews med par af ledere og medarbejdere understøttede undersøgelsesdesignet et komparativt studie af lokal adaption og meningsskabelse i PA på tværs af fire forskellige funktionsområder over et helt års PA cyklus, i en organisation hvor en standard PA proces er søgt implementeret på en ensartet måde i hele organisationen. Vigtige forskelle i de konkrete omgivelser og i ledere og medarbejders handlingsmønstre blev identificeret. En bedre forståelse af disse mønstre er ikke blot vigtigt for vores forståelse af hvordan PA systemer påvirker individers og organisationers præstationsevne. Det er også mere fundamentalt vigtigt for vores forståelse af dynamikkerne bag de problemer i PA, som har været afdækket, analyseret og forsøgt løst over de sidste 20 års forskning i PA, indtil videre uden entydigt fremskridt.

Austin's (1996) opdeling mellem ledelse ved delegering og ledelse ved måling blev anvendt som en typologi over forskelle i arbejdsindhold i salg, forskning, produktion og stabe i en case organisation. Selv om PA er udviklet indenfor ledelse ved måling paradigmet viste den komparative undersøgelse, at der er vigtige forskelle imellem funktionsområderne. Særligt salg og forskning udgjorde kontrasterende cases i forhold til karakteristika ved arbejdsindhold og karakteristika ved PA adaption og meningsskabelse. Studiet demonstrerer således hvordan en standardiseret tilgang til PA kan blive tilpasset i en sådan grad at man i forskningsområdet i nogle sammenhænge kan se flere ligheder med ledelse ved delegering end med ledelse ved måling. Afhandlingen diskuterer endvidere implikationerne af dette samt giver ideer til hvor yderligere forskning kunne fokusere.

12.3 Appendix 3: Letter to Selected HR Partners

To: The HR partner

Request for assistance with identifying interviewees for PhD research project at CBS - Performance Appraisal in PharmaComp

The research project aims to critically examine the theory and practice of utilizing the process of Performance Appraisal, in PharmaComp known as 3P (People Performance Process).

A preliminary review of the last 15 years research into the subject indicates among other things that there is a significant overweight of research in an Anglo-American context and a heavy empirical focus on managers and staff/white collar workers. Comparative studies across workforces are missing, leaving opportunities to develop deeper insights of any systematic differences between staff, production, sales and R&D workforces.

Thus, the preliminary research question is centred on: How is the practice of performance appraisal perceived across different workforces and what is the perceived meaning, outcome, cost and value to the individual and the organisation from the process?

The research design is a so-called *single-case study* where many different sources of data will be used. A key element is qualitative interviews with pairs of managers and employees from different workforces in the organisation. The managers and employees will participate in multiple individual interviews over a two-year period (2011-2012) in connection to the three yearly major process milestones: Goal setting, midyear review and yearly appraisal.

The participants should be ready to talk openly about their personal experiences and thoughts around the PA process. They will also be asked to take personal notes during the two years, whenever they have reflections on the process. Each interview is expected to last up to 1 hour and will take place at the participant's location. The interviews will for practical reasons be recorded and transcribed. The participants will have the opportunity to review the full transcriptions and correct any

misunderstandings. All participants will be protected by confidentiality which will also be guaranteed in writing in a formal confidentiality declaration letter. It is my hope that the participants will also personally benefit from participation, through reflection and discussion with the researcher.

A qualitative research based on interviews will naturally not aim to reach any statistical representativeness. Instead, the interviewees should ideally be identified on the basis of their likely contribution with new insights or reflections around performance appraisal. To identify such candidates, a dialogue with HR Partners in R&D, Staffs, PS and Sales will be conducted. The interview person identification process will be conducted during November and December 2010. It is the intention to identify two pairs of managers and employees within each of the four mentioned workforces.

The research project has a formal deadline in 2015. Before that, significant findings will be shared with the participants and PharmaComp in accordance with good research practice and with adherence to ethical and confidentiality considerations.

I look forward to hear your feedback on the above request. If you have any questions or comments, please contact me.

Kind regards

Morten Rossing

12.4 Appendix 4: Letters to Participating Managers and Employees

To: Participants in interviews regarding the performance appraisal process in PharmaComp

Declaration of Confidentiality

Thank you for your participation in this research project regarding the 3P process! As part of the research project, interviews with 6-8 pairs of managers and employees will be conducted over a two-year period (2011-12). The interviews will be individual and confidential meetings between the employee or manager and the researcher. The interviews will take place at your location.

Each interview will be recorded and there will be taken notes during the interview. Recordings and notes will be kept electronically and in accordance with good research practice and guidelines from Datatilsynet. This means that:

- Data is kept confidential in relation to the identity of each interviewee to protect the interviewee as a person
- Data is treated with confidentiality, i.e. it is only the involved researchers that will have access to recordings and transcriptions
- Data will be stored safely, meaning that recordings and transcriptions will be stored in a safe, protected way and only in electronic format.

The outcome of the research will be published in a way so that protection of the participants and PharmaComp is respected. The outcome is expected to be published in a PhD thesis and a number of articles in scientific journals or conference papers.

If you have any questions or comments do not hesitate to contact me via phone xxxx xxxx or email: mortenrossing@gmail.com.

Kind regards

Morten Rossing

12.5 Appendix 5: Interview Guides

Below the reader will find the two interview guides for semi-structured interviewing was used during the final appraisal interviews. Similar, though adjusted, guides were made for interviews during Target Setting and Midyear Review.

Interview guide for Managers regarding Final Appraisal

Intro

- Interview within 4 workforces, pairs of mgr/employees, research question
- Confidentiality (delivered doc), Semi-structured interview, Use of recording
- Confirm ok?

Warm-up

- Have you used the "diary" I forwarded after our first interview? Why not? / How did that work out?
- How would you overall describe the recent final appraisal meeting?
- Are you satisfied with the meeting?

Deep dive

- What was happening at the meeting? What went well/bad? Something you would have done differently?
- Rating first/last? What was the overall rating? What was the employee's response to this?
- What is the distribution in your area?
- What will be the impact for the employee?
- How did the rating come about? Can you describe the calibration process that has been used?
- Has the employee been in control of the target outcomes or depending on others? What impact has this had? Do you think the employee sees this the same way?
- Have you got a co-responsibility for the employee's score? Describe.
- Which other biases do you identify in the process you have just been through? (e.g. personal relations, recency/memory, not measuring the right thing, manipulation, lack of info, leniency, central tendency, attribution errors (pos attributed to oneself / neg attributed to others), self-serving bias)
- Are these biases ok and manageable - if yes, how do you manage them?
- Has the process and rating in your view been fair to the employee? To yourself?
- How have you prepared for the meeting? How has the employee prepared?
- Have you collected any input from others? Who?
- What key performance events were you discussing?
- Does the rating relate to results or efforts?

- How did behaviour influence the evaluation? Was this transparent to the employee?
- Were targets removed or updated? Who took initiative to this?
- What do you think your employee thinks of the meeting?
- How would you describe your feelings up to, during and after the meeting?
- How do you imagine the employee was feeling?
- Has the 3P process a motivation effect for the employee? For you? How?
- Is this the best way to achieve that?
- What is your overall view on the way the company and you as a manager do performance management?
- Is 3P a fair process for you and your people? Why/why not?
- Does the calibration process secure an accurate representation of performance across the org?
- Is it a transparent process? Would it be better or worse without the final rating?
- What is the purpose of the 3P process? Is it successful?

Particularly for new managers (manager change)

- How did you make sure to cover the full year? Describe the handover activities done.
- What issues are there for you in taking over an employee in the middle of a performance year?
- For the employee?
- How have you handled that?
- What do you think the employee feels about the 3P handover process?

Finalise

- Have you had any other thoughts on the 3P process since we spoke last and which you would like to share? Any improvement suggestions to the process?
- Any important things left out?
- Any questions or concerns?

Interview guide for Employees regarding Final Appraisal

Intro

- Interview within 4 workforces, pairs of mgr/employees, research question
- Confidentiality (delivered doc), Semi-structured interview, Use of recording
- Confirm ok?

Warm-up

- Have you used the "diary" I forwarded after our first interview? Why not? / How did that work out?
- How would you overall describe the recent final appraisal meeting?
- Are you satisfied with the meeting?

Deep dive

- What was happening at the meeting? What went well/bad? Something you would have done differently?
- What was the overall rating? Is this the right rating in your opinion?
- What is the distribution in your area?
- What will be the impact for you?
- How did the rating come about? Can you describe the calibration process that has been used?
- Have you been in control of the target outcomes or depending on others? What impact has this had? Do you think your manager sees this the same way?
- Has your manager got a co-responsibility for your score? Describe.
- Which other biases do you identify in the process you have just been through? (e.g. personal relations, recency/memory, not measuring the right thing, manipulation, lack of info, leniency, central tendency, attribution errors (pos attributed to oneself / neg attributed to others), self-serving bias)
- Are these biases ok and manageable - if yes, how has your manager managed them?
- Has the process and rating in your view been fair to yourself?
- How have you prepared for the meeting? How has the manager prepared?
- Has the manager collected any input from others? Who?
- What key performance events were you discussing?
- How did behaviour influence the evaluation? Was this transparent to you in your targets?
- Were targets removed or updated? Who took initiative to this?
- What do you think your manager thinks of the meeting?
- How would you describe your feelings up to, during and after the meeting?
- How do you imagine your manager was feeling?
- Has the 3P process a motivation effect for you? How?

- Is this the best way to achieve that?
- What is your overall view on the way the company and you as an employee do performance management?
- Is 3P a fair process for you and your colleagues? Why/why not?
- Does the calibration process secure an accurate representation of performance across the org?
- Is it a transparent process? Would it be better or worse without the final rating?
- What is the purpose of the 3P process? Is it successful?

Particularly for employees with new managers (manager change)

- How did you make sure to cover the full year? Describe the handover activities done.
- What issues are there for you in a new manager in the middle of a performance year?
- For the manager?
- How have you handled that?
- What do you think the manager feels about the 3P handover process?

Finalise

- Have you had any other thoughts on the 3P process since we spoke last and which you would like to share? Any improvement suggestions to the process?
- Any important things left out?
- Questions or concerns?

12.6 Appendix 6: Manager's and Employees' Diaries

Performance appraisal diary for managers

The purpose of this diary is to conduct a systematic documentation of important performance appraisal related events, and your own reflections on these.

The diary is, naturally, confidential. It is covered by the declaration of confidentiality you have previously received.

Guide to diary:

Notes are made on an ongoing basis in connection to activities relating to performance appraisal. Examples of such events are:

The recurring PA meetings; target setting, midyear and yearly appraisal

- *Situations where the employee makes contributions to PA targets. E.g. project milestones, customer satisfaction surveys, contract signoffs etc.*
- *Situations where the employee is performing outside PA targets. I.e. where the employee delivers important contributions that were not documented in his/her PA.*
- *Other feedback, e.g. at 1:1 meetings, informal individual or collective feedback or other kind of praise or critique from you or other managers.*
- *Calibration meetings regarding PA, bonus, salary adjustments*
- *Other events you find are interesting in the PA context*

The sooner you make notes of your reflections after an event, the better you will be able to remember important details.

Initials:		
Date	Description of event	Own reflections on event

Performance appraisal diary for employees

The purpose of this diary is to conduct a systematic documentation of important performance appraisal related events, and your own reflections on these.

The diary is, naturally, confidential. It is covered by the declaration of confidentiality you have previously received.

Guide to diary:

Notes are made on an ongoing basis in connection to activities relating to performance appraisal. Examples of such events are:

- The recurring PA meetings; target setting, midyear and yearly appraisal
- Situations where you make contributions to PA targets. E.g. project milestones, customer satisfaction surveys, contract signoffs etc.
- Situations where you are performing outside PA targets. I.e. where you deliver important contributions that were not documented in your PA.
- Other feedback, e.g. at 1:1 meetings, informal individual or collective feedback or other kind of praise or critique from your own or other managers.

Other events you find are interesting in the PA context

The sooner you make notes of your reflections after an event, the better you will be able to remember important details.

Initials:		
Date	Description of event	Own reflections on event

12.7 Appendix 7: List of Journals Represented in Literature Review

1,002 journal articles with title, keywords or abstract containing the word combination "Performance Appraisal" were identified in a database search. Out of these, 125 articles we selected for coding. Below, the 61 journals represented through the 125 selected articles are alphabetically listed.

Applied Psychology - An International Review
Asian Business & Management
British Journal of Industrial Relations
Career Development International
Community College Journal of Research and Practice
Current Directions in Psychological Science
Employee Relations
European Journal of Innovation Management
Health Manpower Management
Human Relations
Human Resource Development International
Human Resource Management International Digest
Human Resource Management Journal
Human Resource Management Review
Industrial and Commercial Training
Industrial and Organizational Psychology
Industrial Relations Journal
Industrial Relations
International Journal of Business Performance Management
International Journal of Career Management
International Journal of Data Analysis Techniques and Strategies
International Journal of Human Resource Management
International Journal of Indian Culture and Business Management
International Journal of Management Reviews
International Journal of Manpower
International Journal of Operations & Production Management
International Journal of Productivity and Performance Management
International Journal of Selection and Assessment

Journal of Business and Psychology
Journal of Communication Disorders
Journal of Educational Administration
Journal of International Management
Journal of Knowledge Management
Journal of Management History
Journal of Management Studies
Journal of Management
Journal of Managerial Psychology
Journal of occupational and organizational psychology
Journal of Product and Brand Management
Journal of Property Research
Journal of Quality Management
Journal of the American Pharmaceutical Association
Leadership and Organization Development Journal
Management and Organization Review
Management Decision
Management Development Review
Managerial Auditing Journal
Managing Human Resources in the Human Services
Measuring Business Excellence
Organizational Behavior and Human Decision Processes
Personnel Psychology
Personnel Review
Qualitative Research in Organizations and Management
Social Justice Research
Social Science Japan Journal
Team Performance Management
The Journal of Management Development
The Journal of Workplace Learning
The TQM Magazine
Training for Quality
Work Study

13 Literature

Abraham S.E., Karns L.A., Shaw K., Mena M.A. (2001): Managerial competencies and the managerial PA process. *The Journal of Management Development*, Volume 20, Number 10, 2001, pp. 842-852.

Abu-Doleh J., Weir D. (2007): Dimensions of PA systems in Jordanian private and public organizations. *International Journal of Human Resource Management*, Volume 18, Number 1, January 2007, pp. 75-84.

Ahmed, S. (1999): The emerging measure of effectiveness for human resource management: An exploratory study with PA. *The Journal of Management Development*, Volume 18, Number 6, 1999, pp. 543-556.

Alvesson, M., Sköldbörg, K. (2009): *Reflexive Methodology. New Vistas for Qualitative Research*. London.

Alvesson, M., Kärreman, D. (2007): Unraveling HRM: Identity, Ceremony, and Control in a Management Consulting Firm. *Organization Science*, Vol. 18, No. 4, July–August 2007, pp. 711–723.

Amabile, T., Conti, R. (2011): *Motivation*. Oxford.

Amba-Rao S. C., Petrick J. A., Gupta J. N. D., Embse T.J.V.D. (2000): Comparative PA practices and management values among foreign and domestic firms in India . *International Journal of Human Resource Management*, Volume 11, Number 1, 1 January 2000, pp. 60-89.

Appelbaum, S. H., Nadeau, D., Cyr, M. (2009): Performance evaluation in a matrix organization: a case study (part 3). *Industrial and Commercial Training*, Volume 41, Number 1, 2009, pp. 9-14.

Appelbaum, S. H., Nadeau, D., Cyr, M. (2008): Performance evaluation in a matrix organization: a case study (Part Two). *Industrial and Commercial Training*, Volume 40, Number 6, 2008, pp. 295-299.

Appelbaum, S. H. (1996): SelfEfficacy as a mediator of goal setting and performance. *Journal of Managerial Psychology*, Vol. 11 Issue: 3, pp 33-47.

Armstrong, M., Angela, B. (2005): *Managing Performance*, London.

- Armstrong, M. (2006): Performance Management – Key Strategies and practical Guidelines, London.
- Ashton, D., Felstead, A. (1995): Training and development. In Storey (ed): Human Resource Management - a critical text, London.
- Austin, R. D. (1996): Measuring and Managing Performance in Organizations, New York.
- Bacal, R. (2004): Performance Management – 24 Lessons for Improving Performance, Madison.
- Bacal, R. (1999): Performance Management, Madison.
- Bai, X., Bennington, L. (2005): PA in the Chinese state-owned coal industry . International Journal of Business Performance Management, Volume 7, Number 3, 4 April 2005, pp. 275-287.
- Baines, A. (1991): Appraisal-based Pay. Work Study, Volume 40, Number 2, 1991 , pp.8-10.
- Barnard, C. I. (1938): The Functions of the Executive. Cambridge Massachusets.
- Baruch, Y. (1996): Self PA vs direct-manager appraisal: A case of congruence . Journal of Managerial Psychology, Volume 11, Number 6, 1996, pp. 50-65.
- Bates, R.A., Holton E.F (1995): Computerized performance monitoring: A review of human resource issues. Human Resource Management Review, Volume 5, Number 4, Winter 1995, pp. 267-288.
- Blau, G. (1999): Testing the Longitudinal Impact of Work Variables and Performance Appraisal Satisfaction on Subsequent Overall Job Satisfaction. Human Relations, Volume 52, Number 8, 1 August 1999, pp. 1099-1113.
- Blau, P. M. (1955): The Dynamics of Bureaucracy. Chicago.
- Boice, D. F., Kleiner, B. H. (1997): Designing effective PA systems. Work Study, Volume 46, Number 6, 1997, pp. 197-201.
- Bowen, P. (1995): Winners or losers? Ranking or co-operation? Training for Quality, Volume 3, Number 4, 1995, pp. 5-8.
- Brown, M., Benson, J. (2003): Rated to exhaustion? Reactions to PA processes . Industrial Relations Journal, Volume 34, Number 1, March 2003, pp. 67-81.

Brown, M., Heywood, J. S. (2005): Performance Appraisal Systems: Determinants and Change. *British Journal of Industrial Relations*, Volume 43, Number 4, December 2005, pp. 659-679.

Brown, M., Hyatt, D., Benson, J. (2010): Consequences of the PA experience. *Personnel Review*, Volume 39, Number 3, 2010, pp. 375-396.

Brown, W. (1968): *Exploration in Management*. Middlesex.

Buchner, T. (2007): Performance management theory: A look from the performer's perspective with implications for HRD. *Human Resource Development International*, Volume 10, Number 1, March 2007, pp. 59-73.

Burke, R. J. (1996): Performance evaluation and counselling in a professional services firm . *Leadership and Organization Development Journal*, Volume 17, Number 3, 1996, pp. 21-26.

Cameron, D. (1993): Performance Appraisal & Review. *Management Decision*, Volume 19, Number 6, 1993, pp. 3-54.

Capelli, P. (2008): Talent Management for the Twenty-First Century. *Harvard Business Review*, March 2008, pp 1-8.

Cardy R.L., Sutton C.L., Carson K.P., Dobbins G.H. (1998): Person and System Effects in Performance Appraisal: Ratings as a Function of the Degree of Performance Responsibility and Errorfulness. *Journal of Quality Management*, Volume 3, Number 1, 1998, pp. 79-99.

Caruth, D. L., Humphreys, J. H. (2008): PA: essential characteristics for strategic control. *Measuring Business Excellence*, Volume 12, Number 3, 2008, pp. 24-32.

Catano, V. M., Darr, W., Campbell, C. A. (2007): Performance Appraisal of Behavior-Based Competencies: A Reliable and Valid Procedure. *Personnel Psychology*, Volume 60, Number 1, Spring 2007, pp. 201-230.

Chang, T., Liao, J., Wang, X. (2008): The effect of team-based PA on knowledge sharing: constructing and verifying an influencing model. *International Journal of Data Analysis Techniques and Strategies*, Volume 1, Number 2, 4 November 2008, pp. 153-172.

- Cheng, K. H. C., Cascio, W. (2009): Performance-Appraisal Beliefs of Chinese Employees in Hong Kong and the Pearl River Delta. *International Journal of Selection and Assessment*, Volume 17, Number 3, September 2009, pp. 329-333.
- Claus, L., Briscoe, D. (2009): Employee performance management across borders: A review of relevant academic literature. *International Journal of Management Reviews*, Volume 11, Number 2, June 2009, pp. 175-196.
- Coens, T., Jenkins, M. (2002): *Abolishing Performance Appraisals*, San Francisco.
- Colella A., DeNisi, A.S., Varma, A. (1997): Appraising the performance of employees with disabilities: A review and model . *Human Resource Management Review*, Volume 7, Number 1, Spring 1997, pp. 27-53.
- Conger, J.A., Fulmer, R. M. (2004): *How Great Organizations use Succession Management to Sustain Competitive Advantage*. New York.
- Cook, M. (1995): PA and true performance . *Journal of Managerial Psychology*, Volume 10, Number 7, 1995, pp. 3-7.
- Creelman, J. (1995): Pay and performance drive human resource agendas . *Management Development Review*, Volume 8, Number 3, 1995, pp. 6-9.
- Deadrick, D.L., Gardner, D.G. (1999): Performance Distributions: Measuring Employee Performance Using Total Quality Management Principles. *Journal of Quality Management*, Volume 4, Number 2, March 1999, pp. 225-241.
- Deci, E. L., Koestner, R., Ryan, R.M. (2001): Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research*.
- DeNisi, A.S., Pritchard, R.D. (2006): Performance Appraisal, Performance Management and Improving Individual Performance: A Motivational Framework . *Management and Organization Review*, Volume 2, Number 2, July 2006, pp. 253-277.
- Desselle, S.P., Vaughan, M., Faria, T. (2002): Creating a Performance Appraisal Template for Pharmacy Technicians Using the Method of Equal-Appearing Intervals . *Journal of the American Pharmaceutical Association*, Volume 42, Number 5, 1 September 2002, pp. 768-779.

- Dewberry C. (2001): Performance disparities between whites and ethnic minorities: Real differences or assessment bias? . *Journal of Occupational and Organizational Psychology*, Volume 74, Number 5, December 2001, pp. 659-673.
- Douglas, M. (1986): *How institutions think*. New York.
- Drach-Zahavy A., Erez, M. (2002): Challenge versus threat effects on the goal–performance relationship. *Organizational Behavior and Human Decision Processes*, Volume 88, Number 2, July 2002, pp. 667-682.
- Drucker, P. F. (2007): *The Essential Drucker*. Oxford.
- Edmonstone, J. (1996): Appraising the state of PA. *Health Manpower Management*, Volume 22, Number 6, 1996, pp. 9-13.
- Emerson, R. M., Fretz, R.I., Shaw, L.L. (1995): *Writing Ethnographic Fieldnotes*. Chicago.
- Endo K. (1998): 'Japanization' of a PA system: a historical comparison of the American and Japanese systems . *Social Science Japan Journal*, Volume 1, Number 2, October 1998, pp. 247-262.
- Eriksson, T., Westergaard-Nielsen, N. (2009): Productivity Differences in an International Pharmaceutical Firm. *International Differences in the Business Practices & Productivity of Firms*. pp. 173-192.
- Espinosa, A.C.C., Flint, D.H. (1999): The role of organizational justice in multi-source PA: theory-based applications and directions for research. *Human Resource Management Review*, Volume 9, Number 1, Spring 1999, pp. 1-20.
- Fink, L.S., Longenecker, C. O. (1998): Training as a PA improvement strategy . *Career Development International*, Volume 3, Number 6, 1998, pp. 243-251.
- Fisher, C.M. (1994): The Differences between Appraisal Schemes: Variation and Acceptability- Part I . *Personnel Review*, Volume 23, Number 8, 1994, pp. 33-48.
- Fitz-Enz, J. (2000): *The ROI of Human Capital: Measuring the Economic Value of Employee Performance*. New York.
- Fletcher C. (2001): PA and management: The developing research agenda . *Journal of occupational and organizational psychology*.

- Flyvbjerg, B. (2006): Five Misunderstandings about case-study research. *Qualitative Inquiry*. vol. 12, no. 2, April 2006, pp. 219-245.
- French, J. R. P. (1950): Field Experiments: Changing Group Productivity, in James G. Miller (Ed.), *Experiments in Social Process: A Symposium on Social Psychology*.
- Frey, B. S. (1997): *Not Just For the Money*. Northampton Massachusetts.
- Furnham, A. (2005): *The Psychology of Behavior at Work*. London.
- Giddens, A. (1984): *The Constitution of Society*. Berkeley.
- Goffman, E. (1967): *Interaction Ritual. Essays in Face-to-Face Behavior*. Chicago.
- Greller M.M. (1998): Participation in the Performance Appraisal Review: Inflexible Manager Behavior and Variable Worker Needs . *Human Relations*, Volume 51, Number 8, 1 August 1998, pp. 1061-1083.
- Grint, K. (2005): Problems, problems, problems: The social construction of 'leadership'. *Human Relations*, Volume 58 (11), pp. 1467-1494. London.
- Grote, D. (2005): *Forced Ranking*. Boston.
- Guest, D. (1997): Human resource management and performance - a review and research. *The International Journal of Human Resource Management*, Volume 8, Issue 3, 1997, pp. 263-276.
- Gutting, G. (2005): *Foucault – A Very Short Introduction*. Oxford.
- Halse, N. (2012): *Presentation to management in relation to Organisational Audit 2012*. Unpublished.
- Halse, N., Smeets, V., Warzynski, F. (unknown year): *International Differences in Subjective Performance Evaluation, Compensation and Career Dynamics in a Global Company*. Unpublished.
- Harrington, H.J. (2000): Was W. Edwards Deming wrong?. *Measuring Business Excellence*, Volume 4, Number 3, 2000, pp. 35-41.
- Hatch, M. J. (1997): *Organisation Theory*, Oxford.
- Hempel, P.S. (2001): Differences between Chinese and Western managerial views of performance . *Personnel Review*, Volume 30, Number 2, 2001, pp. 203-226.

- Henriksen, T. D. (2010): MUS: fra stivnet ritual til strategisk mulighed. *Erhvervspsykologi: tidsskrift om udvikling, dialog, ledelse organisation* 8, 2010 (2), pp. 36-59.
- Herzberg, F., Mausner, B., Snyderman, B. (1993): *The Motivation to Work*. New York.
- Heslin, Peter A., and Walle, D. (2008): Managers' Implicit Assumptions About Personnel. *Current Directions in Psychological Science*, Volume 17, Number 3, June 2008, pp. 219-223.
- Hofstede, G. (1983): The cultural relativity of organizational practices and theories. *Journal of International Business Studies*, Vol. 14, Issue 2, pp. 75-89.
- Holbrook Jr. R.L. (1999): Managing Reactions to Performance Appraisal: The Influence of Multiple Justice Mechanisms. *Social Justice Research*, Volume 12, Number 3, September 1999, pp. 205-221.
- Holbrook R.L. (2002): Contact points and flash points: Conceptualizing the use of justice mechanisms in the PA interview. *Human Resource Management Review*, Volume 12, Number 1, Spring 2002, pp. 101-123.
- Ilstad, S., Nystuen, K. (1988): *Socialpsykologi*. Oslo.
- Cook, J., Crossman, A. (2004): Satisfaction with PA systems: A study of role perceptions. *Journal of Managerial Psychology*, Volume 19, Number 5, 2004, pp. 526-541.
- Jones, O. (1996): I was upward appraised... and survived!. *Career Development International*, Volume 1, Number 2, 1996, pp. 47-48.
- Thurston, Jr, P.W., McNall, L. (2010): Justice perceptions of PA practices . *Journal of Managerial Psychology*, Volume 25, Number 3, 2010, pp. 201-228.
- Kane, B. (2010): *Performance Talk*. PhD Thesis. Copenhagen.
- Kaplan, R. S., Norton, D.P. (2001): *The Strategy-focused organization*. Boston.
- Katsanis, L.P., Pitta, D.A. (1999): The PA process of pharmaceutical product managers in Canada: an empirical study. *Journal of Product and Brand Management*, Volume 8, Number 6, 1999, pp. 463-493.

- Kessler, I. (1995): Reward Systems. In Storey (ed): Human Resource Management - a critical text. London.
- Klitzmann, R., Chin, L.J., Rifai-Bishjawish, H., Kleinert, K., Cheng-Shiun, L.: Disclosures of funding sources and conflicts of interest in published HIV/AIDS research conducted in developing countries. *Journal of Medical Ethics*. August 2010, 36(8), pp. 505–510.
- Knippen, J.T., Green, T. B. (1995): Responding to an unfair PA . *Management Development Review*, Volume 8, Number 5, 1995, pp. 29-31.
- Kohn, A. (1993): Why Incentive Plans Cannot Work. *Harvard Business Review*, Sep 1993.
- Kohn, A. (1999): Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise and Other Bribes. Boston.
- Korsgaard, A. M., Roberson, L. (1995): Procedural justice in performance evaluation: the role of instrumental and non-instrumental voice in PA discussions. *Journal of Management*, Volume 21, Number 4, 1995, pp. 657-669.
- Kremer, M. (2002): Pharmaceuticals and the Developing World. *Journal of Economic Perspectives*. Volume 16, Number 4. pp. 67-90.
- Kuvaas, B. (2006): PA satisfaction and employee outcomes: mediating and moderating roles of work motivation . *International Journal of Human Resource Management*, Volume 17, Number 3, March 2006, pp. 504-522.
- Langan-Fox, J., Waycott, J., Morizzi, M., Mcdonald, L. (1998): Predictors of Participation in Performance Appraisal: A Voluntary System in a Blue-Collar Work Environment. *International Journal of Selection and Assessment*, Volume 6, Number 4, October 1998 , pp. 249-260.
- Latham, G. P., Budworth, M., Yanar, B., Whyte, G. (2008): The Influence of a Manager's Own Performance Appraisal on the Evaluation of Others. *International Journal of Selection and Assessment*, Volume 16, Number 3, September 2008, pp. 220-228.
- Lazear, E. P., Gibbs, M. (2009): *Personnel Economics in Practice*. Hoboken.

Lee M., Son, B. (1998): The effects of appraisal review content on employees' reactions and performance . *International Journal of Human Resource Management*, Volume 9, Number 1, 1 February 1998, pp. 203-214.

Lefkowitz, J. (2000): The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. *Journal of Occupational and Organizational Psychology*, Volume 73, Number 1, March 2000, pp. 67-85.

Levy P.E., Cawley, B.D., Foti, R.J. (1998): Reactions to Appraisal Discrepancies: Performance Ratings and Attributions . *Journal of Business and Psychology*, Volume 12, Number 4, 1998, pp. 437-455.

Longenecker, C. O. (1997): Why managerial PAs are ineffective: causes and lessons. *Career Development International*, Volume 2, Number 5, 1997, pp. 212-218.

March, J. G., Olsen, J.P. (1989): *Rediscovering Institutions*. New York.

Markle, G. L. (2000): *Catalytic Coaching – The End of the Performance Review*, London.

Marques, J. (2008): It's time employees had more say in evaluating their supervisors: Reciprocal evaluation provides path to greater effectiveness. *Human Resource Management International Digest*, Volume 16, Number 5, 2008, pp. 35-37.

Martin, D.C., Bartol, K.M. (2003): Factors influencing expatriate PA system success: an organizational perspective. *Journal of International Management*, Volume 9, Number 2, 2003, pp. 115-132.

McAllister P., Baum A., Crosby N., Gallimore, P., Gray, A. (2003): Appraiser behaviour and appraisal smoothing: some qualitative and quantitative evidence . *Journal of Property Research*, Volume 20, Number 3, September 2003, pp. 261-280(20).

McGregor, Douglas (2006): *The Human Side of Enterprise*. New York.

Mendibil, K., MacBryde, J. (2006): Factors that affect the design and implementation of team-based performance measurement systems. *International Journal of Productivity and Performance Management*, Volume 55, Number 2, 2006, pp. 118-142.

Millward, L.J., Asumeng, M., McDowall, A. (2010): "Catch me if you can?": A psychological analysis of managers' feedback seeking. *Journal of Managerial Psychology*, Volume 25, Number 4, 2010, pp. 384-407.

Miyamoto, D., Higuchi, J. (2007): Paying for Success: Performance-Related Pay Systems and its Effects on Firm Performance in Japan. *Asian Business & Management*, Volume 6, Supplement 1, December 2007, pp. S9-S31.

Mjoen, K. (2010): *The HRM-Performance Trap*. PhD Thesis. Oslo.

Molleman E., Timmerman, H. (2003): Performance management when innovation and learning become critical performance indicators. *Personnel Review*, Volume 32, Number 1, 2003, pp. 93-113(21).

Murphy, K.R. (2008): Perspectives on the Relationship Between Job Performance and Ratings of Job Performance. *Industrial and Organizational Psychology*, Volume 1, Number 2, June 2008, pp. 197-205.

Narcisse, S., Harcourt, M. (2008): Employee fairness perceptions of PA: a Saint Lucian case study. *International Journal of Human Resource Management*, Volume 19, Number 6, June 2008, pp. 1152-1169.

Nikolaou I. (2003): Fitting the person to the organisation: examining the personality-job performance relationship from a new perspective. *Journal of Managerial Psychology*, Volume 18, Number 7, 2003, pp. 639-648.

Nurse, L. (2005): PA, employee development and organizational justice: exploring the linkages . *International Journal of Human Resource Management*, Volume 16, Number 7, Number 7/July 2005, pp. 1176-1194.

Orpen, C. (1995): Employee job performance and relations with superior as moderators of the effect of appraisal goal setting on employee work attitudes . *International Journal of Career Management*, Volume 7, Number 2, 1995, pp. 3-6.

Oxford Dictionary Online

Paik Y., Vance C.M., Stage H.D. (2000): A test of assumed cluster homogeneity for PA management in four Southeast Asian countries. *International Journal of Human Resource Management*, Volume 11, Number 4, 1 August 2000, pp. 736-750.

Payne, S.C., Horner, M.T., Boswell, W.R., Schroeder, A.N., Stine-Cheyne, K.J. (2009): Comparison of online and traditional PA systems . *Journal of Managerial Psychology*, Volume 24, Number 6, 2009, pp. 526-544.

Perlmutter, F.D., Bailey, D., Netting, F.E. (2000): Motivating, Appraising, and Rewarding. *Managing Human Resources in the Human Services*, July 2000, pp. 144-165.

Pettijohn L.S., Parker R.S., Pettijohn C.E., Kent O.L. (2001): PAs: usage, criteria and observations. *The Journal of Management Development*, Volume 20, Number 9, 2001, pp. 754-771.

Pickett, L. (2003): Transforming the annual fiasco . *Industrial and Commercial Training*, Volume 35, Number 6, 2003, pp. 237-240.

Piercy, N.F. (1994): The Real Strategic Issues versus Organizational Bureaucracy and Administrative Neatness . *Management Decision*, Volume 32, Number 4, 1994, pp. 5-8.

DiMaggio, P. J., Powell, W. (1983): The iron cage revisited. Institutional isomorphism and collective rationality in organizational fields, *American Sociological Review*, 48 (1983), pp. 147-60.

Poon J.M.L. (2004): Effects of PA politics on job satisfaction and turnover intention. *Personnel Review*, Volume 33, Number 3, 2004, pp. 322-334.

Powell, W.W., DiMaggio, P.J. (ed.) (1991): *The New Institutionalism in Organisational Analysis*. Chicago.

Powers, M. (2004): Counting, Control and Calculation: Reflections on Measuring and Management. *Human Relations*, June 2004 vol. 57 no. 6. pp. 765-783.

Prendergast, C., Topel, R. (1993): Discretion and bias in performance evaluation. *European Economic Review* 37. pp. 355-365.

Prowse, P., Prowse, J. (2009): The dilemma of PA . *Measuring Business Excellence*, Volume 13, Number 4, 2009, pp. 69-77.

Ragin, C.C., Becker, H.S. (1992): *What is a Case?*. Cambridge.

Ragin, C.C. (1997): Turning the Tables - How Case-oriented Research challenges Variable-oriented Research. *Comparative Social Research*, Volume 16. pp. 27-42.

- Randle, K. (1997): Rewarding failure: operating a performance-related pay system in pharmaceutical research . *Personnel Review*, Volume 26, Number 3, 1997, pp. 187-200.
- Rasch, L. (2004): Employee Performance Appraisal and the 95/5 Rule. *Community College Journal of Research and Practice*, Volume 28, Number 5, June 2004, pp. 407-414.
- Raymark, P.H., Balzer, W.K., DeLaTorre, F. (1999): A Preliminary Investigation of the Sources of Information Used by Raters When Appraising Performance . *Journal of Business and Psychology*, Volume 14, Number 2, 1999, pp. 319-339.
- Repstad, P. (2004): *Mellom Nærhet og Distanse*. Oslo.
- Ring, V. (1989): Formal and informal dimensions of transactions. In Van de Ven, Angle and Poole (Eds), *Research on the Management of Innovation: The Minnesota Studies*. New York.
- Roch, S.G., Sternburgh, A.M., Caputo, P.M. (2007): Absolute vs Relative Performance Rating Formats: Implications for fairness and organizational justice . *International Journal of Selection and Assessment*, Volume 15, Number 3, September 2007, pp. 302-316.
- Ryan, R.M., Deci, E.L. (2000): Intrinsic and Extrinsic Motivations - Classic Definitions and New Directions. *Contemporary Educational Psychology* 25, pp. 54-67.
- Salimäki, A., Jämsén, S. (2010): Perceptions of politics and fairness in merit pay . *Journal of Managerial Psychology*, Volume 25, Number 3, 2010, pp. 229-251.
- Samarakone, P. (2010): Improving PAs using a real-time talent management system: The advantages of a real-time talent management system . *Human Resource Management International Digest*, Volume 18, Number 4, 2010, pp. 35-37.
- Scherer, F.M. (1993): Pricing, Profits, and Technological Progress in the Pharmaceutical Industry. *Journal of Economic Perspectives*. Vol 7, no. 3, pp. 97-115.
- Schweiger, I., Sumners, G.E. (1994): Optimizing the Value of Performance Appraisals. *Managerial Auditing Journal*, Volume 9, Number 8, 1994, pp. 3-7.
- Scott, R.W. (2008): *Institutions and Organisations - Ideas and Interests*. Los Angeles.

Shaikh, T.S. (1995): Appraising job performance - to be or not to be? An Asiandilemma . *International Journal of Career Management*, Volume 7, Number 5, 1995, pp. 13-18.

Shibata H. (2000): The transformation of the wage and PA system in a Japanese firm. *International Journal of Human Resource Management*, Volume 11, Number 2, 1 April 2000, pp. 294-313.

Shibata, H. (2002): Wage and Performance Appraisal Systems in Flux: A Japan-United States Comparison. *Industrial Relations*, Volume 41, Number 4, October 2002, pp. 629-652.

Shore T.H., Adams J.S., Tashchian, A. (1998): Effects of Self-Appraisal Information, Appraisal Purpose, and Feedback Target on Performance Appraisal Ratings. *Journal of Business and Psychology*, Volume 12, Number 3, 1998, pp. 283-298.

Sillup, G.P., Klimberg, R. (2010): Assessing the ethics of implementing PA systems. *The Journal of Management Development*, Volume 29, Number 1, 2010 , pp. 38-55.

Simmons, J. (2002): An "expert witness" perspective on PA in universities and colleges . *Employee Relations*, Volume 24, Number 1, 2002, pp. 86-100.

Skinner, B.F. (1976): *About Behaviorism*, New York.

Smith, A.D., Rupp, W.T. (2004): Knowledge workers' perceptions of performance ratings . *The Journal of Workplace Learning*, Volume 16, Number 3, 2004, pp. 146-166.

Smith, A.D., Rupp, W.T. (2003): Knowledge workers: exploring the link among performance rating, pay and motivational aspects. *Journal of Knowledge Management*, Volume 7, Number 1, 2003, pp. 107-124.

Snape, E., Thompson, D., Yan, F.K., Redman. T. (1998): PA and culture: practice and attitudes in Hong Kong and Great Britain. *International Journal of Human Resource Management*, Volume 9, Number 5, 1 October 1998, pp. 841-861.

Soltani, E., Meer, R.B., Gennard, J., Williams, M.T. (2004): Have TQM organisations adjusted their performance management (appraisal) systems? A study of UK-based TQM-driven organisations. *The TQM Magazine*, Volume 16, Number 6, 2004, pp. 403-417.

Soltani, E., Meer, R., Williams, T.M., Lai, P. (2006): The compatibility of PA systems with TQM principles? Evidence from current practice . *International Journal of Operations & Production Management*, Volume 26, Number 1, 2006, pp. 92-112.

Spicer, D.P., Ahmad, R. (2006): Cognitive processing models in PA: evidence from the Malaysian education system. *Human Resource Management Journal*, Volume 16, Number 2, April 2006, pp. 214-230.

Spinks, N., Wells, B., Meche, M. (1999): Appraising the appraisals: computerized PA systems . *Career Development International*, Volume 4, Number 2, 1999, pp. 94-100.

Stark, E., Poppler, P. (2009): Leadership, performance evaluations, and all the usual suspects. *Personnel Review*, Volume 38, Number 3, 2009, pp. 320-338.

St-Onge, S., Morin, D., Bellehumeur, M., Dupuis, F. (2009): Managers' motivation to evaluate subordinate performance . *Qualitative Research in Organizations and Management: An International Journal*, Volume 4, Number 3, 2009, pp. 273-293.

Strauss, A., Corbin, J. (1990): *Basics of Qualitative Research*. London.

Sudarsan, A. (2009): PA criteria: a survey of Indian organisations. *International Journal of Indian Culture and Business Management*, Volume 2, Number 4, 14 April 2009, pp. 373-391.

Sullivan, J. (2002): Did HR Cause The Downfall of Enron? <http://www.drjohnsullivan.com/newsletter-archives/46-did-hr-cause-the-downfall-of-enron>.

Sutcliffe, K.M., Weick, K.E. (2009): *Information Overload Revisited in Weick: Making sense of the Organization*. Chichester.

Sutcliffe, K.M., Obstfeld, D. (2005): Organizing and the Process of Sensemaking. In *Organizational Science* Vol. 16, No. 4, pp. 409-421.

Taylor, F.W. (1998): *The Principles of Scientific Management*. New York.

Taylor, S.T, Collins, C.J. (2000): Organizational Recruitment: Enhancing the Intersection of Research and Practice. In C. L. Cooper & E. A. Locke (Eds.): *Industrial and Organizational Psychology*, pp. 304-334.

Timperley, H.S. (1998): PA: principals' perspectives and some implications . *Journal of Educational Administration*, Volume 36, Number 1, 1998, pp. 44-58.

Tinkham, R., Kleiner, B.H. (1993): New approaches to managing PAs . *Work Study*, Volume 42, Number 7, 1993, pp. 5-7.

Townley, B. (1999): Practical Reason and Performance Appraisal. *Journal of Management Studies*, Volume 36, Number 3, May 1999, pp. 287-306.

Tuchman, B. (1984): *The March of the Folly - from Troy to Vietnam*. London.

Tziner, A., Kopelman, R.E. (2002): Is there a Preferred Performance Rating Format? A Non-psychometric Perspective. *Applied Psychology An International Review*, Volume 51, Number 3, July 2002, pp. 479-503.

Tziner, A., Murphy, K.R. (1999): Additional Evidence of Attitudinal Influences in Performance Appraisal . *Journal of Business and Psychology*, Volume 13, Number 3, 1999, pp. 407-419.

Tziner, A., Murphy, K.R., Cleveland, J.N., Beaudin, G., March S. (1998): Impact of Rater Beliefs Regarding Performance Appraisal and Its Organizational Context on Appraisal Quality . *Journal of Business and Psychology*, Volume 12, Number 4, 1998, pp. 457-467.

Tziner, A., Murphy, K.R., Cleveland, J.N. (2002): Does Conscientiousness Moderate the Relationship Between Attitudes and Beliefs Regarding Performance Appraisal and Rating Behavior? *International Journal of Selection and Assessment*, Volume 10, Number 3, September 2002, pp. 218-224.

Tziner, A., Murphy, K.R., Cleveland, J.N., Roberts-Thompson, G.P. (2001): Relationships Between Attitudes Toward Organizations and Performance Appraisal Systems and Rating Behavior. *International Journal of Selection and Assessment*, Volume 9, Number 3, September 2001, pp. 226-239.

Ubeda, C.L., Santos, A. (2007): Staff development and PA in a Brazilian research centre . *European Journal of Innovation Management*, Volume 10, Number 1, 2007, pp. 109-125.

Ulrich, D., Brockbank, W. (2005): *HR Value Proposition*, Boston.

Ulrich, D. (1997): *Human Resources Champions*, Boston.

Valle, M., Davis, K. (1999): Teams and PA: Using metrics to increase reliability and validity. *Team Performance Management*, Volume 5, Number 8, 1999, pp. 238-244.

- Varma, A., Pichler, S., Srinivas, E.S. (2005): The role of interpersonal affect in PA: evidence from two samples – the US and India. *International Journal of Human Resource Management*, Volume 16, Number 11, Number 11/November 2005, pp. 2029-2044.
- Ven, A.H. Van De (2007): *Engaged Scholarship. A Guide for Organisational and Social Research*. Oxford.
- Vest, M.J., Scott, K.D., Vest J.M., Markham, S.E. (2000): Factors Influencing Employee Beliefs That Pay Is Tied to Performance. *Journal of Business and Psychology*, Volume 14, Number 4, 2000, pp. 553-562.
- Vroom, V. H. (1964): *Work and motivation*. New York.
- Vroom, V.H., Deci, E.L. (ed.) (1989): *Management and Motivation*. London.
- Wagner, B.T., Hess, C.W. (1999): Supervisors' use of social power with graduate supervisees in speech-language pathology - A neglected aspect of the PA interview. *Journal of Communication Disorders*, Volume 32, Number 5, 10 September 1999, pp. 351-368.
- Wagner, S.H., Goffin, R.D. (1997): Differences in Accuracy of Absolute and Comparative Performance Appraisal Methods. *Organizational Behavior and Human Decision Processes*, Volume 70, Number 2, May 1997, pp. 95-103.
- Waite, M.L., Stites-Doe, S. (2000): Removing PA and merit pay in the name of quality: an empirical study of employees' reactions. *Journal of Quality Management*, Volume 5, Number 2, 3rd Quarter 2000 , pp. 187-206.
- Watson, T. (2002): *Organising and Managing Work*. Harlow.
- Weick, K.E., Sutcliffe, K. (2003): Hospitals as Cultures of Entrapment: A Re-analysis of the Bristol Royal Infirmary. *California Management Review*, Vol 45, No 2, pp. 73-84.
- Weick, K.E., Roberts, K.H. (1993): Collective mind in organizations: Heeful interrelating on flight decks. *Administrative Science Quarterly*, 38, pp. 357-381.
- Weick, K.E. (2009): *Making Sense of the Organization. The Impermanent Organization*. Chichester.
- Weick, K.E. (1995): *Sensemaking in Organisations*. London.

Weick, K.E. (1993): The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster. *Administrative Science Quarterly*, 38, 1993, pp. 628-652.

Weick, K.E. (2004): Faith, Evidence, and Action: Better Guesses in an Unknowable World. *Organization Studies*, Issue 27(11), pp. 1723-1736.

Weick, K.E. (2005): Making Sense of Blurred Images: Mindful Organizing in Mission STS-107. In Farjoun M. and Starbuck W.H. (Eds): *Organisation at the Limit: Lessons from the Columbia Disaster*. Blackwell.

White, L.D. (1958): *The Republican Era*. New York.

Wiese, D.S., Buckley, M. R. (1998): The evolution of the PA process . *Journal of Management History*, Volume 4, Number 3, 1998, pp. 233-249.

Williams J.R., Levy, P.E. (2000): Investigating Some Neglected Criteria: The Influence of Organizational Level and Perceived System Knowledge on Appraisal Reactions . *Journal of Business and Psychology*, Volume 14, Number 3, 2000, pp. 501-513.

Wilson, J.P., Western, S. (2001): PA: an obstacle to training and development? *Career Development International*, Volume 6, Number 2, pp. 93-100.

Wood, S. (1999): Human resource management and performance. *International Journal of Management Reviews*, *International Journal of Management Reviews*, Volume 1, Issue 4, pp. 367–413.

Wood, R.E., Marshall, V. (2008): Accuracy and effectiveness in appraisal outcomes: the influence of self-efficacy, personal factors and organisational variables. *Human Resource Management Journal*, Volume 18, Number 3, July 2008, pp. 295-313.

Woods, P. (2003): Performance management of Australian and Singaporean expatriates . *International Journal of Manpower*, Volume 24, Number 5, 2003, pp. 517-534.

Wren, D.A. (1994): *The Evolution of Management Thought*, John Wiley & Sons, Inc., New York.

Wright, A. (1998): Counselling skills: Part II - making sense of PA, coaching and mentoring . *Industrial and Commercial Training*, Volume 30, Number 5, 1998, pp. 176-178.

13. Literature

Wright, R.P., Cheung, F.K. (2007): Articulating appraisal system effectiveness based on managerial cognitions . *Personnel Review*, Volume 36, Number 2, 2007, pp. 206-230.

Yariv, E. (2006): "Mum effect"?: principals' reluctance to submit negative feedback . *Journal of Managerial Psychology*, Volume 21, Number 6, 2006, pp. 533-546.

Yin, R.K. (2009): *Case Study Research. Design and Methods*. London.

TITLER I PH.D.SERIEN:**2004**

1. Martin Grieger
Internet-based Electronic Marketplaces and Supply Chain Management
2. Thomas Basbøll
*LIKENESS
A Philosophical Investigation*
3. Morten Knudsen
*Beslutningens vaklen
En systemteoretisk analyse af moderniseringen af et amtskommunalt sundhedsvæsen 1980-2000*
4. Lars Bo Jeppesen
*Organizing Consumer Innovation
A product development strategy that is based on online communities and allows some firms to benefit from a distributed process of innovation by consumers*
5. Barbara Dragsted
*SEGMENTATION IN TRANSLATION AND TRANSLATION MEMORY SYSTEMS
An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process*
6. Jeanet Hardis
*Sociale partnerskaber
Et socialkonstruktivistisk casestudie af partnerskabsaktørers virkelighedsopfattelse mellem identitet og legitimitet*
7. Henriette Hallberg Thygesen
System Dynamics in Action
8. Carsten Mejer Plath
Strategisk Økonomistyring
9. Annemette Kjærgaard
Knowledge Management as Internal Corporate Venturing
10. Knut Arne Hovdal
*De professionelle i endring
Norsk ph.d., ej til salg gennem Samfundslitteratur*
11. Søren Jeppesen
*Environmental Practices and Greening Strategies in Small Manufacturing Enterprises in South Africa
– A Critical Realist Approach*
12. Lars Frode Frederiksen
*Industriel forskningsledelse
– på sporet af mønstre og samarbejde i danske forskningsintensive virksomheder*
13. Martin Jes Iversen
*The Governance of GN Great Nordic
– in an age of strategic and structural transitions 1939-1988*
14. Lars Pynt Andersen
*The Rhetorical Strategies of Danish TV Advertising
A study of the first fifteen years with special emphasis on genre and irony*
15. Jakob Rasmussen
Business Perspectives on E-learning
16. Sof Thrane
*The Social and Economic Dynamics of Networks
– a Weberian Analysis of Three Formalised Horizontal Networks*
17. Lene Nielsen
Engaging Personas and Narrative Scenarios – a study on how a user-centered approach influenced the perception of the design process in the e-business group at AstraZeneca
18. S.J Valstad
*Organisationsidentitet
Norsk ph.d., ej til salg gennem Samfundslitteratur*

19. Thomas Lyse Hansen
Six Essays on Pricing and Weather risk in Energy Markets
20. Sabine Madsen
Emerging Methods – An Interpretive Study of ISD Methods in Practice
21. Evis Sinani
The Impact of Foreign Direct Investment on Efficiency, Productivity Growth and Trade: An Empirical Investigation
22. Bent Meier Sørensen
Making Events Work Or, How to Multiply Your Crisis
23. Pernille Schnoor
*Brand Ethos
Om troværdige brand- og virksomhedsidentiteter i et retorisk og diskursteoretisk perspektiv*
24. Sidsel Fabech
*Von welchem Österreich ist hier die Rede?
Diskursive forhandlinger og magtkampe mellem rivaliserende nationale identitetskonstruktioner i østrigske pressediskurser*
25. Klavs Odgaard Christensen
*Sprogpolitik og identitetsdannelse i flersprogede forbundsstater
Et komparativt studie af Schweiz og Canada*
26. Dana B. Minbaeva
Human Resource Practices and Knowledge Transfer in Multinational Corporations
27. Holger Højlund
*Markedets politiske fornuft
Et studie af velfærdens organisering i perioden 1990-2003*
28. Christine Mølgaard Frandsen
*A.s erfaring
Om mellemværendets praktik i en transformation af mennesket og subjektiviteten*
29. Sine Nørholm Just
The Constitution of Meaning – A Meaningful Constitution? Legitimacy, identity, and public opinion in the debate on the future of Europe
- 2005**
1. Claus J. Varnes
Managing product innovation through rules – The role of formal and structured methods in product development
2. Helle Hedegaard Hein
Mellem konflikt og konsensus – Dialogudvikling på hospitalsklinikker
3. Axel Rosenø
Customer Value Driven Product Innovation – A Study of Market Learning in New Product Development
4. Søren Buhl Pedersen
*Making space
An outline of place branding*
5. Camilla Funck Ellehave
*Differences that Matter
An analysis of practices of gender and organizing in contemporary work-places*
6. Rigmor Madeleine Lond
Styring af kommunale forvaltninger
7. Mette Aagaard Andreassen
Supply Chain versus Supply Chain Benchmarking as a Means to Managing Supply Chains
8. Caroline Aggestam-Pontoppidan
*From an idea to a standard
The UN and the global governance of accountants' competence*
9. Norsk ph.d.
10. Vivienne Heng Ker-ni
An Experimental Field Study on the

- Effectiveness of Grocer Media Advertising*
Measuring Ad Recall and Recognition, Purchase Intentions and Short-Term Sales
11. Allan Mortensen
Essays on the Pricing of Corporate Bonds and Credit Derivatives
12. Remo Stefano Chiari
Figure che fanno conoscere
Itinerario sull'idea del valore cognitivo e espressivo della metafora e di altri tropi da Aristotele e da Vico fino al cognitivismo contemporaneo
13. Anders McIlquham-Schmidt
Strategic Planning and Corporate Performance
An integrative research review and a meta-analysis of the strategic planning and corporate performance literature from 1956 to 2003
14. Jens Geersbro
The TDF – PMI Case
Making Sense of the Dynamics of Business Relationships and Networks
15. Mette Andersen
Corporate Social Responsibility in Global Supply Chains
Understanding the uniqueness of firm behaviour
16. Eva Boxenbaum
Institutional Genesis: Micro – Dynamic Foundations of Institutional Change
17. Peter Lund-Thomsen
Capacity Development, Environmental Justice NGOs, and Governance: The Case of South Africa
18. Signe Jarlov
Konstruktioner af offentlig ledelse
19. Lars Stæhr Jensen
Vocabulary Knowledge and Listening Comprehension in English as a Foreign Language
- An empirical study employing data elicited from Danish EFL learners*
20. Christian Nielsen
Essays on Business Reporting
Production and consumption of strategic information in the market for information
21. Marianne Thejls Fischer
Egos and Ethics of Management Consultants
22. Annie Bekke Kjær
Performance management i Process-innovation
– belyst i et social-konstruktivistisk perspektiv
23. Suzanne Dee Pedersen
GENTAGELSENS METAMORFOSE
Om organisering af den kreative gørem i den kunstneriske arbejdspraksis
24. Benedikte Dorte Rosenbrink
Revenue Management
Økonomiske, konkurrencemæssige & organisatoriske konsekvenser
25. Thomas Riise Johansen
Written Accounts and Verbal Accounts
The Danish Case of Accounting and Accountability to Employees
26. Ann Fogelgren-Pedersen
The Mobile Internet: Pioneering Users' Adoption Decisions
27. Birgitte Rasmussen
Ledelse i fællesskab – de tillidsvalgte fornyende rolle
28. Gitte Thit Nielsen
Remerger
– skabende ledelseskrafter i fusion og opkøb
29. Carmine Gioia
A MICROECONOMETRIC ANALYSIS OF MERGERS AND ACQUISITIONS

30. Ole Hinz
Den effektive forandringsleder: pilot, pædagog eller politiker?
Et studie i arbejdslederens meningstilskrivninger i forbindelse med vellykket gennemførelse af ledelsesinitierede forandringsprojekter
31. Kjell-Åge Gotvassli
Et praksisbasert perspektiv på dynamiske læringsnettverk i toppidretten
Norsk ph.d., ej til salg gennem Samfundslitteratur
32. Henriette Langstrup Nielsen
Linking Healthcare
An inquiry into the changing performances of web-based technology for asthma monitoring
33. Karin Tweddell Levinsen
Virtuel Uddannelsespraksis
Master i IKT og Læring – et casestudie i hvordan proaktiv proceshåndtering kan forbedre praksis i virtuelle læringsmiljøer
34. Anika Liversage
Finding a Path
Labour Market Life Stories of Immigrant Professionals
35. Kasper Elmquist Jørgensen
Studier i samspillet mellem stat og erhvervsliv i Danmark under 1. verdenskrig
36. Finn Janning
A DIFFERENT STORY
Seduction, Conquest and Discovery
37. Patricia Ann Plackett
Strategic Management of the Radical Innovation Process
Leveraging Social Capital for Market Uncertainty Management
2. Niels Rom-Poulsen
Essays in Computational Finance
3. Tina Brandt Husman
Organisational Capabilities, Competitive Advantage & Project-Based Organisations
The Case of Advertising and Creative Good Production
4. Mette Rosenkrands Johansen
Practice at the top
– how top managers mobilise and use non-financial performance measures
5. Eva Parum
Corporate governance som strategisk kommunikations- og ledelsesværktøj
6. Susan Aagaard Petersen
Culture's Influence on Performance Management: The Case of a Danish Company in China
7. Thomas Nicolai Pedersen
The Discursive Constitution of Organizational Governance – Between unity and differentiation
The Case of the governance of environmental risks by World Bank environmental staff
8. Cynthia Selin
Volatile Visions: Transactions in Anticipatory Knowledge
9. Jesper Banghøj
Financial Accounting Information and Compensation in Danish Companies
10. Mikkel Lucas Overby
Strategic Alliances in Emerging High-Tech Markets: What's the Difference and does it Matter?
11. Tine Aage
External Information Acquisition of Industrial Districts and the Impact of Different Knowledge Creation Dimensions

2006

1. Christian Vintergaard
Early Phases of Corporate Venturing

- A case study of the Fashion and Design Branch of the Industrial District of Montebelluna, NE Italy*
12. Mikkel Flyverbom
*Making the Global Information Society Governable
On the Governmentality of Multi-Stakeholder Networks*
 13. Anette Grønning
*Personen bag
Tilstedevær i e-mail som interaktionsform mellem kunde og medarbejder i dansk forsikringskontekst*
 14. Jørn Helder
*One Company – One Language?
The NN-case*
 15. Lars Bjerregaard Mikkelsen
*Differing perceptions of customer value
Development and application of a tool for mapping perceptions of customer value at both ends of customer-supplier dyads in industrial markets*
 16. Lise Granerud
*Exploring Learning
Technological learning within small manufacturers in South Africa*
 17. Esben Rahbek Pedersen
*Between Hopes and Realities:
Reflections on the Promises and Practices of Corporate Social Responsibility (CSR)*
 18. Ramona Samson
*The Cultural Integration Model and European Transformation.
The Case of Romania*
- 2007**
1. Jakob Vestergaard
*Discipline in The Global Economy
Panopticism and the Post-Washington Consensus*
 2. Heidi Lund Hansen
*Spaces for learning and working
A qualitative study of change of work, management, vehicles of power and social practices in open offices*
 3. Sudhanshu Rai
*Exploring the internal dynamics of software development teams during user analysis
A tension enabled Institutionalization Model; "Where process becomes the objective"*
 4. Norsk ph.d.
Ej til salg gennem Samfundslitteratur
 5. Serden Ozcan
*EXPLORING HETEROGENEITY IN ORGANIZATIONAL ACTIONS AND OUTCOMES
A Behavioural Perspective*
 6. Kim Sundtoft Hald
*Inter-organizational Performance Measurement and Management in Action
– An Ethnography on the Construction of Management, Identity and Relationships*
 7. Tobias Lindeberg
*Evaluative Technologies
Quality and the Multiplicity of Performance*
 8. Merete Wedell-Wedellsborg
*Den globale soldat
Identitetsdannelse og identitetsledelse i multinationale militære organisationer*
 9. Lars Frederiksen
*Open Innovation Business Models
Innovation in firm-hosted online user communities and inter-firm project ventures in the music industry
– A collection of essays*
 10. Jonas Gabrielsen
Retorisk toposlære – fra statisk 'sted' til persuasiv aktivitet

11. Christian Moldt-Jørgensen
Fra meningsløs til meningsfuld evaluering.
Anvendelsen af studentertilfredsheds-målinger på de korte og mellemlange videregående uddannelser set fra et psykodynamisk systemperspektiv
12. Ping Gao
Extending the application of actor-network theory
Cases of innovation in the telecommunications industry
13. Peter Mejlby
Frihed og fængsel, en del af den samme drøm?
Et phronetisk baseret casestudie af frigørelsens og kontrollens sam-eksistens i værdibaseret ledelse!
14. Kristina Birch
Statistical Modelling in Marketing
15. Signe Poulsen
Sense and sensibility:
The language of emotional appeals in insurance marketing
16. Anders Bjerre Trolle
Essays on derivatives pricing and dynamic asset allocation
17. Peter Feldhütter
Empirical Studies of Bond and Credit Markets
18. Jens Henrik Eggert Christensen
Default and Recovery Risk Modeling and Estimation
19. Maria Theresa Larsen
Academic Enterprise: A New Mission for Universities or a Contradiction in Terms?
Four papers on the long-term implications of increasing industry involvement and commercialization in academia
20. Morten Wellendorf
Postimplementering af teknologi i den offentlige forvaltning
Analysen af en organisations kontinuerlige arbejde med informations-teknologi
21. Ekaterina Mhaanna
Concept Relations for Terminological Process Analysis
22. Stefan Ring Thorbjørnsen
Forsvaret i forandring
Et studie i officerers kapabiliteter under påvirkning af omverdenens forandringspres mod øget styring og læring
23. Christa Breum Amhøj
Det selvskabte medlemskab om managementsstaten, dens styringsteknologier og indbyggere
24. Karoline Bromose
Between Technological Turbulence and Operational Stability
– An empirical case study of corporate venturing in TDC
25. Susanne Justesen
Navigating the Paradoxes of Diversity in Innovation Practice
– A Longitudinal study of six very different innovation processes – in practice
26. Luise Noring Henler
Conceptualising successful supply chain partnerships
– Viewing supply chain partnerships from an organisational culture perspective
27. Mark Mau
Kampen om telefonen
Det danske telefonvæsen under den tyske besættelse 1940-45
28. Jakob Halskov
The semiautomatic expansion of existing terminological ontologies using knowledge patterns discovered

- on the WWW – an implementation and evaluation*
29. Gergana Koleva
European Policy Instruments Beyond Networks and Structure: The Innovative Medicines Initiative
 30. Christian Geisler Asmussen
Global Strategy and International Diversity: A Double-Edged Sword?
 31. Christina Holm-Petersen
*Stolthed og fordom
Kultur- og identitetsarbejde ved skabelsen af en ny sengeafdeling gennem fusion*
 32. Hans Peter Olsen
*Hybrid Governance of Standardized States
Causes and Contours of the Global Regulation of Government Auditing*
 33. Lars Bøge Sørensen
Risk Management in the Supply Chain
 34. Peter Aagaard
*Det unikkes dynamikker
De institutionelle mulighedsbetingelser bag den individuelle udforskning i professionelt og frivilligt arbejde*
 35. Yun Mi Antorini
*Brand Community Innovation
An Intrinsic Case Study of the Adult Fans of LEGO Community*
 36. Joachim Lynggaard Boll
*Labor Related Corporate Social Performance in Denmark
Organizational and Institutional Perspectives*
- 2008**
1. Frederik Christian Vinten
Essays on Private Equity
 2. Jesper Clement
Visual Influence of Packaging Design on In-Store Buying Decisions
 3. Marius Brostrøm Kousgaard
*Tid til kvalitetsmåling?
– Studier af indrulleringsprocesser i forbindelse med introduktionen af kliniske kvalitetsdatabaser i speciallægepraksissektoren*
 4. Irene Skovgaard Smith
*Management Consulting in Action
Value creation and ambiguity in client-consultant relations*
 5. Anders Rom
*Management accounting and integrated information systems
How to exploit the potential for management accounting of information technology*
 6. Marina Candi
Aesthetic Design as an Element of Service Innovation in New Technology-based Firms
 7. Morten Schnack
*Teknologi og tværfaglighed
– en analyse af diskussionen omkring indførelse af EPJ på en hospitalsafdeling*
 8. Helene Balslev Clausen
Juntos pero no revueltos – un estudio sobre emigrantes norteamericanos en un pueblo mexicano
 9. Lise Justesen
*Kunsten at skrive revisionsrapporter.
En beretning om forvaltningsrevisionsens beretninger*
 10. Michael E. Hansen
The politics of corporate responsibility: CSR and the governance of child labor and core labor rights in the 1990s
 11. Anne Roepstorff
Holdning for handling – en etnologisk undersøgelse af Virksomheders Sociale Ansvar/CSR

12. Claus Bajlum
Essays on Credit Risk and Credit Derivatives
13. Anders Bojesen
The Performative Power of Competence – an Inquiry into Subjectivity and Social Technologies at Work
14. Satu Reijonen
*Green and Fragile
A Study on Markets and the Natural Environment*
15. Ilduara Busta
*Corporate Governance in Banking
A European Study*
16. Kristian Anders Hvass
*A Boolean Analysis Predicting Industry Change: Innovation, Imitation & Business Models
The Winning Hybrid: A case study of isomorphism in the airline industry*
17. Trine Paludan
De uvidende og de udviklingsparate Identitet som mulighed og restriktion blandt fabriksarbejdere på det aftayloriserede fabriksgulv
18. Kristian Jakobsen
Foreign market entry in transition economies: Entry timing and mode choice
19. Jakob Elming
Syntactic reordering in statistical machine translation
20. Lars Brømsøe Termansen
*Regional Computable General Equilibrium Models for Denmark
Three papers laying the foundation for regional CGE models with agglomeration characteristics*
21. Mia Reinholdt
The Motivational Foundations of Knowledge Sharing
22. Frederikke Krogh-Meibom
*The Co-Evolution of Institutions and Technology
– A Neo-Institutional Understanding of Change Processes within the Business Press – the Case Study of Financial Times*
23. Peter D. Ørberg Jensen
OFFSHORING OF ADVANCED AND HIGH-VALUE TECHNICAL SERVICES: ANTECEDENTS, PROCESS DYNAMICS AND FIRMLEVEL IMPACTS
24. Pham Thi Song Hanh
Functional Upgrading, Relational Capability and Export Performance of Vietnamese Wood Furniture Producers
25. Mads Vangkilde
*Why wait?
An Exploration of first-mover advantages among Danish e-grocers through a resource perspective*
26. Hubert Buch-Hansen
*Rethinking the History of European Level Merger Control
A Critical Political Economy Perspective*
- 2009**
1. Vivian Lindhardsen
From Independent Ratings to Communal Ratings: A Study of CWA Raters' Decision-Making Behaviours
2. Guðrið Weihe
Public-Private Partnerships: Meaning and Practice
3. Chris Nøkkentved
*Enabling Supply Networks with Collaborative Information Infrastructures
An Empirical Investigation of Business Model Innovation in Supplier Relationship Management*
4. Sara Louise Muhr
Wound, Interrupted – On the Vulnerability of Diversity Management

5. Christine Sestoft
Forbrugeradfærd i et Stats- og Livsformsteoretisk perspektiv
6. Michael Pedersen
Tune in, Breakdown, and Reboot: On the production of the stress-fit self-managing employee
7. Salla Lutz
Position and Reposition in Networks – Exemplified by the Transformation of the Danish Pine Furniture Manufacturers
8. Jens Forssbæck
Essays on market discipline in commercial and central banking
9. Tine Murphy
Sense from Silence – A Basis for Organised Action
How do Sensemaking Processes with Minimal Sharing Relate to the Reproduction of Organised Action?
10. Sara Malou Strandvad
Inspirations for a new sociology of art: A sociomaterial study of development processes in the Danish film industry
11. Nicolaas Mouton
On the evolution of social scientific metaphors: A cognitive-historical enquiry into the divergent trajectories of the idea that collective entities – states and societies, cities and corporations – are biological organisms.
12. Lars Andreas Knutsen
Mobile Data Services: Shaping of user engagements
13. Nikolaos Theodoros Korfiatis
Information Exchange and Behavior
A Multi-method Inquiry on Online Communities
14. Jens Albæk
Forestillinger om kvalitet og tværfaglighed på sygehuse
– skabelse af forestillinger i læge- og plejegrupperne angående relevans af nye idéer om kvalitetsudvikling gennem tolkningsprocesser
15. Maja Lotz
The Business of Co-Creation – and the Co-Creation of Business
16. Gitte P. Jakobsen
Narrative Construction of Leader Identity in a Leader Development Program Context
17. Dorte Hermansen
“Living the brand” som en brandorienteret dialogisk praksis: Om udvikling af medarbejdernes brandorienterede dømmekraft
18. Aseem Kinra
Supply Chain (logistics) Environmental Complexity
19. Michael Nørager
How to manage SMEs through the transformation from non innovative to innovative?
20. Kristin Wallevik
Corporate Governance in Family Firms
The Norwegian Maritime Sector
21. Bo Hansen Hansen
Beyond the Process
Enriching Software Process Improvement with Knowledge Management
22. Annemette Skot-Hansen
Franske adjektivisk afledte adverbier, der tager præpositionssyntagmer indledt med præpositionen à som argumenter
En valensgrammatisk undersøgelse
23. Line Gry Knudsen
Collaborative R&D Capabilities
In Search of Micro-Foundations

24. Christian Scheuer
*Employers meet employees
Essays on sorting and globalization*
25. Rasmus Johnsen
*The Great Health of Melancholy
A Study of the Pathologies of Performativity*
26. Ha Thi Van Pham
*Internationalization, Competitiveness
Enhancement and Export Performance
of Emerging Market Firms:
Evidence from Vietnam*
27. Henriette Balieu
*Kontrollbegrebets betydning for kausalalternationen i spansk
En kognitiv-typologisk analyse*
- 2010**
1. Yen Tran
*Organizing Innovation in Turbulent
Fashion Market
Four papers on how fashion firms create and appropriate innovation value*
2. Anders Raastrup Kristensen
*Metaphysical Labour
Flexibility, Performance and Commitment in Work-Life Management*
3. Margrét Sigrún Sigurdardóttir
*Dependently independent
Co-existence of institutional logics in the recorded music industry*
4. Ásta Dis Óladóttir
*Internationalization from a small domestic base:
An empirical analysis of Economics and Management*
5. Christine Secher
E-deltagelse i praksis – politikernes og forvaltningens medkonstruktion og konsekvenserne heraf
6. Marianne Stang Våland
What we talk about when we talk about space:
7. Rex Degnegaard
*Strategic Change Management
Change Management Challenges in the Danish Police Reform*
8. Ulrik Schultz Brix
*Værdi i rekruttering – den sikre beslutning
En pragmatisk analyse af perception og synliggørelse af værdi i rekrutterings- og udvælgelsesarbejdet*
9. Jan Ole Similå
*Kontraktsledelse
Relasjonen mellom virksomhetsledelse og kontraktshåndtering, belyst via fire norske virksomheter*
10. Susanne Boch Waldorff
Emerging Organizations: In between local translation, institutional logics and discourse
11. Brian Kane
*Performance Talk
Next Generation Management of Organizational Performance*
12. Lars Ohnemus
*Brand Thrust: Strategic Branding and Shareholder Value
An Empirical Reconciliation of two Critical Concepts*
13. Jesper Schlamovitz
Håndtering af usikkerhed i film- og byggeprojekter
14. Tommy Moesby-Jensen
*Det faktiske livs forbindtlighed
Føroskratisk informeret, ny-aristotelisk ἦθος-tænkning hos Martin Heidegger*
15. Christian Fich
*Two Nations Divided by Common Values
French National Habitus and the Rejection of American Power*

16. Peter Beyer
Processer, sammenhængskraft og fleksibilitet
Et empirisk casestudie af omstillingsforløb i fire virksomheder
17. Adam Buchhorn
Markets of Good Intentions
Constructing and Organizing Biogas Markets Amid Fragility and Controversy
18. Cecilie K. Moesby-Jensen
Social læring og fælles praksis
Et mixed method studie, der belyser læringskonsekvenser af et lederkursus for et praksisfællesskab af offentlige mellemledere
19. Heidi Boye
Fødevarer og sundhed i sen-modernismen
– En indsigt i hyggefænomenet og de relaterede fødevarerpraksisser
20. Kristine Munkgård Pedersen
Flygtige forbindelser og midlertidige mobiliseringer
Om kulturel produktion på Roskilde Festival
21. Oliver Jacob Weber
Causes of Intercompany Harmony in Business Markets – An Empirical Investigation from a Dyad Perspective
22. Susanne Ekman
Authority and Autonomy
Paradoxes of Modern Knowledge Work
23. Anette Frey Larsen
Kvalitetsledelse på danske hospitaler
– Ledelsernes indflydelse på introduktion og vedligeholdelse af kvalitetsstrategier i det danske sundhedsvæsen
24. Toyoko Sato
Performativity and Discourse: Japanese Advertisements on the Aesthetic Education of Desire
25. Kenneth Brinch Jensen
Identifying the Last Planner System
Lean management in the construction industry
26. Javier Busquets
Orchestrating Network Behavior for Innovation
27. Luke Patey
The Power of Resistance: India's National Oil Company and International Activism in Sudan
28. Mette Vedel
Value Creation in Triadic Business Relationships. Interaction, Interconnection and Position
29. Kristian Tørning
Knowledge Management Systems in Practice – A Work Place Study
30. Qingxin Shi
An Empirical Study of Thinking Aloud
Usability Testing from a Cultural Perspective
31. Tanja Juul Christiansen
Corporate blogging: Medarbejderes kommunikative handlekraft
32. Malgorzata Ciesielska
Hybrid Organisations.
A study of the Open Source – business setting
33. Jens Dick-Nielsen
Three Essays on Corporate Bond Market Liquidity
34. Sabrina Speiermann
Modstandens Politik
Kampagnestyling i Velfærdsstaten.
En diskussion af trafikkampagners styringspotentiale
35. Julie Uldam
Fickle Commitment. Fostering political engagement in 'the flighty world of online activism'

36. Annegrete Juul Nielsen
Traveling technologies and transformations in health care
37. Athur Mühlen-Schulte
Organising Development Power and Organisational Reform in the United Nations Development Programme
38. Louise Rygaard Jonas
Branding på butiksgulvet Et case-studie af kultur- og identitetsarbejdet i Kvickly
8. Ole Helby Petersen
Public-Private Partnerships: Policy and Regulation – With Comparative and Multi-level Case Studies from Denmark and Ireland
9. Morten Krogh Petersen
'Good' Outcomes. Handling Multiplicity in Government Communication
10. Kristian Tangsgaard Hvelplund
Allocation of cognitive resources in translation - an eye-tracking and key-logging study

2011

1. Stefan Fraenkel
Key Success Factors for Sales Force Readiness during New Product Launch A Study of Product Launches in the Swedish Pharmaceutical Industry
2. Christian Plesner Rossing
International Transfer Pricing in Theory and Practice
3. Tobias Dam Hede
Samtalekunst og ledelsesdisciplin – en analyse af coachingsdiskursens genealogi og governmentality
4. Kim Pettersson
Essays on Audit Quality, Auditor Choice, and Equity Valuation
5. Henrik Merkelsen
The expert-lay controversy in risk research and management. Effects of institutional distances. Studies of risk definitions, perceptions, management and communication
6. Simon S. Torp
Employee Stock Ownership: Effect on Strategic Management and Performance
7. Mie Harder
Internal Antecedents of Management Innovation
11. Moshe Yonatany
The Internationalization Process of Digital Service Providers
12. Anne Vestergaard
Distance and Suffering Humanitarian Discourse in the age of Mediatization
13. Thorsten Mikkelsen
Personlighedens indflydelse på forretningsrelationer
14. Jane Thostrup Jagd
Hvorfor fortsætter fusionsbølgen udover "the tipping point"? – en empirisk analyse af information og kognitioner om fusioner
15. Gregory Gimpel
Value-driven Adoption and Consumption of Technology: Understanding Technology Decision Making
16. Thomas Stengade Sønderkov
Den nye mulighed Social innovation i en forretningsmæssig kontekst
17. Jeppe Christoffersen
Donor supported strategic alliances in developing countries
18. Vibeke Vad Baunsgaard
Dominant Ideological Modes of Rationality: Cross functional

- integration in the process of product innovation*
19. Throstur Olaf Sigurjonsson
Governance Failure and Iceland's Financial Collapse
 20. Allan Sall Tang Andersen
Essays on the modeling of risks in interest-rate and inflation markets
 21. Heidi Tscherning
Mobile Devices in Social Contexts
 22. Birgitte Gorm Hansen
Adapting in the Knowledge Economy Lateral Strategies for Scientists and Those Who Study Them
 23. Kristina Vaarst Andersen
Optimal Levels of Embeddedness The Contingent Value of Networked Collaboration
 24. Justine Grøn bæk Pors
Noisy Management A History of Danish School Governing from 1970-2010
 25. Stefan Linder
Micro-foundations of Strategic Entrepreneurship Essays on Autonomous Strategic Action
 26. Xin Li
Toward an Integrative Framework of National Competitiveness An application to China
 27. Rune Thorbjørn Clausen
Værdifuld arkitektur Et eksplorativt studie af bygningers rolle i virksomheders værdiskabelse
 28. Monica Viken
Markedsundersøkelser som bevis i varemerke- og markedsføringsrett
 29. Christian Wymann
Tattooing The Economic and Artistic Constitution of a Social Phenomenon
 30. Sanne Frandsen
Productive Incoherence A Case Study of Branding and Identity Struggles in a Low-Prestige Organization
 31. Mads Stenbo Nielsen
Essays on Correlation Modelling
 32. Ivan Häuser
Følelse og sprog Etablering af en ekspressiv kategori, eksemplificeret på russisk
 33. Sebastian Schwenen
Security of Supply in Electricity Markets
- 2012**
1. Peter Holm Andreasen
The Dynamics of Procurement Management - A Complexity Approach
 2. Martin Haulrich
Data-Driven Bitext Dependency Parsing and Alignment
 3. Line Kirkegaard
Konsulenten i den anden nat En undersøgelse af det intense arbejdsliv
 4. Tonny Stenheim
Decision usefulness of goodwill under IFRS
 5. Morten Lind Larsen
Produktivitet, vækst og velfærd Industrirådet og efterkrigstidens Danmark 1945 - 1958
 6. Petter Berg
Cartel Damages and Cost Asymmetries
 7. Lynn Kahle
Experiential Discourse in Marketing A methodical inquiry into practice and theory
 8. Anne Roelsgaard Obling
Management of Emotions in Accelerated Medical Relationships

9. Thomas Frandsen
Managing Modularity of Service Processes Architecture
10. Carina Christine Skovmøller
*CSR som noget særligt
Et casestudie om styring og menings-skabelse i relation til CSR ud fra en intern optik*
11. Michael Tell
*Fradragsbeskæring af selskabers finansieringsudgifter
En skatteretlig analyse af SEL §§ 11, 11B og 11C*
12. Morten Holm
*Customer Profitability Measurement Models
Their Merits and Sophistication across Contexts*
13. Katja Joo Dyppe
*Beskatning af derivater
En analyse af dansk skatteret*
14. Esben Anton Schultz
*Essays in Labor Economics
Evidence from Danish Micro Data*
15. Carina Risvig Hansen
"Contracts not covered, or not fully covered, by the Public Sector Directive"
16. Anja Svejgaard Pors
Iværksættelse af kommunikation - patientfigurer i hospitalets strategiske kommunikation
17. Frans Bévort
*Making sense of management with logics
An ethnographic study of accountants who become managers*
18. René Kallestrup
The Dynamics of Bank and Sovereign Credit Risk
19. Brett Crawford
*Revisiting the Phenomenon of Interests in Organizational Institutionalism
The Case of U.S. Chambers of Commerce*
20. Mario Daniele Amore
Essays on Empirical Corporate Finance
21. Arne Stjernholm Madsen
The evolution of innovation strategy Studied in the context of medical device activities at the pharmaceutical company Novo Nordisk A/S in the period 1980-2008
22. Jacob Holm Hansen
*Is Social Integration Necessary for Corporate Branding?
A study of corporate branding strategies at Novo Nordisk*
23. Stuart Webber
Corporate Profit Shifting and the Multinational Enterprise
24. Helene Ratner
*Promises of Reflexivity
Managing and Researching Inclusive Schools*
25. Therese Strand
The Owners and the Power: Insights from Annual General Meetings
26. Robert Gavin Strand
In Praise of Corporate Social Responsibility Bureaucracy
27. Nina Sormunen
*Auditor's going-concern reporting
Reporting decision and content of the report*
28. John Bang Mathiasen
*Learning within a product development working practice:
- an understanding anchored in pragmatism*
29. Philip Holst Riis
Understanding Role-Oriented Enterprise Systems: From Vendors to Customers
30. Marie Lisa Dacanay
*Social Enterprises and the Poor
Enhancing Social Entrepreneurship and Stakeholder Theory*

31. Fumiko Kano Glückstad
Bridging Remote Cultures: Cross-lingual concept mapping based on the information receiver's prior-knowledge
32. Henrik Barslund Fosse
Empirical Essays in International Trade
33. Peter Alexander Albrecht
*Foundational hybridity and its reproduction
Security sector reform in Sierra Leone*
34. Maja Rosenstock
*CSR - hvor svært kan det være?
Kulturanalytisk casestudie om udfordringer og dilemmaer med at forankre Coops CSR-strategi*
35. Jeanette Rasmussen
*Tweens, medier og forbrug
Et studie af 10-12 årige danske børns brug af internettet, opfattelse og forståelse af markedsføring og forbrug*
36. Ib Tunby Gulbrandsen
*'This page is not intended for a US Audience'
A five-act spectacle on online communication, collaboration & organization.*
37. Kasper Aalling Teilmann
Interactive Approaches to Rural Development
38. Mette Mogensen
*The Organization(s) of Well-being and Productivity
(Re)assembling work in the Danish Post*
39. Søren Friis Møller
*From Disinterestedness to Engagement
Towards Relational Leadership In the Cultural Sector*
40. Nico Peter Berhausen
Management Control, Innovation and Strategic Objectives – Interactions and Convergence in Product Development Networks
41. Balder Onarheim
*Creativity under Constraints
Creativity as Balancing 'Constrainedness'*
42. Haoyong Zhou
Essays on Family Firms
43. Elisabeth Naima Mikkelsen
*Making sense of organisational conflict
An empirical study of enacted sense-making in everyday conflict at work*
- 2013**
1. Jacob Lyngsie
Entrepreneurship in an Organizational Context
2. Signe Groth-Brodersen
*Fra ledelse til selvet
En socialpsykologisk analyse af forholdet imellem selvledelse, ledelse og stress i det moderne arbejdsliv*
3. Nis Høyrup Christensen
Shaping Markets: A Neoinstitutional Analysis of the Emerging Organizational Field of Renewable Energy in China
4. Christian Edelvold Berg
*As a matter of size
THE IMPORTANCE OF CRITICAL MASS AND THE CONSEQUENCES OF SCARCITY FOR TELEVISION MARKETS*
5. Christine D. Isakson
*Coworker Influence and Labor Mobility
Essays on Turnover, Entrepreneurship and Location Choice in the Danish Maritime Industry*
6. Niels Joseph Jerne Lennon
*Accounting Qualities in Practice
Rhizomatic stories of representational faithfulness, decision making and control*
7. Shannon O'Donnell
*Making Ensemble Possible
How special groups organize for collaborative creativity in conditions of spatial variability and distance*

8. Robert W. D. Veitch
Access Decisions in a Partly-Digital World
Comparing Digital Piracy and Legal Modes for Film and Music
9. Marie Mathiesen
Making Strategy Work
An Organizational Ethnography
10. Arisa Shollo
The role of business intelligence in organizational decision-making
11. Mia Kaspersen
The construction of social and environmental reporting
12. Marcus Møller Larsen
The organizational design of offshoring
13. Mette Ohm Rørdam
EU Law on Food Naming
The prohibition against misleading names in an internal market context
14. Hans Peter Rasmussen
GIV EN GED!
Kan giver-idealtyper forklare støtte til velgørenhed og understøtte relationsopbygning?
15. Ruben Schachtenhaufen
Fonetisk reduktion i dansk
16. Peter Koerver Schmidt
Dansk CFC-beskatning
I et internationalt og komparativt perspektiv
17. Morten Froholdt
Strategi i den offentlige sektor
En kortlægning af styringsmæssig kontekst, strategisk tilgang, samt anvendte redskaber og teknologier for udvalgte danske statslige styrelser
18. Annette Camilla Sjørup
Cognitive effort in metaphor translation
An eye-tracking and key-logging study
19. Tamara Stucchi
The Internationalization of Emerging Market Firms: A Context-Specific Study
20. Thomas Lopdrup-Hjorth
"Let's Go Outside": The Value of Co-Creation
21. Ana Alačovska
Genre and Autonomy in Cultural Production
The case of travel guidebook production
22. Marius Gudmand-Høyer
Stemningsindssygdommens historie i det 19. århundrede
Omtydningen af melankolien og manien som bipolarere stemningslidelser i dansk sammenhæng under hensyn til dannelsen af det moderne følelselivs relative autonomi.
En problematiserings- og erfarings-analytisk undersøgelse
23. Lichen Alex Yu
Fabricating an S&OP Process
Circulating References and Matters of Concern
24. Esben Alfort
The Expression of a Need
Understanding search
25. Trine Pallesen
Assembling Markets for Wind Power
An Inquiry into the Making of Market Devices
26. Anders Koed Madsen
Web-Visions
Repurposing digital traces to organize social attention
27. Lærke Højgaard Christiansen
BREWING ORGANIZATIONAL RESPONSES TO INSTITUTIONAL LOGICS
28. Lærke Højgaard Christiansen
EGENTLIG SELVLEDELSE
En ledelsesfilosofisk afhandling om selvledelsens paradoksale dynamik og eksistentielle engagement

29. Morten Rossing
*Local Adaption and Meaning Creation
in Performance Appraisal*

TITLER I ATV PH.D.-SERIEN

1992

1. Niels Kornum
Servicesamkørsel – organisation, økonomi og planlægningsmetode

1995

2. Verner Worm
*Nordiske virksomheder i Kina
Kulturspecifikke interaktionsrelationer ved nordiske virksomhedsetableringer i Kina*

1999

3. Mogens Bjerre
*Key Account Management of Complex Strategic Relationships
An Empirical Study of the Fast Moving Consumer Goods Industry*

2000

4. Lotte Darsø
*Innovation in the Making
Interaction Research with heterogeneous Groups of Knowledge Workers creating new Knowledge and new Leads*

2001

5. Peter Hobolt Jensen
*Managing Strategic Design Identities
The case of the Lego Developer Network*

2002

6. Peter Lohmann
The Deleuzian Other of Organizational Change – Moving Perspectives of the Human
7. Anne Marie Jess Hansen
To lead from a distance: The dynamic interplay between strategy and strategizing – A case study of the strategic management process

2003

8. Lotte Henriksen
*Videndeling
– om organisatoriske og ledelsesmæssige udfordringer ved videndeling i praksis*
9. Niels Christian Nickelsen
Arrangements of Knowing: Coordinating Procedures Tools and Bodies in Industrial Production – a case study of the collective making of new products

2005

10. Carsten Ørts Hansen
Konstruktion af ledelsesteknologier og effektivitet

TITLER I DBA PH.D.-SERIEN

2007

1. Peter Kastrup-Misir
Endeavoring to Understand Market Orientation – and the concomitant co-mutation of the researched, the researcher, the research itself and the truth

2009

1. Torkild Leo Thellefsen
*Fundamental Signs and Significance effects
A Semeiotic outline of Fundamental Signs, Significance-effects, Knowledge Profiling and their use in Knowledge Organization and Branding*
2. Daniel Ronzani
When Bits Learn to Walk Don't Make Them Trip. Technological Innovation and the Role of Regulation by Law in Information Systems Research: the Case of Radio Frequency Identification (RFID)

2010

1. Alexander Carnera
*Magten over livet og livet som magt
Studier i den biopolitiske ambivalens*