

Rigorous, transparent, and eye-catching:
Exploring the universalistic parameters of impactful
theory building in management

A dissertation presented by
Lakshmi Balachandran Nair

Supervised by
Prof. Michael Gibbert

Submitted to the
Faculty of Communication Sciences
Università della Svizzera italiana

For the degree of
Ph.D. in Communication Sciences

February 2016

Board

Prof. Michael Gibbert, Università della Svizzera italiana, Switzerland

Prof. Rebecca Piekkari, Aalto University, Finland

Prof. Luis Araujo, The University of Manchester, United Kingdom

Dedicated to

my parents, for introducing me to the world of words;

my husband, for his unconditional support and love;

my in-laws, for always believing in me;

my supervisor, for his advice and encouragement. I would not be here without him.

Co-author statement from first author

I hereby declare that I am aware that the work in the paper/manuscript of which I am a co-author, would form a part of the PhD thesis of Ms. Lakshmi Balachandran Nair, who contributed to the work in the research phase in the following way:

Title: What passes as rigorous replication logic in case study research? (and earlier versions of the paper with different headings, all of them dealing with replication logic, internal and external validity)

major proportional minor

and who contributed to the work in the writing phase in the following way:

Title: What passes as rigorous replication logic in case study research? (and earlier versions of the paper with different headings, all of them dealing with replication logic, internal and external validity)

major proportional minor

Signature:



Name: Prof. Michael Gibbert,

Dean, Faculty of Communication Sciences

Date: 04.02.2016

Co-author statement from first author

I hereby declare that I am aware that the work in the paper/manuscript of which I am a co-author, would form a part of the PhD thesis of Ms. Lakshmi Balachandran Nair, who contributed to the work in the research phase in the following way:

Title: "Oops, I've got an outlier in my data – what now?" A call for using model mis-fit as an opportunity for theory building in organization science (and earlier versions of the paper with different headings, all of them dealing with outliers in qualitative and quantitative management field research)

major proportional minor

and who contributed to the work in the writing phase in the following way:

Title: "Oops, I've got an outlier in my data – what now?" A call for using model mis-fit as an opportunity for theory building in organization science (and earlier versions of the paper with different headings, all of them dealing with outliers in qualitative and quantitative management field research)

major proportional minor

And the book we have co-authored together,

Book Title: Strategic Innovation. The definitive guide to outlier strategies

to which she has contributed to the research phase in the following way:

major proportional minor

and the writing phase in the following way:

major proportional minor

Signature:



Name: Prof. Michael Gibbert,

Dean, Faculty of Communication Sciences

Date: 04.02.2016



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Munich School of Management
Institute for Leadership and Organization (ILO)



Co-author statement from first author

I hereby declare that I am aware that the work in the paper/manuscript of which I am a co-author, would form a part of the PhD thesis of Ms. Lakshmi Balachandran Nair, who contributed to the work in the research phase in the following way:

Title: What passes as a transparent field study in management? An analysis of qualitative and quantitative research in top management journals (and earlier versions of the paper with different headings, all of them dealing with transparency in management field research)

major proportional minor

and who contributed to the work in the writing phase in the following way:

Title: What passes as a transparent field study in management? An analysis of qualitative and quantitative research in leading management journals (and earlier versions of the paper with different headings, all of them dealing with transparency in management field research)

major proportional minor

Signature: 

Name: Dr. Matthias Weiss, Assistant Professor

Date: 12.12.2015

Abstract

In the management discipline, scholarly impact is most commonly measured using a researcher perspective, by counting the number of times a particular article is mentioned in the references section of other articles (Aguinis, Shapiro, Antonacopoulou, and Cummings, 2014). This approach conceptualizes scholarly impact using a measurable indicator, the citation count an article receives. Several studies have been conducted to examine what drives scholarly impact in the field of management. The originality of the idea, rigor of the study, and clarity of writing have been identified as the most significant universalistic parameters of scholarly impact (Judge, Colbert, Cable, and Rynes, 2007). This dissertation sets out to do a detailed examination of these parameters, with a series of research articles. The six articles included in the thesis do so in two ways: either by offering recommendations for improving these universalistic parameters of scholarly impact or by further exploring the relationship between the universalistic parameters and scholarly impact. The majority of the articles included in this thesis focus either predominantly, and sometimes exclusively on qualitative research.

Our first empirical article, here relayed in Chapter II, focuses on case studies (research designs), and analyzes the methodological rigor of all case studies published during the period 1996-2006. From prior literature, we understand that in management discipline replication logic is either not used at all, or underutilized, i.e. used mainly as a remedy for external validity, at the expense of the more fundamental internal validity. This paper helps to clarify these “misconceptions” (Tsang and Williams, 2012) in the debate surrounding external validity as seemingly the only criterion worth worrying about in case study research. We point out different types of replication logic, and illustrate how their individual research actions have differential

effects on the internal and external validity (in that order of priority) of the emerging theory.

Chapter III follows up on the previous chapter, extending the investigation to quantitative as well as qualitative research, and offers replication logic as a tool for analyzing deviant cases or outliers identified during the course of a qualitative or quantitative study. We call this technique the ‘Deviant Case Method’ (‘DCM’). Previous methodological work has shown that quantitative studies (often firmly set on theory testing and with confirmatory objectives), routinely ignore outliers (Aguinis, Gottfredson, and Joo, 2013; Lewin, 1992), treat them as noise (Lieberman, 2005), sweep them under the proverbial carpet of model fit, and eventually refrain from analyzing them further. These practices are usually either based on the assumption that the outliers are due to errors (measurement or sampling), or done in an effort to tie up “loose ends” and improve statistical power (Kendall and Wolf, 1949; Pearce, 2002). Similarly, qualitative studies typically focus on similarities between cases in a replication logic, rather than cases that do not fit the patterns (Blatter and Haverland, 2012; Eisenhardt, 1989; Gibbert, Ruigrok, and Wicki, 2008). Through this study, we explain the theoretical consequentiality (Aguinis et al., 2013; Cortina, 2002) of analyzing three different kinds of outliers (construct, model fit, and prediction outliers/ deviant cases) and offer DCM for analyzing prediction outliers/deviant cases. Depending on their number and position in the scatterplot, different deviant cases will accommodate different methods of DCM, with varying degrees of theoretical gains.

In Chapter IV, we extend this method to have a look at medium-N studies. Here we focus on inconsistent or deviant cases which turn up during a particular variant of Qualitative Comparative Analysis, namely fuzzy set Qualitative Comparative Analysis (fsQCA). Unlike in qualitative and quantitative studies, the inconsistent cases of QCA usually differ depending

upon whether they turn up during an analysis of ‘necessity’ or ‘sufficiency’. Depending upon their position in the theoretical prediction, they are again subdivided into ‘deviant cases of consistency (degree)’, ‘deviant cases of consistency (kind)’, and ‘deviant cases of coverage’. We offer a method called ‘Comparative Outlier Analysis’ (‘COA’) which combines DCM and Mill’s canons (1875) to examine these multitude of inconsistent cases. Since suitable examples from management are not available, we explicate this using exemplars from fields like politics, marketing, and education. Chapters II, III and IV contribute to the discussion on rigor and novelty of ideas (theory generation).

In order to assess the rigor of research procedures (whether qualitative or quantitative), to appreciate a study’s results, and to want to integrate them into the body of scholarly knowledge, they need to be transparently relayed. Unlike in other disciplines or methods, it is far from clear what the label ‘transparent research procedures’ constitutes in management field studies, with adverse effects during write-up, revision, and even after publication. To rectify this, in Chapter V, we review field studies (again both qualitative as well as quantitative) across seven major management journals (1997-2006) in order to develop a transparency index, and link it to article impact.

Chapter VI is a sequel to the previous chapter. It focuses on the auditability of a popular qualitative research method, grounded theory research. Iterative methods like grounded theory rely on several cycles of comparing empirical phenomena with theoretical inclinations until saturation is reached. We propose a new method for assessing the methodological rigor of these procedures ex-post using an audit trail perspective. While existing research on the methodological sophistication of grounded theory was typically done from the perspective of the author or producer of the research, our perspective is customer-centric, both in terms of the end-customer (i.e.

the reader or other author), as well as the intermediate customer (i.e. reviewers and editors). This perspective is helpful in that it focuses attention on the concrete research actions which need to be transparently relayed in the final manuscript, so that an article's chain of evidence can be conveniently assessed *ex post*.

The last empirical article in the thesis, Chapter VII, focuses on yet another parameter influencing impact: the style of academic writing. Specifically, we focus on the attributes of article titles and their subsequent influence on the citation count. The question, what makes a "good title", i.e. one which produces scholarly impact via citations in the academic community, has so far not provided conclusive results. Our study is a first step towards building an overall model of article characteristics and linking them with impact. For the same, we create a coding sheet of title attributes, compiled from previous studies in other disciplines. At this early stage of theory development on article titles, we do this in the specific application context of management science.

Overall, that is, via its single papers as well as in its entirety, the current thesis makes a major contribution to the discourse on generating and developing novel solutions for addressing the paucity of rigor, transparency, and clarity of reporting in the field of management. We demonstrate the prominence of the involved research inquiries by illustrating their contributions with the help of exemplars (from other disciplines where these techniques are more evolved) or by indicating their relationship with scholarly impact in management. We conclude with Chapter VIII where we sum up the findings and implications of all preceding studies and put forth suggestions for future research.

Acknowledgements

First and foremost, I would like to thank Prof. Michael Gibbert, my supervisor, for giving me the opportunity to pursue my PhD on this interesting topic. I am passionate about this topic and I owe it all to you. Over the past years, you have seen me through each stage of the PhD process, encouraged my ideas, and provided invaluable advice both for my research as well as my career. You have been a wonderful mentor and a perfect role model to me, encouraging my research and allowing me to grow as a researcher.

I would like to extend my thanks to Institute of Marketing and Communication Management (IMCA) and Università della Svizzera italiana, for providing me all the facilities to conduct my research effectively. I am very grateful to Prof. Luis Araujo and Prof. Rebecca Piekkari for serving as my committee members. Thank you for accommodating my defense into your busy schedules. I thank Prof. Martin Hoegl and Prof. Liisa Valikangas for being great mentors during my visiting scholarship at Ludwig-Maximilians-Universität München, Germany and Hanken School of Economics, Finland respectively. It has been a great pleasure working with both teams.

On the same note, I extend my greatest gratitude to Swiss National Science Foundation for granting me the Docmobility funding to spend a year abroad in Germany. This stay has greatly enhanced my exposure to academic practices outside Switzerland, and helped me to work closely with one of the best institutes in my field. I thank USI Equal Opportunities Service for funding my stay in Finland. It was a great opportunity to collaborate with the esteemed academic team on publishing our book. I would also like to thank my co-authors, research partners, and colleagues in both these institutions. You made me feel at home and provided unconditional help and support.

I appreciate my wonderful colleagues and friends at USI, whose presence in my life made my PhD life quite enjoyable and productive. My special thanks to Dr. Matthias Weiss, Salvatore Maione, and Jost Reinhold for helping me understand the nuances of quantitative research methodology. I thank my beloved work friends (Rukhshona, Kubra, Monica, and Helene) for always being there for me. And my delightful friends outside work (too many in number to list out by name, but you know who you all are), you are like my family. Thank you for never letting me forget what a wonderful life it is.

Speaking of family, I thank my parents who instilled an interest for research in me. My gratitude to my in-laws for having me in their prayers always. Thank you to my dearest young brother Sreehari Balachandran Nair, my uncles and aunts, my grandmothers, and all my cousins for encouraging me to fulfill my ambitions. I saved the best for the last- I am deeply grateful to my husband Harikrishna Nair, for cherishing every moment and every achievement of my PhD life with me, while being immensely encouraging. I am indeed blessed to have you in my life.

Table of contents

Co-author statements	3
Abstract	9
Acknowledgements	13
Index of tables	18
Index of figures	20
Chapter I	
Introduction	21
1.1. Scholarly impact in management research	22
1.2. Replication and replicability in management research	25
1.3. Main objectives and theoretical framework	27
1.4. Research questions and organization of the study	28
PART ONE. IDEA AND RIGOR	35
Chapter II	
What passes as rigorous replication logic in case study research?	37
2.1. Abstract	38
2.2. Introduction	39
2.3. A roadmap for rigorous replication logic in case study research	41
2.4. Methodology	53
2.5. Results	56
2.6. Discussion and conclusion	58
2.7. Suggestions for further research	61
Chapter III	
“Oops, I’ve got an outlier in my data – what now?” A call for using model mis-fit as an opportunity for theory building in organization science	63
3.1. Abstract	66
3.2. Introduction	67
3.3. Outliers as an attractive nuisance?	69
3.4. Deviant case analysis: How to use prediction outliers for theory building	76
3.5. Methodology	80
3.6. Discussion of results: theory-building potential realized vs . foregone	84
3.7. General discussion	101
3.8. Conclusion	106
Chapter IV	

Analyzing inconsistent cases in management fsQCA studies: A methodological manifesto	107
4.1. Abstract	108
4.2. Introduction	109
4.3. Case oriented strategies in QCA	110
4.4. Why analyze the inconsistent cases in QCA?	110
4.5. COA- How to analyze outliers in a QCA	112
4.6. Analysis of multiple outliers in fsQCA	113
4.7. Concrete examples of analysis of deviant cases using COA	121
4.8. Discussion	124
4.9. Limitations	125
4.10. Conclusion	126
PART TWO. RIGOR AND WRITING	127
Chapter V	
What passes as a transparent field study in management? An analysis of qualitative and quantitative research in top management journals	129
5.1. Abstract	130
5.2. Introduction	131
5.3. Operationalizing transparency	133
5.4. Methods	140
5.5. Results	148
5.6. Discussion	157
5.7. Limitations	164
Chapter VI	
Hot on the audit trail: How to assess methodological transparency of grounded theory in management?	165
6.1. Abstract	166
6.2. Introduction	167
6.3. Auditability in qualitative research	169
6.4. Auditability of research procedures: In the eye of the second-party?	172
6.5. Methods (general)	178
6.6. Main study	179
6.7. Results	181
6.8. Discussion	183
6.9. Theoretical implications and future directions	186
6.10. Conclusion	187
6.11. Appendix A: Pilot study	190
PART THREE. WRITING	193

Chapter VII	
Judging by its title: The link between article title attributes and citation count in management research	195
7.1. Abstract	196
7.2. Introduction	197
7.3. Article title attributes	199
7.4. Methods	204
7.5. Results	209
7.6. Discussion	218
7.7. Limitations and conclusion	223
Chapter VIII	
Conclusion	225
8.1. General summary of the findings	226
8.2. Implications of the findings and suggestions for future research	230
References	235

Index of tables

PART ONE. IDEA AND RIGOR

Chapter II	What passes as rigorous replication logic in case study research?	
	2.1. Comparison of different replication logic designs	43
	2.2. Main research actions involved in rigorous replication logic	44
	2.3. Replication logic and research strategies undertaken: Examples	55
	2.4. Prevalence of empirical case studies in selected journals	57
	2.5. Share of empirical articles applying the three replication logic designs	57
	2.6. Results for hypotheses 1-6	58
	2.7. Results for external validity indicators	59
Chapter III	“Oops, I’ve got an outlier in my data – what now?” A call for using model mis-fit as an opportunity for theory building in organization science	
	3.1. Different types of outliers and their theoretical heft	70
	3.2. Examples where deviant cases’ theory-building potential is fully realized	84
	3.3. Examples where deviant cases’ theory-building potential is partially realized	92
Chapter IV	Analyzing inconsistent cases in management fsQCA studies: A methodological manifesto	
	4.1. COA in fsQCA Necessity	114
	4.2. COA in fsQCA Sufficiency	115
	4.3. Examples of COA from prior literature	122

PART TWO. RIGOR AND WRITING

Chapter V	What passes as a transparent field study in management? An analysis of qualitative and quantitative research in top management journals	
	5.1. Sample description	142
	5.2. Codes for transparency of quantitative field studies	144
	5.3. Codes for transparency of qualitative field studies	145
	5.4. Reporting frequency of individual codes (percentages)	149
	5.5. Descriptive statistics and variable correlations	152

	5.6. Results of regression analyses on consequences of article transparency (article citations per year)	153
	5.7. Results of regression analyses on consequences of article transparency (total article citations)	154
	5.8. Results of regression analyses on antecedents of article transparency	156
Chapter VI	Hot on the audit trail: How to assess methodological transparency of grounded theory in management?	
	6.1. New classifications of audits	172
	6.2. Second-party audit trail parameters for grounded theory articles	174
	6.3. Reporting frequency of audit trail parameters pertaining to stages of research process	182
	6.4. Second-party audit trail in management research	183
	6.5. Second-party audit trail parameters on an exemplar article	188
	A1. Pilot study- Descriptives	191
 PART THREE. WRITING		
Chapter VII	Judging by its title: The link between article title attributes and citation count in management research	
	7.1. Article title attributes	207
	7.2. Article title attributes: Frequencies	210
	7.3. Article title attributes: Descriptives	210
	7.4. Correlations	211
	7.5. Step-wise regression analysis (ISI citation count)	213
	7.6. Step-wise regression analysis (ISI citation count per year)	214
	7.7. Step-wise regression analysis (Google scholar citation count)	215
	7.7. Step-wise regression analysis (Google scholar citation count per year)	216

Index of figures

Chapter 1	Introduction	
	1.1. Universalistic attributes of scholarly impact	25
	1.2. Outline of the dissertation	33
PART ONE. IDEA AND RIGOR		
Chapter II	What passes as rigorous replication logic in case study research?	
	2.1. Assumed relationship between the factor X and dependent variable	47
	2.2. An alternative possibility for the relationship between factor X and dependent variable	47
Chapter III	“Oops, I’ve got an outlier in my data – what now?” A call for using model mis-fit as an opportunity for theory building in organization science	
	3.1. Search process for prediction outliers (“deviant cases”)	82
Chapter IV	Analyzing inconsistent cases in management fsQCA studies: A methodological manifesto	
	4.1. Types of cases in fsQCA (Necessity)	116
	4.2. Types of cases in fsQCA (Sufficiency)	119
PART TWO. RIGOR AND WRITING		
Chapter V	What passes as a transparent field study in management? An analysis of qualitative and quantitative research in top management journals	
	5.1. Stages of the qualitative research process	138
	5.2. Plot of interaction effect	155
PART THREE. WRITING		
Chapter VII	Judging by its title: The link between article title attributes and citation count in management research	
	7.1. Plot of interaction effect	218

Chapter I*

Introduction

Lakshmi Balachandran Nair

* Conferences

Nair, L.B. (2015, June). EURAM Doctoral Colloquium, Warsaw, Poland.

1.1. Scholarly impact in management research

Management academia is passionate about creating and sustaining scholarly impact. From the early days, researchers were interested in conceptualizing and measuring their contribution to the discipline. Impactful articles document the history of an investigation, create and disseminate new knowledge, and map the future trajectory of a research area (Judge, Colbert, Cable, & Rynes, 2007). It thus has a huge effect on shaping the direction of the entire management discipline and subsequently plays a significant role in human resource management decisions such as promotion and tenure (Tahai and Meyer, 1999) as well.

One of the most common ways of measuring and quantifying scholarly impact is through the number of citations an article or a publication receives (Bergh, Perry, & Hanke, 2006; Judge et al., 2007; Aguinis, Shapiro, Antonacopoulou & Cummings, 2014). This conceptualization is as old as the discipline itself. The popularity of citation count is to the extent that it has been referred to as ‘frozen footprints on the landscape of scholarly achievement’ (Cronin, 1984). Peers, recruiting committees, and promotional review boards evaluate a scholar’s influence, recognition, and standing based on summed citation frequency counts (e.g., Garfield, 2006). As a result, management scholars are quite interested in understanding what drives scholarly impact.

The universalistic view of science considers an article cite-worthy if it offers original contributions to science (new theory generation) and exhibits high quality standards (Cole & Cole, 1972; Judge et al., 2007; Merton, 1968). An exploration of the universalistic parameters have pointed out that the two main requirements of a high citation rate are: originality of the idea and clarity of presentation of one’s work. In this dissertation, we provide strategies for original rigorous theory building as well as examine various factors pertaining to the clarity of presentation of scientific articles and their influence upon scholarly impact measured via cumulated citation counts. In section 1.3, we describe each of them in detail. Figure 1.1. shows the three universalistic predictors of scholarly impact we discuss henceforth.

Universalistic predictors of scholarly impact

Idea

Originality of an idea brings recognition and esteem to the scientists (Judge et al., 2007 citing Merton, 1957). Studies in different fields like psychology (Gottfredson, 1978), management (Beyer, Chanove, & Fox, 1995; Eisenhardt, 1989; Judge et al, 2007) has found evidence for this. Exploratory studies which look into new paradigms, bring paradigms to a new territory (Newman & Cooper, 1993), find new variables (Aguinis et al., 2013), explore puzzling observations (Schneider & Rohlfing, 2013) etc. were found to have a higher citation count than studies that are just refining or extending an existing idea. From this perspective, qualitative studies (for example, case study research) are more exploratory than quantitative studies, and are considered most appropriate in the early phases of new theory building. However, originality alone would not guarantee a higher scholarly impact. Lack of rigor, especially in the case of exploratory, inductive studies is highly problematic since relevance cannot be claimed without rigor (Scandura and Williams, 2000).

Rigor

As per the positivistic, natural science model, rigor refers to the methodological sophistication of a scientific study, i.e. its validity (internal, external, construct) and reliability. Although this model is positivistic, the tangible research actions underlying these criteria have been adapted by interpretivists as well. These research actions are comparable to the interpretivistic criterion of credibility, dependability, conformability, and transferability as well. Lack of rigor, especially in exploratory studies might be quite problematic. A deficiency of rigor in an inductive, theory building study would cause ripple effects not only in the study in hand, but also in subsequent studies which elaborate and test the findings of the former (Eisenhardt and Graebner, 2007).

The link between impact and rigor in management discipline (case study research, in particular) has been explored by prior studies (Gibbert et al., 2008; Gibbert and Ruigrok, 2010). The results showed that rigorous case studies do indeed tend to be

more impactful than their less rigorous counterparts. This prompts us to explore different techniques to assess and improve the rigor of research methods. However, what stands in the way of assessing the rigor of a study is the vagueness of reporting. Transparent or coherent writing is thus a precondition for assessing the rigor of a study.

Writing

Writing a report clearly is important for ensuring its contributions to the academic community and its good reception. Clarity or transparency of writing initially influences decisions regarding the acceptance or rejection at journals (Beyer, Chanove & Fox, 1995). Once accepted, a clear presentation of one's study has an influence on scholarly impact (Judge et al., 2007). A transparent research report walks the reader through the various stages of the argument, in an effort to make the logic, reasoning, and causalities evident (Dubois & Gibbert, 2010). A non-transparent research report curbs the chances of another scientist to replicate the procedure and verify the observations (Glass, 1965). Careful documentation and clarification of the involved research procedures (Dubois & Gibbert, 2010) would ensure comprehensibility and easy communication of the ideas to the audience (Beyer et al., 1995, Judge et al., 2007).

Another important feature of writing which might have an influence on scholarly impact is the attractiveness of presentation. An article should be able to attract the attention of the intended audience while communicating its essence (Moore, 2010). The decision to read or not read an article is based on several 'cues' (Sagi & Yechiam, 2008) concerning the way its contents are portrayed. The title of the said article is one of the most important of these cues (Hartley, 2007 b) which individualize the publication, summarize its content, and makes it appealing to the audience (Soler, 2010). Several studies in various fields other than management have explored the influence of titles in enhancing readership (Paiva, Lima, & Paiva, 2012; Subotic & Mukherjee, 2014). Together, these two features of writing (transparency and attractiveness of title cue) have been found to individually contribute to scholarly impact.

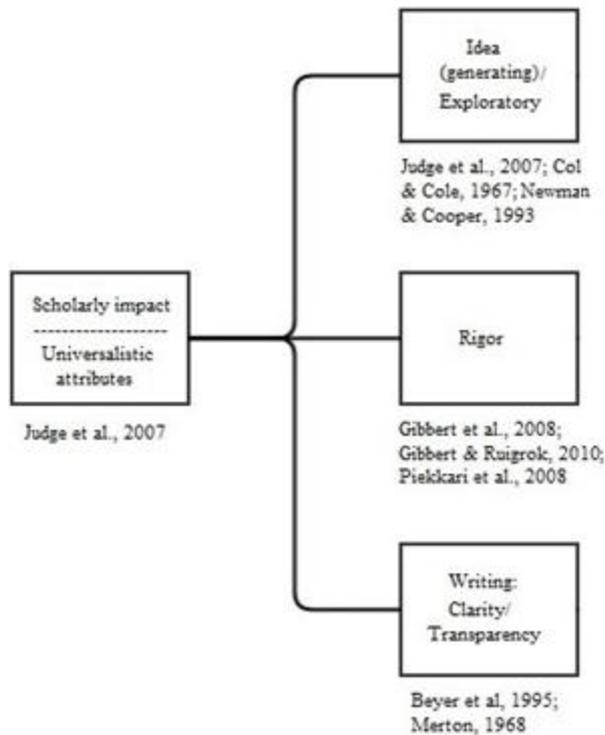


Figure 1.1. Universalistic attributes of scholarly impact

1.2. Replication and replicability in management research

Replication is often referred to as an ‘indispensable ingredient’ in the scientific process (Eden, 2002). Without replicating, it is impossible to understand whether the findings of a particular study are just “isolated coincidences” (Tsang & Kwan, 1999) or not. A replication design attempts to duplicate the results of a previous study (Berthon, Pitt, Ewing, and Carr, 2002; Rowley, 2002) and make sure the research findings and measurements are reliable. Since reliability is the most basic rigor parameter (the validity parameters cannot be established until a study is proven to be free of random error, i.e. reliable), a replication design which ensures it is considered as a cornerstone of science.

When compared to natural sciences, social sciences lag behind in replicability (Tsang & Kwan, 1999). However, recent developments in management research has seen an uncharacteristically high interest in replication and replicability. Failure to replicate the results or observations of studies have led to discovery of manipulated

results, which in turn have led to rejection and, if published already, the retraction of scholarly papers. This has prompted various top journals to become interested in the topic of replication and even invite calls for articles undertaking replication (for example, *Strategic Management Journal*, special issue ‘Replication in Strategic Management’).

Having said that, replication is not a simple process. Different researchers have suggested different types of replication designs that can be undertaken based on the study’s data set, measurement, analysis etc. (Hendrick, 1990; Tsang & Kwan, 1999). Replication also differs according to the research designs. For example, unlike in experimental studies, replication in case study research is part of the research design itself. That is, replication happens within the study, before it is completed rather than outside the study (using another data set, measurement, or analysis). This brings us to ‘replication logic’.

Replication logic

Replication logic is a practice for identifying new cases to which the finding of an original case can be replicated. It is central to theory building (Eisenhardt & Graebner, 2007) in case study research. Each new case selected would act like a new ‘experiment’ which when replicated provides confirmation (similar cases) or contradiction (dissimilar cases). A study which uses replication logic to select cases similar to the initial case undertakes ‘replication logic without difference condition’ (Eisenhardt, 1989), also known as ‘literal replication’ (Yin, 1994). ‘Replication logic with difference condition’ (Eisenhardt, 1989) otherwise known as ‘theoretical replication’ (Yin, 1994) meanwhile selects cases which predict different results, but due to predictable reasons (dissimilar cases). Depending upon whether the replication occurs within a single case or across different cases, the case study design becomes embedded single, multiple holistic, or embedded holistic (Yin, 1994; Gerring, 2007). Eisenhardt (1989) has made substantial contributions to the field of management by pondering upon the potential of replication logic to enhance external validity. In this thesis, we focus primarily on the contributions

of replication logic in enhancing internal validity and building rigorous theory (through the analysis of multiple case studies, deviant cases etc). A non-transparent research report curbs the chances of a study being completely understood and replicated. Therefore, as a next step, we discuss transparency as a factor affecting the replicability of a study.

1.3. Main objectives and theoretical framework

The main objective of the study is to explore the different universalistic predictors of scholarly impact, in the light of replication logic and replicability. As can be seen in Figure 1.2, the first three studies focus on the universalistic parameters of ‘idea’ (exploratory design, theory building prowess, multiple case studies) and ‘rigor’ (mainly internal validity, external validity). To be clear, these three studies do NOT explore the connection between scholarly impact and the universalistic parameters. Rather, they use ‘replication logic’ as a tool for increasing the rigor and theory building potential of management research.

In Chapter II, we focus on the role of replication logic in enhancing internal and external validity of case study research. In Chapter III, replication logic is used as a foundation for enhancing the theory building potential of qualitative and quantitative research, through the qualitative analysis of deviant cases. Chapter IV builds upon the earlier articles and proposes a replication logic-based design for analyzing post-QCA inconsistent cases of necessity and sufficiency. QCA is a popular research method in political science which has quite recently been introduced to the field of management (e.g., Fiss, 2007). There has been calls in political science QCA research to focus on the analysis of post-QCA deviant cases (Schneider and Rohlfing, 2013). Chapter IV of the dissertation introduces this technique to the field of management and expands it to encompass a comparative study of identified deviant cases using replication logic.

The next study (Chapter V) explores ‘transparency’ which is an aspect of the universalistic parameter ‘writing’, and a precondition for replicability and assessment of rigor. Transparency is explored using a positivistic (Chapter V) model here. We examine

whether the transparency parameters we include in the study has an influence on the scholarly impact of the corresponding articles (measured in terms of citation counts). Chapter VI follows up on this study and examines the second party auditability of grounded theory studies, borrowing the concept from audit literature. This study expands other work on auditing in grounded theory research by providing an index of audit parameters for second-party auditors (independent readers and reviewers) of published studies.

The final empirical study, Chapter (VII), explores another ‘writing’ parameter which has so far been unexplored in management: the attributes of titles. Many studies in several fields like medicine, biology, and even social sciences (psychology) have explored this topic. However, no such effort has been made in management so far. So through this study we replicate similar studies in other disciplines, and examine whether the attributes of article titles have any relationship with scholarly impact.

1.4. Research questions and organization of study

This cumulative dissertation consists of six individual studies which were carried out at the Institute of Marketing and Communication Management at Università della Svizzera italiana in Lugano between the years 2012 and 2016. All of them are either accepted for publication (2 articles) or are under review (3 articles) or are being prepared for submission (1 article) to prominent journals in the field of management or research methodology. All the papers have been presented at international conferences (either as traditional papers or as working papers). Individually, all these articles explore a distinct research question which pertains to management research methodology and scholarly impact. Together they all contribute to the main objectives: enhancing the methodological sophistication of research in management by refining existing methods and adopting new methods from other disciplines.

Idea and rigor parameters: Role of replication logic

Disciplines such as political science, biology etc. have pondered upon the importance of replication logic for enhancing the rigor of research. Prior studies have established the

role of rigor and idea generation (exploratory designs) in enhancing scholarly impact (Gibbert et al., 2008; Gibbert & Ruigrok, 2010; Judge et al., 2007). Therefore in this thesis we are not exploring this link. Rather, we are focusing on studies to improve the rigor and idea generation potential of management research. Previous studies in the field of management have explored the role of replication logic in enhancing the external validity of a study. Yet so far, replication logic has not been examined in the context of internal validity.

Research Question 1. To what extent can replication logic be used to build rigorous (internally and externally valid) theory?

To answer this question, in Chapter II we a) distinguish three types of replication logic: literal replication, theoretical replication and no replication, (b) discuss the propensity of each type for rigorous theory building, and (c) describe the boundary conditions in terms of various factors of internal validity (theoretical sampling, discussion of units of analysis, discussion of control, discussion of focal variables), and external validity (usage of multiple cases, elaboration of case study context). Finally, we check their corresponding effects on internal vs. external validity.

In Chapter III, we use a replication logic-based technique ‘Deviant Case Method’ (DCM) to explore the theory building potential of outliers. We discuss three different types of outliers with varying degrees of theoretical building potential. Consequently, we suggest three replication logic-based strategies for utilizing the outliers with the greatest theory building potential (deviant cases). Finally, we discuss their relative methodological sophistication in comparison with each other, and illustrate their theoretical prowess empirically (with reference to published qualitative and quantitative studies in management).

QCA is a relatively new method in management. Chapter IV focuses on the theory-building prowess of inconsistent or deviant cases which turn up during a fsQCA study. This chapter adopts case analysis strategies (‘most similar systems design’ and ‘most different system design’) from comparative politics, introduce them to the analysis

of post-QCA deviant cases, develop them further using a replication logic-based template, and offer them for application in management academia. Thus we propose two new strategies which we collectively term ‘Comparative Outlier Analysis’ (COA) for improving the neat theory-data fit characteristic of QCA. COA takes into consideration inconsistent cases of both necessity and sufficiency, and also incorporate the concepts of ‘method of agreement’ and ‘method of difference’ (Faure, 1994; Lijphart, 1975).

Rigor and writing: Role of transparency

The role of writing in scholarly impact has been explored by prior researchers (Judge et al., 2007; Merton, 1968). Features like clarity of presentation, limitations, implications, and number of pages have been explored in the context of citation counts. However, previous studies have not exhaustively explored various features which are to be reported in a study to ensure its transparency. There is a lack of understanding and codified parameters for ensuring transparency at each stage of an empirical study. This poses a difficulty for researchers in understanding the rigorousness of a study, integrating it into the body of scholarly knowledge, and extending it by further replications. Also, since the examined parameters of writing (previously mentioned) have an influence on the citation count, it is logical to assume that transparency of research procedures might also have an effect on scholarly impact. In this section, we focus on the following research question.

Research question 2. How are research procedures reported in top management journal articles?

In Chapter V, we theoretically develop and empirically identify a transparency index, following the approach of Aytug et al. (2012) in the domain of meta-analyses, for capturing which methodological choices and data properties are reported in qualitative and quantitative field research, and link transparency to its impact in the scholarly community (in terms of the citation count). For studying this, we investigate published papers in leading management journals over an extended period of time.

Chapter VI builds upon the previous chapter, and provides a conceptual overview of the reporting parameters particular to grounded theory research. Since Chapter V used a positivistic approach which might not be suitable for a purely interpretivistic research methodology like grounded theory, we create a coding criteria specifically suited for grounded theory method for analyzing these studies. Since grounded theory in management is considered to be in a ‘state of confusion’ (Jones and Noble, 2007), increased auditability regarding different stages of the research process would help decrease the perception that grounded theory research possesses inadequate validity and objectivity (Denk, Kaufmann, and Carter; 2012; George and Bennett, 2005), and make it easier to decipher the study’s rigor. As of now, there are several studies which explore the auditability of grounded theory studies from the researcher’s perspective as well as independent third party’s perspective. However, there is no such study which explores auditability from the perspective of readers or reviewers aka. the ‘customers’ of an article. The study included in the thesis is a starting point in this direction. Borrowing the concept from audit literature, we introduce a checklist for ensuring auditability by an article’s customers i.e. ‘second party auditability’. To validate this auditability index, we explore the reporting standards of grounded theory articles across management journals of various ranks over a short time frame. We then follow up this pilot study with a more detailed examination of second-party auditability of top management journal grounded theory articles during a four decade period (1970-2010).

Writing: Role of article titles

Chapter VII explores another aspect of writing which has not been investigated in the context of management research before: Article titles. Studies in other disciplines like psychology (Subotic and Mukherjee, 2014), information science (Diener, 1984), biomedicine (Lewison and Paraje, 2004), biology, medicine, physics (Lewison and Hartley, 2005), and marketing (Stremersch, Verniers and Verhoef, 2007) have explored the effects of title attributes on invoking readers’ interest and subsequently, citation count.

Research question 3. Does other aspects of writing, like the way an article title is formulated, influence readership and scholarly impact?

In this article, we put together attributes previously examined individually in various studies, add a few management-field related attributes, and examine their influence on the article's impact. This is the first study in management academia to explore the role of titles in facilitating readership, and also the first study to explore a wide range of title attributes in the same model.

Finally, in Chapter VIII we sum up the findings of the six chapters preceding it. We also discuss the overall implication of our studies, the theoretical contributions, limitations, and future directions for research. Figure 1.2. gives an outline of the whole dissertation.

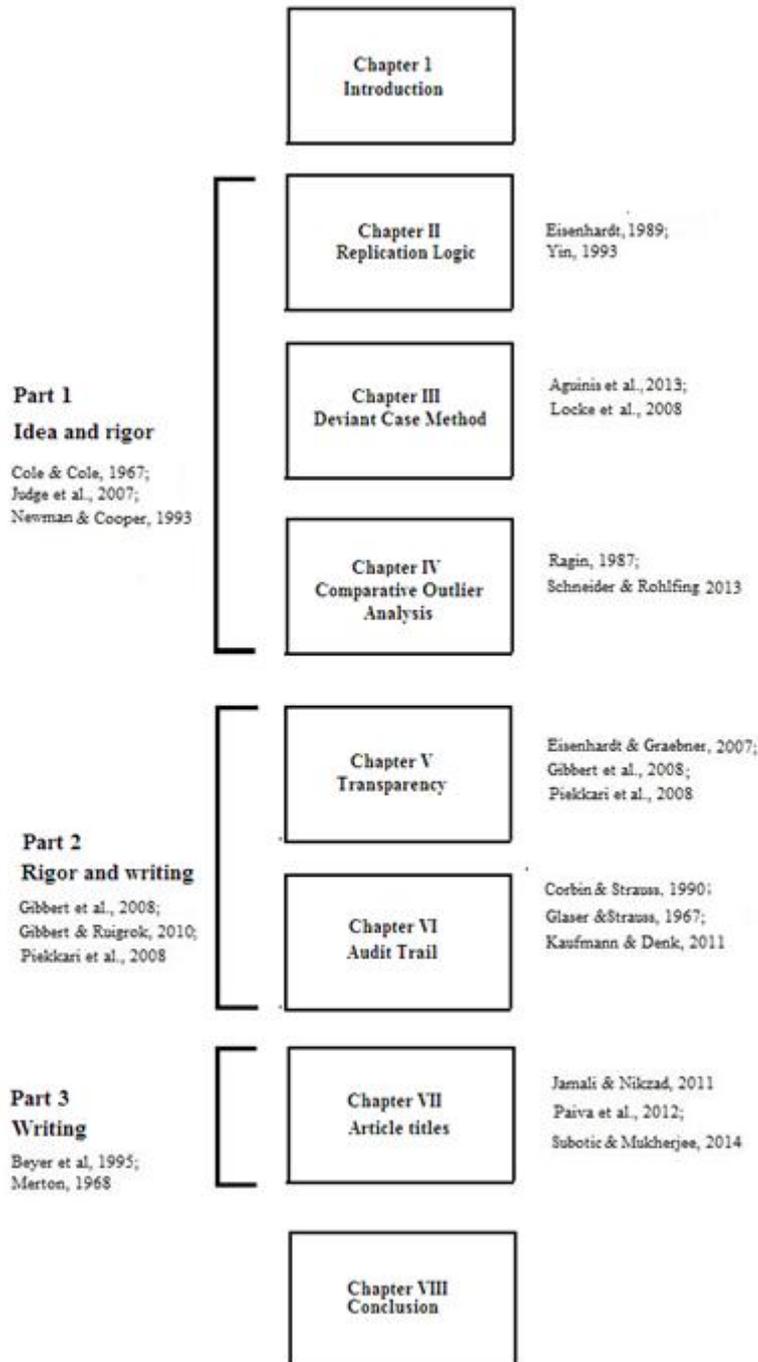


Figure 1.2. Outline of the dissertation.

PART ONE

IDEA AND RIGOR

Chapter II*

What passes as rigorous replication logic in case study research?

Prof. Michael Gibbert

Lakshmi Balachandran Nair

Prof. Winfried Ruigrok

* Journal

Strategic Management Journal (Under review)

Conferences

Gibbert, M., Nair, L.B. (2016, August). 76th Academy of Management Annual Meeting, Anaheim, USA (Awaiting decision)

Gibbert, M., Nair, L.B., & Ruigrok, W. (2015, October). Strategic Management Society Conference, Denver, USA.

Gibbert, M., Nair, L.B. (2013, August). 73rd Academy of Management Annual Meeting, Florida, USA.

Gibbert, M., Nair, L.B. (2013, June). European Academy of Management Conference, Istanbul, Turkey.

2.1. Abstract

This article focuses on research designs using multiple cases in a replication logic, and analyzes the methodological rigor of all case studies (N=184) published in the period 1996–2006 in *Academy of Management Journal*, *Administrative Science Quarterly*, *Organization Science*, *Management Science*, and *Strategic Management Journal*. In the majority of cases, replication logic is either not used at all, or underutilized, i.e. used mainly as a remedy for external validity, at the expense of the more fundamental internal validity. To rectify this, we empirically differentiate three types of replication logic, each with different degrees of methodological rigor, and discuss the boundary conditions in terms of theoretical sampling, units of analysis, controls, as well as focal variables.

Key words: *replication logic, internal validity, external validity, case study research, rigor*

2.2. Introduction

Each method comes with its own strengths and weaknesses. Strategy researchers interested in theory building need to carefully balance possible trade-offs when it comes to internal vs. external validity, and explicitly motivate the “fit” of their method (Edmondson and McManus, 2007), especially in qualitative research (Bettis, Gambardella, Helfat, Mitchell, 2014; Gibbert and Ruigrok, 2010). A popular qualitative method is case study research, and one of its key strengths is internal validity, i.e. the in-depth tracing of causal processes within or between cases and units of analysis (Gerring, 2007; March, Sproull, and Tamuz, 1991). Since case study research analyzes only one or a few cases, its corollary weakness is external validity, and so it seems only natural for authors to problematize whether an emerging theory is idiosyncratic to the case at hand, or may be repeated across multiple cases, improving generalizability (Tsang, 2014).

Using multiple case designs multiple cases in a so-called *replication logic* (Dubois and Araujo, 2007; Eisenhardt, 1989, 2014; Yin, 1981) constitutes a widely-used, though not uncontroversial, method. On the one hand, researchers maintain that replication logic enables comparisons within and between units of analysis, which clarifies whether an emerging theory is idiosyncratic to a single case or can be consistently replicated across several cases, enhancing the theory’s external validity (Eisenhardt and Graebner, 2007: 27). On the other hand, other researchers argue that the usage of replication logic seems to have evolved narrowly as a remedy for limited generalizability, often without putting it in the context of other, and more important manifestations of rigor, in particular internal validity (e.g. Campbell and Stanley, 1963; Gibbert, Ruigrok, and Wicki, 2008; Lee and Baskerville, 2003; Tsang and Williams, 2012).

The controversy surrounding replication logic and its main benefits (internal vs. external validity) stems, to our minds, from a lack of clarity regarding different types of replication logic, the main steps involved in rigorously applying replication logic, and the implications of these steps for methodological rigor. On the one hand, there are those who start with a selection of similar cases and use replication logic mainly to demonstrate that findings of one particular case are replicable across other cases. The

cases under investigation therefore essentially provide the very same insights. This interpretation of replication logic is similar to replication in a laboratory setting, where an exact repetition of the research design and procedures is expected to lead to the same results. For instance, Bradach (1997) investigated the research question, how do restaurant chains achieve the objectives of uniformity and system wide adaptation? He studied five large restaurant chains, all of which corroborate the same finding, namely that simultaneous use of company and franchise units provides control and innovation processes which reduces some of the weaknesses, and increases the strengths of the company and franchise arrangements.

On the other hand, there are those who use replication logic in a covariational sense, namely to show that changes in a theoretically-motivated cause lead to corresponding changes in some outcome variable. Here, researchers start with a selection of cases which are different, but for predictable reasons. For example, Schweizer (2005) studied the research question, how is a biotech company integrated into a pharmaceutical company which is seeking to gain access to the former's technology, innovative capabilities and know-how? Schweizer's (2005) selection of cases included extreme or polar research sites of successful and unsuccessful deals of mergers/acquisitions between biotech companies and pharmaceutical companies. Successful cases were characterized by a more sophisticated post-acquisition integration approach than unsuccessful cases.

Overall, there seems to be limited shared understanding in the community of case-study researchers regarding (1) how different types of replication logic can be used for ensuring internal vs. external validity, and (2) what rigorous replication logic entails in terms of concrete research actions or steps. Our purpose is to address these issues by providing an overview of the relevant research actions in terms of theoretical sampling, the discussion of units of analysis, controls and focal variables. We empirically demonstrate how different types of replication logic address each of these criteria by examining all case studies (N=184) published in five top management journals (*Academy of Management Journal*, *Administrative Science Quarterly*, *Organization*

Science, Management Science, and Strategic Management Journal) in the period 1996 – 2006.

The paper is structured as follows. We next discuss a roadmap for rigorous replication logic and subsequently outline different types of replication logic. Hypotheses regarding the links between individual replication logic types and theoretical sampling, units of analysis, controls, as well as focal variables are subsequently tested empirically.

2.3. A roadmap for rigorous replication logic in case study research

Replication logic in case studies is often compared to experimental studies: in a series of experiments, one duplicates the exact theoretical framework of the preceding experiment and predicts similar results. In this case, Yin speaks of *literal replication* (LR). Literal replication is considered fundamental in establishing the external validity of an individual case across other contexts and situations. Basically, when we find the same outcomes in a variety of cases, we can claim that these results are generalizable across the cases in the sample. Eisenhardt (1989) provided much-used practical guidance by suggesting that “a number of 4 - 10 cases usually works well” (Eisenhardt, 1989: 545). Each case thus stands on its own as an independent unit of analysis. With fewer than four cases, it may be difficult to generate a theory with much depth and its generalizability remains unconvincing, whereas with more than 10 cases the marginal benefit of adding yet another replication is outweighed by having to cope with the complexity and volume of the data (Eisenhardt, 1989: 545 - 546). For example, Bingham and Eisenhardt (2011) study “simple rules” heuristics, replicating the findings across cases from Singapore, Finland, and the U.S. This use of multiple cases in multiple countries enables a more generalizable theory to emerge rather than simply a single-country perspective (Eisenhardt, 2014: 4).

Yin uses the term *theoretical replication* (TR) when we expect different results, but for predictable reasons. Theoretical replication is fundamental in establishing internal validity. TR involves designs, where cases from *both* extremes of a theoretical prediction (e.g. good and bad outcomes, corresponding to high/low values on an independent

variable) are deliberately chosen. The internal coherence of such a theoretical framework then depends directly on the cogency of the claim that observed outcomes are actually caused by the independent variables of interest, rather than other variables. For example, Tripsas and Gavetti (2000) study only one company (Polaroid), but in three different phases (units of analysis) illustrating how it shifted from analog to digital imaging, as well as the corresponding changes in the outcome variable (organizational capabilities and adaptation).

Finally, *no replication* (NR) applies to situations where $N=1$, i.e. where only one case or unit of analysis is investigated (Yin, 2003: 47), in order to provide an in-depth insight about processes within one case without sub-units of analysis. Replication logic finds a critical test in such single case designs, i.e. situations commonly referred to as “ $N=1$ ”. These are often considered the prototypical case study design (e.g. March, Sproull, and Tamuz, 1991). For instance, Brusoni and Prencipe (2005) studied one organization, Pirelli, to answer the research question, how does new knowledge enable technological and organizational evolution? Table 2.1 provides an overview of the three types of replication logic.

Overall, Yin and Eisenhardt’s often-cited replication logic refers to a situation where a researcher is conducting a series of case studies (Yin, 2003; Eisenhardt, 1989; Glaser and Strauss, 1967). The origins of this approach go back to John Stuart Mill’s (Mill, 1875) method of difference, where researchers set out to establish the effect of one specific factor in a series of cases or units of analysis. To establish this effect plausibly, other factors must be held constant, so that the difference in outcome can be ascribed to a corresponding difference in the selected causes of interest (Blatter and Haverland, 2012; Gerring, 2007; King, Keohane, and Verba, 1994; Lijphart, 1975; Mahoney and Goertz, 2006). How will individual replication logic types address these important issues of explicitly sampling the relevant unit of analysis and addressing explanatory and control variables? Table 2.2. provides an overview of the main steps involved in rigorous replication logic, and the ensuing sections discuss the likelihood of individual replication logic types to address each of the research actions.

	Theoretical replication	Literal replication	No replication
Recent examples	Schweizer (2005) Tinsas and Gavetti (2000)	Grant (2003) Bradach (1997)	Erusoni and Principe (2006) Burgelman (2002)
Research question	What makes an outcome "Y" possible in a series of different cases (with different values on dependent variable "Y"/causal condition "X")?	What makes an outcome "Y" possible in a series of similar cases (with similar values on dependent variable/causal condition "X")?	What makes an outcome "Y" possible in a case (with a particular value on dependent variable/causal condition "X")?
Case selection	Multiple cases exhibiting the phenomenon of interest, from both ends of the theoretical spectrum. E.g.: All cases with high and low values on "X".	Multiple cases exhibiting the phenomenon of interest, from one end of the theoretical spectrum. E.g.: All cases with high value on "X".	One case exhibiting the phenomenon of interest.
Internal validity	High, causal strength tested in variety of cases which vary in theoretically-predictable ways.	Limited to homogenous sample of cases. Subject to sampling bias.	Limited to single case under investigation. Subject to sampling bias.
External validity	Medium-high, if range of variation in sample is similar to range of variation in population. Drawing conclusions about the causal effect of X on Y from the selected cases allows generalizing to a population of cases that are similar in respect to all relevant control variables.	Drawing conclusions about the causal effect of X on Y from the selected cases allows generalizing to a population of cases that are similar in respect to all control variables.	Limited, applies to selected case, only.

Table 2.1. Comparison of different replication logic designs

Step	Activities	Rationale	Recent examples
Discuss criteria for theoretical sampling.	Based on research question, illustrate how cases vary with regard to the relevant explanatory factor.	Focuses efforts on theoretically relevant sampling criteria, i.e. those closely tied to research question.	"We constructed a two-by-two cell design to explore effects of stronger/weaker scientific evidence and the degree of innovation complexity on spread pathways....We undertook theoretical rather than random sampling, choosing a pair of innovations in all four cells, giving us a total of eight cases." (Fertle, Fitzgerald, Wood and Hawkins, 2005; p. 119)
Identify main theoretical constructs and relationships.	Explanatory factors as well as expected outcomes to be clearly indicated and explicitly labeled	Provides theoretical leverage and representativeness of cases, avoiding sampling bias.	The model linked customer demands and technological innovation. The independent variable causal condition here was "customer demands". The dependent variable outcome was "technological innovation" or "disruptive architecture" (Christensen and Bower, 1996).
Problemalyze alternative explanations.	Consider to what extent the cases differ, and if any of these differences might be theoretically consequential.	If they might indeed be consequential, they need to be included in theoretical framework.	The authors chose to study identity dynamics within the same medical education training center. "By controlling for organization, we were better able to see differences in identity construction that were attributable to residency type as opposed to those caused by differences between medical centers." (Pratt et al., 2006; p. 238)
Focus on explicit unit of analysis.	Clearly label the object(s) under scrutiny, and how they fit theoretical expectations.	Enables within or between case comparisons, ensures that reader can appreciate empirical basis of emerging theory.	"The SBU was selected as the unit of analysis because of its centrality in the product innovation process." (Brown and Eisenhardt, 1997; p. 4)
Provide details of case study context.	Illustrate any contextual variables that are not directly relevant to the emerging theory, but which might enhance external validity of case.	Contextual variables allow for comparisons with cases not studied in sample.	The authors provide detailed contextual information about the three teams they observed (Coo, Ico, and Hoo). The details include physical condition of the offices, team composition, nature of work, and various kinds of software engineering tasks that each team undertook (Perlow et al., 2004).

Table 2.2. Main research actions involved in rigorous replication logic

Discuss criteria for theoretical sampling and problematize alternative explanations

Cook and Campbell (1979) coined the widely accepted term of “internal validity” to refer to covariational relationships between independent variables and outcomes. The cases or units of analysis we study should therefore exhibit strong differences with respect to the main independent variables, corresponding differences on the outcome variable, and be otherwise as similar as possible. John Stuart Mill pointed out that “we can either find an instance in nature suited to our purpose, or, by artificial arrangement of circumstances make one” (Mill, 1875: 249). Since case study research by definition precludes manipulation, the emphasis here is on “finding” the right cases (i.e. those that exhibit the characteristics necessary to make causal claims). The characteristics of the selected cases then directly determine the possibility of the case researcher making internally valid inferences about independent variables and their effects, which is why cases need to be sampled theoretically (i.e. so that their characteristics allow for a test of a causal argument, focusing on the variables of theoretical interest, while keeping others constant, or “controlled”). As such, theoretical sampling¹ constitutes a precondition for successful replication and internal validity.

Rigorous replication logic in case study research would therefore not only stipulate explicitly the degree of variance underlying the focal variables of theoretical interest (i.e. independent variables), but also those, which plausibly provide alternative explanations (and therefore need to be controlled for, i.e. the control variables, or

¹ Theoretical sampling is often contrasted with random sampling. The goal of random sampling is to control for an infinite number of alternative explanations without specifying explicitly what any of them are. It never completely controls for additional variables that may have an effect, but renders them implausible to a measurable degree. By contrast, theoretical sampling specifically selects cases that are similar with regard to these variables that are believed to *not* cause the observed effect, rendering alternative explanations based on these factors implausible (Yin, 2003). Several methodologists have stressed that random sampling is not applicable to small-N research, precisely as it might lead to selecting cases with insufficient variation in the independent variable of interest, or may vary on the control variables which the theoretical framework sets out to keep constant (e.g. Blatter and Haverland, 2012; King, Keohane, and Verba, 1994). By contrast, the central importance of theoretical sampling for small-N research is generally accepted across epistemological traditions (as evidenced, for instance, in Glaser and Strauss, 1967, as well as Gerring, 2007, devoting entire chapters to it). In particular, Glaser and Strauss suggested that “the adequate theoretical sample is easily judged on the basis of how *widely and diversely* the analyst chose his groups for *saturating categories according to the type of theory* he wished to develop” (Glaser and Strauss, 1967: 63, emphasis added).

“controls”). Given the pivotal importance of discussing theoretical sampling and the explicit consideration of control variables, we expect that case study authors who use a replication logic involving cases from both extremes of a theoretical prediction such as good and bad outcomes, or high/low values on an independent variable (i.e. TR) to explicitly discuss the criteria used for theoretical sampling, as well as the control variables. This is in contrast to authors who use a research design that does not involve replication logic (no replication, i.e. NR) or those using only cases with the same variance on dependent and independent variable (i.e. literal replication, or LR).

*Hypothesis 1: Case studies which use “theoretical replication” (TR) are more likely to discuss **theoretical sampling** than the case studies that use “literal replication” (LR) or “no replication” (NR).*

*Hypothesis 2: Case studies which use TR are more likely to discuss **controls** than case studies that use LR or NR.*

Identify main theoretical constructs and relationships

Notwithstanding the benefits of theoretical sampling, several methodologists have alerted us to its potential downside: selection bias, in particular selection bias stemming from selecting on the dependent variable (Dion, 1989; Geddes, 1990). Selection bias refers to the situation where a non-random sample of cases results in inferences that are not representative of the population. Deliberate selection of cases by the investigator poses the problem of overrepresentation of cases on one end of the distribution of a key variable, suggesting that a relationship between two variables exists for cases A–B as well as C–D (see Figure 2.1., below), when, in fact, this relationship only applies to cases A–B (see Figure 2.2., below).

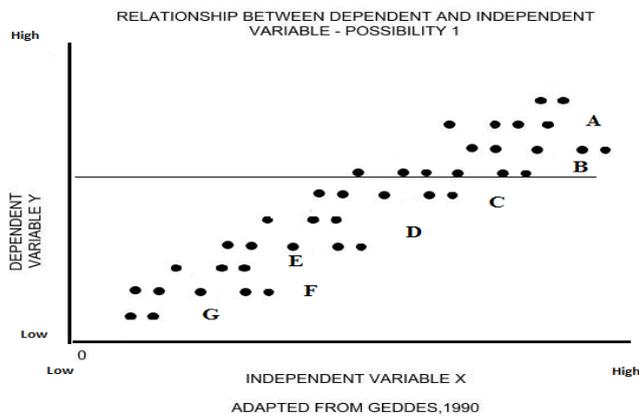


Figure 2.1. Assumed relationship between the factor X and dependent variable (Geddes, 1990: 133)

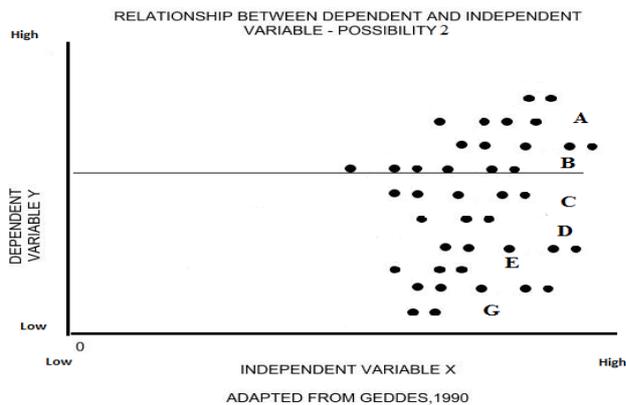


Figure 2.2. An alternative possibility for the relationship between factor X and dependent variable (Geddes, 1990: 133)

The risk of sampling bias in small-N research is particularly acute when using a criterion for theoretical sampling commonly called “extreme” cases (Eisenhardt: 1989: 537; Yin, 1994: 38–40; Glaser and Strauss, 1967: 47). “Extreme” or “polar” cases in this regard might be misunderstood as displaying extremeness on the dependent variable, only (e.g. come from the upper right-hand corner of Figure 2.1. Instead, extremeness should be interpreted with regard to situations where the independent variable and the

outcome variable co-vary (i.e. involve also cases in the lower left hand corner of Figure 2.1). As a consequence, case study researchers might effectively misinterpret calls for theoretically sampling extreme cases with the result of selection bias (as illustrated in Figure 2.2). In particular, Eisenhardt's (1989) suggestion to build on multiple case designs using replication logic to avoid results that are idiosyncratic to a single case might easily be misconstrued. As she notes,

“given the limited number of cases which can usually be studied, it makes sense to choose cases such as extreme situations and polar types in which the process of interest is ‘transparently observable’. Thus, the goal of theoretical sampling is to choose cases which are likely to replicate or extend emergent theory. [This is in contrast to] statistical sampling, where researchers randomly select the sample from the population. In this type of study, the goal of the sampling process is to obtain accurate statistical information about the *distributions of variables* within the population” (Eisenhardt, 1989: 537, emphasis added).

The risk of sampling bias for theory building is particularly acute when it comes to case study designs relying on LR (only). Unless LR within extreme cases from one end of the prediction (e.g. those where the independent variable is high and the dependent variable is high, also) is combined with TR across extreme cases at the other end of the co-variational spectrum (e.g. those where the independent variable is low and the dependent variable is low, also, for instance comparing cases A–B with cases F–G in Figure 2.1) the sample might over-represent just one kind of “extreme” cases. Put differently, without comparing cases sampled purposefully from opposite ends of the same spectrum, it is hardly plausible to establish a causal link between independent and dependent variables across a minimum range of variance, and results may not be representative of the population, curtailing external validity. The internal validity of such a theory would also be minimal, as we do not know whether there is (Figure 1) or is not (Figure 2.2) a relationship between the variables of interest, and the resulting theory is predictive only with regard to the narrow range characterizing the sample. We suggest that authors who are using TR are more likely to explicitly discuss the relationship

between dependent and independent variables than authors using NR, or authors using LR. Hence, we hypothesize that authors using theoretical replication will be more likely to discuss their focal dependent and independent variables explicitly.

*Hypothesis 3: Case studies which use TR are more likely to discuss **dependent and independent variables** than case studies that use LR or NR.*

Focus on explicit unit of analysis

How many cases or sub-units of analysis are needed for rigorous replication logic? In research designs using more than one case (i.e. in both LR and TR), the researcher can stop replicating, that is, choosing and analyzing additional cases, when she has reached theoretical saturation. Glaser and Strauss provide a straightforward definition of theoretical saturation as when the new data do not provide any additional theoretical insight (Glaser and Strauss, 2006: 61). Logically, though, the quest for saturation – and by implication the need to study yet another case – is infinite, for a theory remains valid until it is falsified (Popper, 1934). Gerring notes that the very rationale of case studies is to analyze a few cases in depth (qualitatively) and points out that beyond a certain (small) number of cases, “in order to reach meaningful conclusions about [such a] pile of data, it will be necessary to reduce informational overload, which is why God gave us statistics” (Gerring, 2007: 33). Thus, whereas it is logically impossible to stipulate with certainty the minimum number of replications needed to reach theoretical saturation, replication logically requires at least two cases or sub-units of analysis within a case.

Authors of positivistic as well as constructivist persuasions are in full agreement that replication requires an $N > 1$, and not $N = 1$. Gerring states that “the evidentiary basis upon which case studies rely is plural, not singular” (Gerring, 2007: 27). Glaser and Strauss argue that “saturation can never be attained by studying one incident in one group. What is gained by studying one group is at most the discovery of some basic categories and a few of their properties” (Glaser and Strauss, 2006: 62). A research design where there is only one case studied, and without replication among theoretically relevant units of analysis within that case is

“not logically feasible. A single case observed at a single point in time without the addition of within-case observations offers no evidence whatsoever of a causal proposition. In trying to intuit a causal relationship from this snapshot one would be engaging in a truly random operation, since an infinite number of lines might be drawn” (Gerring, 2007: 31).

How will the three types of replication logic fare with regard to making their evidentiary basis explicit? In particular, how likely is each one to explicitly mention the *unit of analysis*? Both LR and TR by definition are replicating something. And what exactly that “something” (i.e. the case(s) or sub-units of analysis within case) is, needs to be obviated in both TR and LR, otherwise the reader cannot appreciate just what is being replicated in the first place. As such, we do not expect significant differences between TR and LR, when it comes to mentioning the unit of analysis. The main contrast should be between cases studies that by design do not set out to replicate anything (NR case studies) and those that replicate “something” (LR and TR designs). Thus,

Hypothesis 4: Case studies which use TR and LR are more likely to mention the unit of analysis explicitly than case studies that use NR.

To summarize thus far, rigorous replication logic represents a tool first and foremost for enhancing the possibility of making causal inferences, i.e. internal validity. Specifically, TR enhances internal validity in terms of theoretical sampling, controlling for some variables, explicitly mentioning the focal dependent and independent variables, as well as pointing clearly to the unit of analysis. TR is based on the method of difference, which sets out to compare cases which are comparable with regard to plausible alternative explanations, but differ with regard to the independent variables of interest which are believed to cause the observed effects. By contrast, LR replicates similar (rather than different) cases and is used mainly in the context of external validity. NR, whereas by definition does not replicate results across cases and units of analysis. We suggest that the desirable characteristics of case studies should be spelled out *ex ante*, so readers can appreciate the rationale for case selection and the cogency of the

causal argument, and that TR designs are most likely to do so. We therefore predict by way of summary:

Hypothesis 5: Case studies which use TR are more likely to use any of the internal validity measures than case studies that use LR or NR.

Provide details of case study context

Once internal validity is ensured, a subsequent consideration in using replication logic is enhanced external validity. “External validity”, or “generalizability” is grounded in the intuitive belief that theories must be shown to account for phenomena not only in the setting in which they are studied, but also in other settings (e.g. Calder, Phillips and Tybout, 1982; McGrath and Brinberg, 1983). The two most commonly used understandings of generalization are within-population and cross-population generalization. According to Tsang and Williams (2012), within-population generalization refers to the degree to which findings based on a sample apply to the corresponding population. In contrast, cross-population generalization refers to the extent to which findings from one sample in one population apply to members of other populations existing in similar contexts and in similar periods of time. Cross-population generalization can further be detailed into temporal and contextual generalization, with the former referring to populations in different context, but similar time periods, and the latter with different time periods but similar contexts (Tsang and Williams, 2012: 741).

All of these types of generalization are what Yin refers to as “statistical generalization”, where “an inference is made about a population [...] on the basis of empirical data collected about a sample.” (Yin, 2003: 32–33). Empirical findings based on TR will stipulate both how cases are different as well as how they are similar (which represent the very criteria why they were sampled in the first place). Notably, by explicitly spelling out the theoretical rationale for selecting cases, we automatically also spell out how representative they are of other, not-studied cases. That is, we coincidentally also spell out their relevance *outside* the studied sample, a.k.a. their external validity. Slater and Zibblatt point out that TR achieves external validity by “explaining variation in outcomes across closely matched cases rather than an individual

outcome in a single case” (Slater and Zibblatt, 2013: 7). Consider the example of Putnam’s (1993) study of government performance in the case of Italy. Putman’s goal, so Slater and Zibblatt, was not to explain Italian government performance as a single case with no replication (i.e. an NR design), but to explain variation in government performance across closely matched cases, namely the provinces of northern Italy (which is well-governed) and southern Italy (which is much less well-governed), respectively. In our terms, Putman used a TR design. By comparing two sets of provinces in one country, Putman controls for a wide array of national-level variables which might provide alternative explanations (e.g. Catholicism, parliamentarism, fascist legacies). Furthermore, the cases chosen (northern and southern Italy) are extreme in the sense that their variation in outcomes is so vast that it approximates the full range of variation in industrialized democracies at large (northern Italy in fact is as well governed as any other country in the OECD, whereas southern Italy is among the worst managed (Slater and Zibblatt, 2013: 8, original emphasis).² Thus, explicitly spelling out the

² In principle, generalizing from case study research follows the same logic as generalizing from large-N research, which is why “it makes sense to speak of “statistical generalization” when it comes to characterizing generalizability of case studies (Blatter and Haverland, 2012: 69). Large-N research, by virtue of using random sampling, logically has a wider generalizability (since it “controls” for a theoretically infinite number of alternative explanations without explicitly specifying them). Case study research on the other hand can plausibly only be generalized to unstudied cases as long as they correspond to the studied cases in terms of the control variables, and the range of variation in terms of theoretically relevant factors. Thus, the generalizability of case studies relative to large-N research is a matter of degree, not a matter of kind. Unhelpfully, over the past five decades, authors of various epistemological persuasions (Lee and Baskerville, 2003; Tsang and Williams, 2012; Yin, 2003; Glaser and Strauss, 1967) have fervently argued that different types of generalizability apply to case studies and quantitative research. A case in point is Yin, who stresses on several occasions that “a fatal flaw in doing case studies is to conceive of statistical generalization as the method of generalizing results of a case study” (Yin, 2003: 32). In particular, Yin distinguishes statistical generalization from “analytical generalization”. Analytical generalization refers to a situation where “the investigator is striving to generalize a particular set of results to some broader theory. [...] The generalization is not automatic, however. A theory must be tested by replicating the findings in a second or third [setting] where the theory has specified the same result should occur” (Yin, 2003: 37). The correct mode of generalization to utilize in case studies, so Yin, is not statistical generalization, but analytic generalization. It must be appreciated, though, that *analytic generalization* and *replication* are used almost interchangeably: “analytic generalization, in which a previously developed theory is used as a template with which to compare the empirical results of the case study. If two or more cases are shown to support the same theory, replication can be claimed” (Yin, 2003: 32-33). At the same time, analytic “generalization” as Yin conceptualizes it is about causal relationships between variables, rather than the applicability or validity of these relationships beyond the immediate case(s) studied. Yin thus actively dismisses the usual interpretation of “generalization” as an appropriate methodological criterion for case studies and puts one in its place that is not at all about the

theoretical rationale for case selection in terms of their variation along theoretically relevant factors allows for an efficient assessment of any given study's representativeness or similarity with an unstudied case which, in turn, allows for an assessment of its external validity. We expect authors using TR to be more aware of the central importance of explicitly discussing measures taken to enhance external validity than authors using LR or NR. Consequently,

*Hypothesis 6: Case studies which use TR are more likely to discuss measures related to **external validity** than case studies that use LR or NR.*

2.4. Methodology

We selected qualitative case studies which were published in five top management journals in the period 1996–2006: *Strategic Management Journal*, *Academy of Management Journal*, *Administrative Science Quarterly*, *Organization Science* and *Management Science*. These journals were used previously in studies on methodological rigor (e.g. Gibbert et al., 2008; Gomez, Mejia and Belkin, 1992; Podsakoff, MacKenzie, Bachrach, D. G., and Podsakoff, 2005; Tahai and Meyer, 1999). To identify articles, we followed the criteria set forth by a previous study on case study rigor (Gibbert et al., 2008). “Qualitative research involves the studied use and collection of a variety of empirical materials – case study, personal experience, introspective, life story, interview, observational, historical, interactional and visual texts – that describe routine and problematic moments and meanings in individuals' lives” (Denzin and Lincoln, 1994: 3). To identify the articles which meet the relevant criteria from among these journals, we performed a key word search involving the keywords: qualitative, case study, grounded theory, triangulation, archival data, interview, observation, coding, theoretical sampling, and ethnography. Articles which used both qualitative and quantitative methods simultaneously (mixed methods articles) were excluded. A candidate list was then compiled which included the author names, year of publication, journal name, volume, issue, page numbers, etc. Our final sample consisted of 184 qualitative papers.

plausibility of causal relationships between variables beyond the cases studied but refers instead squarely to the causalities between variables in the studied cases. Plainly put, analytic generalization is about internal validity, rather than external validity.

Coding

The 184 papers were then coded by two blind coders. The coding sheet included six dimensions and research actions taken to ensure the internal and external validity of studies along with the three replication logic dimensions. The codes were dichotomous: a paper which met a particular code was marked with “1” and a paper which did not, was marked with “0”.

Internal validity involved the following dimensions which were adapted from prior studies on rigor (Cook and Campbell, 1979; Gerring, 2007; Gibbert et al., 2008; Gibbert and Ruigrok, 2010; Glaser and Strauss, 1967; Yin, 1994) of qualitative research: if authors had clearly indicated the unit of analysis, this was code 1. Code 2 was assigned to papers which discussed theoretical sampling. Articles explicitly discussing controls were assigned code 3. Case studies in which the dependent and independent variables were identified, were marked for code 4. Codes 5 to 7 dealt with the three types of replication logic.

Based on Yin’s and Eisenhardt’s recommendations on replication logic, we developed a set of three codes to ascertain which specific type of replication logic was used to design the case study research. Code 5 (literal replication) was applied when authors chose “extreme” cases from one end of a range of possible outcomes, only (typically, these were those where the outcome variable was “high”, corresponding to high levels on the input variable). Code 6 (TR) was applied to research designs where the range of variance includes not just the high/high, but also low/low cases. Both types of replication logic could involve either cross-case designs (where the replication happened on the level of the case), or embedded designs, where there was only one case, but with several units of analysis (one company with several teams, for instance). Code 7 (no replication logic) applied to scenarios where there was no replication either on the level of the case, or on embedded sub-units of analysis. Table 2.3. includes examples and research strategies for each type of replication logic.

#	Reference	RL	Research question/focus	Research design	Results
1	Schweizer (2005)	TR	How (is) a biotech company integrated into a pharmaceutical company which is seeking to gain access to the former's technology, innovative capabilities and know-how?	Five polar cases: Successful and unsuccessful biotech and pharmaceutical company deals.	The successful firms have a post-acquisition integration approach which take the complexity and multifacetedness into account.
2	Tripsas and Gavetti (2000)	TR	How does managerial cognition affect the adaptive intelligence of organizations?	One company (Polaroid), but with three different phases (units of analysis) illustrating how it shifted from analog to digital imaging, as well as the corresponding changes in the outcome variable (organizational capabilities and a daptation).	In a new environment, managerial cognition directs search processes and organizational capabilities, eventually aiding organizational adaptation.
3	Grant (2003)	LR	(How) do companies perform strategic planning in increased environmental turbulence?	Ten similar cases: Vertically integrated, diversified, large multinational (oil and gas) companies in a turbulent environment.	In turbulent corporate environments, strategic planning plays the role of a coordination and performance management mechanism.
4	Bradach (1997)	LR	How do chain organizations achieve the dual objectives of uniformity and system wide adaptation to changing markets?	Five similar cases: Large plural-form (simultaneously using company and franchise units) chain restaurant organizations.	Processes involved in plural forms balance the similarity and differences between the two structures involved, thus ensuring uniformity and system wide adaptation.
5	Brusoni and Prencipe (2006)	NR	How does new knowledge enable technological and organizational evolution?	Single case: One organization (Pirelli tires).	New knowledge or technology creates new needs and opportunities in an organization, which ultimately leads to a new open-ended evolving design.
6	Burgelman (2002)	NR	What are the implications of extraordinary success and coevolutionary lock-in for organizational adaptation?	Single case: Longitudinal study of a company (Intel) during the tenure of a CEO (Andy Grove).	Long term success can result in coevolutionary lock-in, which, in turn, could impede new business development.

Table 2.3. Replication logic and research strategies undertaken: Examples

The research actions to ensure external validity (generalizability) were assessed using two codes from Gibbert et al. (2008): A case study which involved a cross-case analysis was assigned code 8. Code 9 was assigned when details of the specific context of the case being studied such as financial data, business cycle, organizational structure etc. were presented. Each one of the articles was coded independently by two researchers. Thus we had 2 x 184 individual coding sheets (no. of raters x no. of papers) and a final consensus coding sheet. Following Gibbert et al. (2008), the whole article was read and coded, rather than just the methods section. This was necessary since some of authors either presented their methodology in the appendix (this practice was common, for instance, in *Organization Science* papers prior to 2000), or discussed methodological considerations in other sections of the paper (e.g. many authors chose to discuss generalizability issues in the limitations section). To assess inter-rater reliability of the coding process, a consensus coding approach was followed with coders discussing and agreeing upon the most correct one. During instances of differences in the initial coding, the coders went back to the concerned papers and discussed them till there was consensus. Overall pre-consensus inter-rater agreement was high (87.9%).

To predict the likelihood of an article (discussing a particular variable) belonging to any of the three types of replication logics, binary and multinomial logistic regression analyses were carried out. For the multinomial logistic analysis, the reference category was NR. The composites were scaled ordinally in both analyses.

2.5. Results

Descriptive statistics reveal some interesting tendencies. First, in the five journals studied, qualitative papers constituted 5.37% of the total number of papers. As table 2.4. shows, the percentage of relevant qualitative papers in each journal varied within the range of 0.42% to 11.08%. Second, 59.23% of the case studies in the sample either do not use replication logic (NR) or use replication logic primarily in the context of external validity (LR), without much consideration about internal validity. Table 2.5. discusses the share of empirical articles applying the three replication logic designs.

Journal	Published articles	Share of empirical case studies
Organization Science	641	11.08% (71)
Administrative Science Quarterly	716	5.45% (39)
Academy of Management Journal	754	5.44% (41)
Strategic Management Journal	778	3.47% (37)
Management Science	1425	0.42% (6)
TOTAL	4314	5.37% (184)

Table 2.4. Prevalence of empirical case studies in selected journals

	Total number of articles using replication logic	Percentage of articles using replication logic (to total number of articles)
No Replication	61	33.15
Literal Replication	48	26.09
Theoretical Replication	75	40.76
Total	184	100

Table 2.5. Share of empirical articles applying the three replication logic designs

With respect to data analysis, initially a binary logistic regression analysis was carried out. The regression models showed that coefficients were positive (with $p < .01$) for hypotheses H1, H2, H3, H5 and H6. This meant that a higher score on the corresponding scale was more likely to occur in TR (analyzed group) than in LR+NR (the reference group). The model thus postulated that the likelihood of a paper discussing theoretical sampling, controls, dependent and independent variables, and internal and external validity measures altogether of belonging to TR was higher when compared to either LR or NR group.

For H4, multinomial logistic regression analyses was carried out. The results showed that the regression coefficients for model TR vs. NR and LR vs. NR were positive (with $p < .01$), when it comes to discussing “units of analysis” (Table 2.6.). Thus the likelihood of a paper addressing “unit of analysis” to be in TR or LR group was higher than its likelihood to be in the NR group.

Hypotheses	TR vs LR			TR vs NR			LR vs NR			TR vs (LR+NR)		
	Beta	S.E.	Sig.	Beta	S.E.	Sig.	Beta	S.E.	Sig.	Beta	S.E.	Sig.
Theoretical sampling(H1)	0.952	-0.492	0.053	1.793	-0.538	0.001	0.841	-0.481	0.080	1.330	-0.449	0.003
Discussion of controls (H2)	1.088	-0.461	0.018	3.240	-0.827	0.000	2.152	-0.829	0.009	1.665	-0.435	0.000
Discussion of dependent& independent variables (H3)	1.456	-0.484	0.003	1.911	-0.642	0.003	0.456	-0.658	0.489	1.603	-0.448	0.000
Explicit mentioning of the unit of analysis (H4)	0.156	-0.456	0.733	1.325	-0.506	0.009	1.170	-0.438	0.008	0.603	-0.415	0.146
Internal validity measure (composite)(H5)	0.993	-0.210	0.002	1.639	-0.28	0.000	0.646	-0.246	0.009	1.212	-0.198	0.000
External validity measure (composite)(H6)	0.509	-0.416	0.221	2.554	-0.579	0.000	2.045	-0.519	0.000	1.127	-0.378	0.003

Table 2.6. Results for hypotheses 1–6

2.6. Discussion and conclusion

The objective was to point differentiate different types of replication logic, and how their individual research actions contribute to the internal and external validity (in that order of priority) of the emerging theory. Overall, TR represents a particularly sophisticated methodological design, as can be seen from the model comparing TR with LR and NR (Table 2.6). The primary aim of TR is to produce a theory that plausibly predicts effects between variables of interest (i.e. is internally valid). When only cases within one of the extreme ends of the spectrum are studied (as in LR), the resultant theory may be generalizable across these cases (resulting in enhanced external validity – only). More importantly, since it is not clear if the focal variables do in fact constitute the cause of the observed outcome, its internal validity would also be compromised. An even weaker theory would result from a single, illustrative and descriptive case, with no replication whatsoever (an NR design). Similarly, TR generally provides the relevant information in terms of details of the case study context as well as cross case analysis as a means to demonstrate the representativeness of the cases studies relative to the population of cases that are not studied, but which the researcher would like to generalize to (Table 2.7.).

Indicators	TR vs LR			TR vs NR			LR vs NR			TR vs (LR+NR)		
	Beta	S.E.	Sig.	Beta	S.E.	Sig.	Beta	S.E.	Sig.	Beta	S.E.	Sig.
Involvement of a cross-case analysis	0.163	-0.465	0.726	2.151	-0.631	0.001	1.988	-0.575	0.001	0.748	-0.429	0.081
Presentation of case context details	2.152	-1.016	0.034	4.240	-1.158	0.000	2.088	-0.752	0.005	2.703	-0.960	0.005

Table 2.7. Results for external validity indicators

This paper helps to clarify “misconceptions” (Tsang and Williams, 2012) in the debate surrounding external validity as seemingly the only criterion worth worrying about in case study research. First, a recent empirical analysis of published case studies revealed that authors (probably derailed by comments in seminars as well as reviewers about the limited generalizability of their findings) prioritized external validity at the expense of the more important internal validity in within and between case comparisons (Gibbert et al., 2008). Our results demonstrate how and why this tendency of (over-) emphasizing external validity is counterproductive. Several authors have argued that since case studies employ theoretical sampling (which controls for a limited, explicitly mentioned number of confounds only), they logically have a lower external validity than quantitative studies, which use random sampling (and controls for an unlimited number of confounds without ever explicating any of them, see Yin, 1994; Gerring, 2007). At the same time, a key strength of case studies is the in-depth analysis of causal factors with a degree of sensitivity to unexpected findings unmatched by other methods. In fact, this key strength of the case study allows for incorporating unexpected variables, and altering the design and hypotheses to accommodate the resulting “outliers”. Thus case studies use model misfit in a constructive manner, creating a theory with greater internal validity (Sullivan, 2011; Lindesmith, 1968; Katz, 1988). In contrast, quantitative studies typically see outliers as a nuisance and find myriad excuses to exclude them, (For e.g.: “The outliers are based on measurement error or some other anomalies”) in an effort to tie up “loose ends” and improve statistical power (Kendall and Wolf, 1949; Pearce, 2002). Thus, it seems that authors are underselling the case method by denying it its key strength (internal validity), while emphasizing one of its key limitations (external validity).

Second, the preoccupation with external validity by case study authors seems to have led to its widespread usage as a kind of “melting pot” of other validity types, in particular internal validity. We pointed to “replication logic” as a much misunderstood methodological means in this regard, as a kind of fire under the proverbial melting pot. Yin’s (1981), and later Eisenhardt’s (1989) influential calls for replication logic which employs multiple cases, so as to help theory building and enhance external validity (Eisenhardt, 1989: 537; Yin, 2003: 37) seem to have been widely misunderstood as calls for focusing on external validity, only. Multiple cases certainly help enhancing external validity by enabling comparisons within and between units of analysis that clarify whether an emergent finding is simply idiosyncratic to a single case or consistently applicable across several cases, i.e. externally valid (Eisenhardt and Graebner, 2007: 27). But unless a very specific design (TR) is used, authors cannot reap the full benefits of replication logic. By enabling comparisons across purposefully selected cases which vary systematically in theoretically relevant dimensions, replication logic has the potential to enhance internal validity, above and beyond ensuring only generalizability.

Third, our discussion of the relative superiority of TR is not intended to denigrate the inherent value of the other two designs. To begin with, LR, by virtue of involving comparisons across cases or units of analysis is still in a position to claim external validity (especially when compared with the NR design, which by definition does not involve such comparisons). Compared with TR, for instance, LR is not faring so badly when it comes to accommodating key measures of external validity (see Table 2.6 where there is no significant difference between TR and LR when it comes to discussing external validity measures). Note, in particular, that there were also two instances with no significant difference in the model comparing TR with LR, underscoring empirically the merits of LR. First, theoretically, this lack of difference between TR and LR (H4) comes as no surprise, since both TR and LR are about replicating “something” (i.e. a case or sub-unit of analysis). The only difference is that in the TR design, these units of analysis are sampled from both ends of a correlational spectrum (e.g. where independent variables are low, the outcomes should be low also, and where independent variables are high, so should be the outcomes), whereas LR replicates from within one of these two

extremes, only. Consequently, since both TR and LR authors are replicating “something”, it would appear that both types of authors also relay just what is being replicated, in their methods section. This reasoning finds support also when comparing TR with NR. Since NR authors by definition do not replicate units of analysis, they are least likely to bother about problematizing what their unit of analysis might be. Second, with regard to H1, LR authors also seem to take nearly as much care in discussing theoretical sampling of these units of analysis as do TR authors (see the model TR vs. LR in Table 2.6). Notwithstanding, we encourage authors to (where possible) theoretically sample cases across a minimum range of variance so as to move beyond the “usual suspect” (Gibbert et al., 2008: 1473), i.e. generalizability.

Finally, NR, which is also referred to as “causal process tracing” (e.g. by Blatter and Haverland, 2012; George and Bennett, 2005; George, 1979; Bennett and Elman, 2006), constitutes a design where one outcome is analyzed in as much depth as possible and with regard to one single case, only. The focus of tracing the causal process involves consideration either prior to entering the field, or during the field study, of just which independent variables are responsible for the observed outcome. Consequently, it seems plausible that NR authors actually do take care in relaying to the reader just what these variables are. These dynamics are empirically evident in Table 2.6, where NR authors were nearly as likely as LR authors to discuss dependent and independent variables, even though they clearly lagged behind TR authors (see the different models underlying H3 in Table 2.6). Even though NR designs are not unlikely to discuss focal variables (dependent and independent variables), they are nevertheless very unlikely indeed to discuss variables that confound the theoretical expectations. This empirical result fits nicely with methodological theory, in that TR (and LR) authors by virtue of comparing different cases need to demonstrate the comparability of their cases (and consequently discuss control variables).

2.7. Suggestions for further research

Typically, case studies are supposed to be more theory-building (and inductive), rather than theory testing (or deductive). Despite this, theoretical sampling by definition

presupposes the existence of some variables according to which case studies are selected. A fruitful area for further research would be to disentangle this apparent contradiction. How do case study authors motivate their methodological choices and how do they manage tradeoffs with regard to the need to be theory building and inductive, yet clearly anchored in a specific theory which motivates the selection of cases and variables? Research here could investigate cases that, while theoretically chosen, deviate from the theory used for case selection. Explicitly following up on – and including into the research design – any additional variables of interest that a deviant case may suggest constitutes key strategy for enhancing the internal validity of case studies above and beyond literal and theoretical replication. The deviant case may not disprove a theory. Rather, at times, it may help refining the theory, for instance, by providing additional variables which act as boundary conditions moderating the main effect of the key variables of interest. Eisenhardt suggests that replication logic involving deviant cases “confirms, extends, and sharpens theory. The result is often a theory with strong internal validity, wider generalizability and higher conceptual level” (Eisenhardt, 1989: 544). If deviant cases are explicitly included into the data collection and analysis (rather than ignored), then this may lead to the discovery of additional variables that may act as boundary conditions to the original design, potentially resulting in a richer, more plausible, and internally valid theory.

Chapter III*

“Oops, I’ve got an outlier in my data – what now?” A call for using model mis-fit as an opportunity for theory building in organization science

Prof. Michael Gibbert

Lakshmi Balachandran Nair

Dr. Matthias Weiss

Prof. Martin Hoegl

* Awards and recognitions

Academy of Management Annual Meeting: Research Methods Division & Sage Publications Best paper Award (2014).

Journal

Academy of Management Review (to be submitted)

Conferences

Gibbert, M., Nair, L.B., & Weiss, M. (2014, August). 74th Academy of Management Annual Meeting, Philadelphia, USA.

Nair, L.B., Gibbert, M., Weiss, M., & Hoegl, M. (2014, June). European Academy of Management Conference, Valencia, Spain.

Nair, L.B., Gibbert, M., Weiss, M., & Hoegl, M. (2014, March). Western Academy of Management Conference, California, USA.

Other related publications

Book

Valikangas, L., Gibbert, M., Nair, L.B., Paukku, M., & Peixoto, I. (2015). Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. (ISBN-13: 978-0-13-398944-1)

Book chapters

Nair, L.B. (2015). Gengo: Winning in translation through crowdsourcing. Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. Page 140-148.

Nair, L.B. (2015). Kaggle: Getting quant brains to play data games. Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. Page 158-170.

Nair, L.B. (2015). Organovo: Leaving 3D bioprints for others to follow. Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. Page 198-205.

Nair, L.B. (2015). Scoopshot: Pulling the scoop on the trendy photo crowdsourcing platform. Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. Page 69-76.

Nair, L.B. (2015). Spire: Launching crowd funding beyond earth. Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. Page 189-198.

Nair, L.B. (2015). TaskRabbit: Hop online and pull a rabbit out of your service network. Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. Page 131-140.

Nair, L.B. (2015). ZenRobotics: Riveting robots to reduce and recycle. Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. Page 181-189.

Nair, L.B., Paukku, M. (2015). BioCurious: The curious case of crowd-sourced science. Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. Page 124-131.

Nair, L.B., Valikangas, L. & Paukku, M. (2015). Robin Hood: The merry adventures of an asset management cooperative. Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. Page 217-223.

Valikangas, L., Gibbert, M., Nair, L.B., Paukku, M., & Peixoto, I. (2015). Outliers from a Theoretical and Methodological Perspective. Strategic Innovation. The definitive guide to outlier strategies. Financial Times Press. Page 17-30.

3.1. Abstract

While cleaning datasets of outliers is a common (and often sensible) practice, routinely sweeping all outliers under the proverbial carpet of model fit is problematic from a theory-building point of view. In fact, intently perusing outliers off the regression line, in particular those with large model residuals (i.e., “deviant cases”) presents us – in theory – with a valuable opportunity to improve model correspondence with empirical realities in organization science. In practice, however, we find that such opportunities are underutilized: In six major management journals over 20 years, while 318 papers explicitly reported outliers in their dataset (less than 5% of all empirical articles published), only six papers problematized their theoretical potential. To rectify this, we discuss which outliers are most propitious theory-building wise, offer a methodological frame to leverage their potential, and illustrate the concrete theoretical gains of deviant case analysis.

Key words: *Outlier; deviant case; regression; theory building; theoretical contribution*

3.2. Introduction

As organization science researchers, we all know the “oops-moment” that comes with finding outliers in our data set, either before analysis (e.g., when observations are characterized by extreme values on predictor or outcome variables), or during analysis (e.g., when we find extreme observations with or without model residuals). The present essay is a call to organization science researchers to not flinch at these incongruities, but to seize upon model *mis*-fit as an opportunity for theory building (Shapira, 2011; Locke, 2007).

Previous methodological work has shown that there are good reasons in quantitative studies (often firmly set on theory testing and with confirmatory objectives), to routinely ignore outliers (Aguinis et al., 2013; Lewin, 1992), treat them as noise (Lieberman, 2005), sweep them under the proverbial carpet of model fit, and eventually refrain from analyzing them further. These practices are usually either based on the assumption that the outliers are due to errors (measurement or sampling), or are performed in an effort to tie up “loose ends” and improve statistical power (Kendall and Wolf, 1949; Pearce, 2002). Similarly, qualitative studies typically focus on similarities between cases in a replication logic (Gibbert and Nair, 2013), rather than cases that do not fit the patterns (Blatter and Haverland, 2012; Eisenhardt, 1989; Gibbert et al., 2008).

From a theoretical (rather than methodological) perspective, this practice wants rectifying because intently perusing outliers may improve our theoretical purchase on empirical realities. Certain outliers (in particular those with large model residuals, i.e., “deviant cases”), could signal the need for developing theory to improve model correspondence with empirical phenomena (see Katz, 1988; Kuhn, 1977; Sullivan, 2011), and may even represent a cornerstone in scientific advancement (Aguinis et al., 2013; Popper, 1934). A dedicated engagement with outliers in statistical and qualitative analyses would thus be highly productive from a theory-building perspective, even if the researcher’s initial interest in a study may have focused on better understanding general (or average) patterns. Analyzing outliers represents a complement to a focus on

averages, an opportunity (not an obligation) for those authors interested in theory building in organization science.

Previous research has furthered our methodological understanding about how to best identify and handle outliers in statistical analyses (Aguinis et al., 2013; Lewin, 1992; Pearce, 2002). However, none of the previous treatises on the topic has provided concrete strategies for using outliers constructively from a theoretical perspective, and more specifically for building organization theory. Although there have been influential calls in the organizational sciences to understand what is special about an outlier (Freedman, 2008; King et al., 1994), there is currently a dearth of understanding on just *how* to “make doubt generative” (Locke et al., 2008). This is in stark contrast to other disciplines, where outlier analysis constitutes a widely appreciated and utilized theory-building device in areas as diverse as biology (Hagstrum, 2013), comparative politics (Emigh, 1997; Gerring, 2007b; Lieberman, 2005), health care (Mays and Pope, 2000), law (Gordon, 1947), and criminology (Sullivan, 2011).

Our present objectives are to (1) review the extent to which outliers were (under)-utilized for theoretical locomotion over the last 20 years in six major management journals, (2) discuss three concrete strategies for theory building in organizational research based on deviant cases and (3) empirically illustrate the varying degrees of theoretical progress these strategies may yield (via actual applications in published articles and an appraisal of their respective theoretical contributions). Foreshadowing results, the present analysis of all empirical articles (quantitative, qualitative, as well as mixed methods) published in a 20-year period (1993 to 2012) in *Administrative Science Quarterly*, *Academy of Management Journal*, *Strategic Management Journal*, *Organization Science*, *Journal of Management* and *Management Science*, revealed: although 318 articles (less than 5% of all empirical papers published) reported the existence of outliers in the dataset, only two articles made effective use of them by constructively building new theory, and another four articles used them, albeit slightly less thoroughly (e.g., by mentioning their existence in the limitations section and providing some speculation about their theoretical implications).

The remainder of this essay is structured as follows. Next, we discuss distinctive types of outliers, with varying degrees of theoretical interest. Subsequently, we provide three concrete strategies for utilizing those outliers of greatest theoretical interest (deviant cases), discuss their relative methodological sophistication, and illustrate their theoretical prowess empirically (with reference to published articles).

3.3. Outliers as an attractive nuisance?

Notwithstanding their potentially productive properties, not all outliers do indeed warrant systematic exploratory follow-up analysis (Aguinis et al., 2013; Cortina, 2002). Outlying observations are typically due to three reasons, with only the third reason warranting further perusal. First, an outlier can reflect measurement error. In quantitative field research, for example, researchers investigate responses to an online questionnaire and may find response bias of some kind (e.g., all questions were given the highest rating), which disqualifies the outlier from inclusion in further analysis. Second, an outlier can reflect a data collection error. An example would be a field study on innovation projects, where one case ended up in the sample which actually constitutes much more of a customer service project pursuing routine tasks (and not, as expected, an innovation project). Here, the researcher measures a case other than the one he or she intended to measure, as the study actually set out to test theory domain-specific to innovation projects (rather than customer service projects). On this basis, the case is removed from the sample, given the domain-specific objectives of the study.

There is also a third case, where outlying observations are neither due to measurement issues nor due to data collection issues. Aguinis et al. (2013: 275) termed this kind of outliers “interesting outliers” and defined them as “accurate (i.e., nonerror) data points that lie at a distance from other data points and may contain valuable or unexpected knowledge” (Aguinis et al., 2013: 275). As with non-interesting outliers, there may be good reasons to delete even interesting outliers, without foregoing their theoretical potential, depending on the *type* of outlier encountered. In particular, there are three types of interesting outliers: construct outliers, model fit outliers, and prediction outliers (Aguinis et al., 2013). They differ with regard to whether they are detected prior

or following data analysis, and whether or not they are characterized by extreme values on the studied variables and/or large model residuals (Aguinis et al., 2013). Most importantly, they differ fundamentally in terms of their theory-building potential for organization science, as summarized in Table 3.1. below and discussed in the next three subsections.

	Single-Construct Outliers	Model-Fit Outliers	Prediction Outliers (deviant cases)
<i>When identified?</i>	Before focal data analysis	Focal data analysis	Focal data analysis
<i>Identification strategy</i>	Based on descriptive statistics, distance metrics, etc., from data collected	Multiple construct techniques	By Calculating DFFITS, Cook's <i>D</i> , or DFBETAS values
<i>Location of extreme value</i>	Single variable (predictor or outcome)	Predictor variable	Relationship between a predictor and outcome
<i>Location on the regression line/normal distribution curve</i>	Extreme ends of normal distribution curve	On the regression line, but assumes higher or lower values on the independent variables when compared to the main body of observations	Off the regression line
<i>Model residuals present?</i>	Unknown prior to focal analysis	Residuals are small	Residuals are large
<i>Effects of including outlier for model fit</i>	Unknown prior to focal analysis (could turn out to be model fit or prediction outlier)	Positive: Improves model fit (e.g., R-squared)	Negative: Reduces model fit and parameter estimates.
<i>Costs of excluding outlier</i>	Unknown prior to focal analysis	Reduced explanatory power and model fit	Theoretical bereavement
<i>Theory-testing potential</i>	Unknown prior to focal analysis	High, although ambivalent (outlier improves model fit, but may over-state it)	Deviant case decreases model fit, but residuals may point to additional explanatory variables.
<i>Theory-building potential</i>	Unknown prior to focal analysis (could turn out to be model fit or prediction outlier)	Low	High

Table 3.1. Different types of outliers and their theoretical heft

Theory-Building Potential Not (-yet) Known: Construct Outliers

First, *construct outliers* are simply extreme values on a given variable whether dependent or independent variable, whether in one or in a set of variables, and are based on indicators such as descriptive statistics or distance metrics (Aguinis et al., 2013). Irrespective of the metric used to identify them, they are singled out *before* the focal data analysis on the relationship between independent and dependent variables is run. For instance, in a study linking intelligence with job performance, construct outliers would be individuals with extremely high/low IQs. There might be good reasons for ignoring or deleting extreme values on the raw data level – without the risk of biasing the results of the analysis (Aguinis et al., 2013). Specifically, the research question at hand (e.g., is the researcher interested in average or superior performance?) determines which data are to be used and which data can instead be discarded without further penalty. Construct outliers can then safely be deleted from the dataset, on the basis that they are produced by the presence of some theoretically non-relevant anomalies (Boone and Van Witteloostuijn, 2005; Haynes and Hillman, 2010), and to achieve ‘cleaner’ models (Aguinis et al., 2013; Clemen and Ulu, 2008; O’Boyle and Aguinis, 2012; Stevens, 1984). For example, techniques such as ‘winsorizing’ (e.g., replacing the top and bottom 1 or 5% of the data by the highest value that is not removed) are used to ‘trim’ raw data in means analyses such as OLS regressions, since coefficient estimates will be biased by skewed underlying distributions (Cohen et al., 2003). The potential downside of deleting construct outliers prior to data analysis resides in the infamous “unknown unknowns”: construct outliers may also be characterized by large model residuals and thus point to unexplained variance and theory building potential (more on this below). But an ultimate assessment whether or not these construct outliers are in fact characterized by large model residuals is typically not possible prior to focal data analysis: Their potential for theory building in organization science is not (-yet) known.

In some instances, special research designs and statistical approaches are built around this type of outliers (usually in the dependent variable). In such studies the primary goal is to hone in on – and possibly predict – extreme outcomes by applying econometric methods that are specifically designed to estimate rare or extreme cases

(Singh and Fleming, 2010; Wiggins and Ruefli, 2002; 2005). For instance, quantile regression is a type of regression analysis that aims at estimating the conditional median or other quantiles of the response variable (Koenker and Hallock, 2001). One of the main advantages of the quantile regression approach, relative to OLS regression (whose estimates approximate the conditional mean of the response variable given certain values of the predictor variables), is that its estimates are more robust against “outliers” in the response measurements (Koenker and Hallock, 2001). Quantile regression, or similar approaches specifically geared to study extreme cases, are used to estimate how the coefficient of the predictor variable(s) varies with the specific quantile of the response variable being considered either at one end (Wiggins and Ruefli, 2002; 2005) or at both ends of the spectrum of possible outcomes (Singh and Fleming, 2010; St. John and Harrison, 1999). In two papers by Wiggins and Ruefli (2002; 2005), for example, the authors benefited from a large initial dataset (> 6,500 firms in 45 industries over 25 years) from which a considerable number of construct outliers on one extreme end of the distribution (firms with persistently superior performance) could be extracted and examined in greater depth. As another example, St. John and Harrison (1999) juxtaposed two larger groups of construct outliers on both ends of the outcome spectrum, and examined these outlying low and high performing firms using qualitative methods.

Weak Theory-Building Potential (but strong theory-testing potential): Model-Fit Outliers

Second, *model fit outliers*, i.e., those observations “whose presence influences the fit of the model” (Aguinis et al., 2013: 275) can be identified *after* the focal data analysis. “Pure” model fit outliers are observations with small model residuals but extreme values on the predictor variable. These are thus “on” the regression line, but are outlying in the sense that they assume relatively (i.e., relative to the main body of observations) high/low values. For instance, in the hypothetical study linking intelligence to job performance, these would be observations where (as theoretically expected) particularly intelligent individuals also enjoy particularly high levels of job performance.

What are the costs and benefits of ignoring them? Since these outliers actually improve the fit of the model (e.g., R-squared), potentially even changing the statistical significance of a model fit index from insignificant to significant, researchers have incentives to include, rather than exclude them (Aguinis et al., 2013). As with construct outliers, it ultimately depends on the research question and study design whether or not they should be included. Particularly in qualitative research, where the interest is on comparing extreme cases in a replication mode (e.g., Eisenhardt, 1989), analysis may even focus exclusively on model fit outliers (i.e., those with extremely high/low values on predictor variables and, correspondingly, on outcome variables). Since qualitative research by definition focuses on a few (rather than myriad) theoretically (rather than randomly) sampled cases, there might be good theoretical reasons for focusing only on those cases where the focal variables take on extreme values. Eisenhardt (1989) explicitly promotes this practice, by arguing that in these instances the causal dynamics are transparently observable. Put differently, perusal of pure model fit outliers might reveal more crisply the causal processes underway in a larger body of observations which simply cannot all be studied qualitatively due to time and resource constraints. For instance, in the qualitative case analysis by Zott and Huy (2007), the authors investigated the link between impression management and resource acquisition by start-up companies. An initial set of 26 cases was identified, but analysis then focused only on seven extreme cases.

Thus, there are good reasons for including model fit outliers in the analysis. Despite this, if their inclusion grossly overstates the theoretical significance of the model predictions, it would be opportune to relay the results with and without the outliers. This ensures transparency of the research procedures, and “places the burden for the most accurate conclusions on the reader” (Aguinis et al., 2013: 291), consistent with a “customer centric” view of presenting research results in organization science (Aguinis et al. 2010). In sum, pure model fit outliers are likely to offer only limited theory building potential precisely because they align with the other cases illustrating the relationships between focal variables. Thus, analysis of these cases is unlikely to provide

insights that go *beyond* what a theory based on the less extreme (average) cases would generally predict.

Strong Theory-Building Potential: Prediction Outliers a.k.a. Deviant Cases

Third, *prediction outliers*, i.e., outliers whose presence affects a model's parameter estimates (Aguinis et al., 2013:275) can be identified following data analysis and are characterized by having large model residuals (only) or having large model residuals in addition to having extreme values on the focal variables (similar to pure model fit outliers). Unlike pure model fit outliers, prediction outliers are therefore characterized by model *mis*-fit, and are "off" the regression line; they are typically referred to as "deviant cases", precisely because they usually deviate from theoretical expectations. Their inclusion reduces model fit, and affects parameter estimates (e.g., slope and/or intercept coefficients). Henceforth, we use the terms "prediction outlier" and "deviant case" interchangeably.

In terms of the hypothetical study linking intelligence to job performance, there might be cases which, despite having a high IQ nevertheless score low on job performance. Conversely, there might be low-IQ cases with high job performance, and both cases defy the logic of original predictions as well as the distribution of the main body of observations. What are the theory-building implications of keeping or deleting the deviant cases?

They are often deleted from the dataset, though for different reasons. In an attempt to produce robust regressions (Stevens, 1984), data sets with deviant cases are "cleaned up": cases outside the confidence interval are deemed not to be of theoretical interest and often treated as unexplained "noise" (Lieberman, 2005: 444). This practice leads to cleaner models (without any deviant cases beyond a certain statistical threshold of model residuals), regularly increasing the explanatory power of a given model (as indicated, for instance, in terms of R-squared), and attenuating the "undue" influence of outlying cases on the statistical models (e.g., Aguinis et al., 2013; Cohen et al., 2003; Judge et al., 1988). Second, researchers may conduct their hypotheses tests with the deviant cases included. In contrast to their outright deletion, this practice, however, tends

to result in underestimated explanatory power (e.g., R-squared), given the “untidy” dataset. Usually the choice of this option is accompanied either by performing robustness checks to assure readers and reviewers of the negligible impact of the retained outliers or by using certain data trimming methods such as winsorizing.

Notwithstanding the previously discussed reasons for excluding such deviant cases from theory testing, altogether ignoring them might be highly problematic from a theory-building perspective (Kendall and Wolf, 1949; Pearce, 2002). In particular, analysis of these cases may shed new light on systematic inconsistencies and theory-data conflicts emerging whilst trying to reconcile theoretical predictions with real-world observations (Lieberman, 1992). Aguinis et al. (2013) suggested model respecification by adding additional terms to the regression specification (e.g., squared terms or moderating effects). Model-data conflicts may also point to variables that were omitted from analysis, but which would actually improve the correspondence between theory and data. Deviant cases do not necessarily invalidate theories outright, but they may reveal boundary conditions, contingencies, and conditional effects (George and Bennett, 2005; Gerring, 2007). For instance, when it comes to linking intelligence with job performance, we might decide to study those individuals who, despite having a low IQ, are high in job performance. In-depth investigation of these individuals might reveal a so-far overlooked (and therefore unmeasured) variable: emotional intelligence. High emotional intelligence might compensate for negative effects of low IQ (Côté and Miners, 2006). In criminology, Giordano (1989) used the same logic of deviant case analysis to identify youths who, based on the theorizing at the time, should have been delinquent but were not. Perusal of the outlying observations pointed to a so-far overlooked variable: the strength of ties with parents. This insight contributed to criminology by suggesting that even if all other factors point to delinquency, strong ties with parents might attenuate these factors, and ultimately curb delinquency. If similar deviant cases recur (as in fact they did in Giordano’s study on youth delinquency), a researcher may use that finding to establish (and possibly push) the boundary for the utility of the theory of interest (see also Sullivan, 2011: 907-908). If the terms or variables added post-hoc significantly improve model fit or predictive power, then

including outliers can help the researcher engage in inductive theory building (Eisenhardt, 1989; Locke, 2007), a type of research heavily underutilized in organizational theory (Aguinis and Joo, 2014).

Because deviant cases are typically small in number, a qualitative methodology would appear to be most appropriate for revealing hitherto neglected, but theoretically consequential variables, thereby inductively refining the predictive power of a theory (Eisenhardt, 1989; Locke, 2007), as well as expanding the population of theoretically well-understood observations. It should be noted, however, (sample size and number of deviant cases allowing) that a quantitative analysis using statistical methods may be perfectly feasible, too. Whether qualitative or quantitative techniques are employed, Freedman (2008) suggested that analyzing deviant cases is akin to “detective work” in an effort to understand whether they are in fact unremarkable or whether they are theoretically useful. The next section discusses the degree of methodological sophistication of three research strategies designed to leverage the theory-building potential of deviant cases.

3.4. Deviant case analysis: How to use prediction outliers for theory building

From a theory-building perspective, the most promising outliers are prediction outliers, as only they actually represent *deviant* cases that diverge from general trends in the relationships between variables (and do so not because of error). To what extent their theory-building prowess can be harnessed depends on the type of method for analyzing them. In this section, we discuss three variants of the deviant case method (DCM), which provide increasing levels of theory-building sophistication.

Deviant Cases in Their Own Right: Single DCM

The most basic DCM research strategy examines one, single deviant case (we call this the Single DCM). The research design and analytic method here is qualitative, in-depth analysis in an attempt to uncover additional variables and their relationships, which might alter the theoretical model, providing a better fit between theory and data.

Uncovering the variables causing the observed deviation then holds the clue for refining an existing theory or establishing its boundary conditions. Thus, in the Single DCM the

investigators ask what factors lead to a concrete outcome or which preconditions are necessary and sufficient in order to make a specific kind of outcome possible. They search for causal conditions that are individually necessary and, in combination with other causal conditions, sufficient for the outcome. Blatter and Haverland (2012) suggested that this kind of research is “Y-oriented”, that is, it works backwards from the outcome to find explanatory factors leading to that (deviant) outcome. It would help in pointing out the boundary conditions of a particular theory that, prior to the discovery of the deviant case, may have been both unchallenged and even prominent (Gerring, 2007). Tracing the process leading from a causal factor to an outcome therefore also constitutes a method for enhancing the study’s causal claims, since the Single DCM allows investigators to ask what factors lead to a concrete outcome or which contingencies are necessary and sufficient in order to make a specific kind of outcome possible (Gerring, 2007).

The weakness of a single DCM relates to the lack of generalizable conclusions. Its validity is “internal” to the single case studied, it may be idiosyncratic and not replicable across similar cases. Glaser and Strauss point out that “saturation can never be attained by studying one incident in one group. What is gained by studying one group is at most the discovery of some basic categories and a few of their properties” (Glaser and Strauss, 1967: 62). If in additional studies similar deviant cases surface in larger numbers, the argument of simple randomness vanishes as commonalities between these cases may provide clues as to likely causes for this group of deviant cases and more convincing theoretical grounding (Eisenhardt, 1989).

Comparing Deviant Cases with Similar Outcomes: Multiple DCM

A second DCM research strategy is the comparative analysis not just of one, but several deviant cases which have similar outcomes, that is, to replicate the results from studying one deviant case in a second or further deviant case (we call this method Multiple DCM). The replication logic underlying this second DCM is analogous to the replication logic used in experimental studies: in a series of experiments, some might duplicate the exact theoretical framework of the preceding study and we would predict similar results.

In this case, Yin (1994) speaks of literal replication. As such, literal replication involves a research design where several deviant cases were deliberately chosen because there is little or no variation in outcomes.

As in the first research strategy (Single DCM), Multiple DCM can be done in the context of a study where deviant cases are deliberately chosen and compared using a qualitative or, the number of deviant cases permitting, quantitative methodology. The Multiple DCM may also be part of a qualitative research design, where individual cases cohere theoretically, while others do not follow the expected pattern and are systematically compared both within each case (i.e., by comparing different units of analysis within the same case) as well as across cases to establish generality of the research findings.

The previously-discussed method, Single DCM, provides a theory that, while internally valid to the case at hand, may not be externally valid beyond the idiosyncrasies of the studied case. The potential for generating a theory that is externally valid beyond the one deviant case studied is therefore greater in Multiple DCM than in Single DCM. Multiple DCM illustrates consequential causal factors that are applicable to multiple deviant cases (Seawright and Gerring, 2008), leading to the generation of a new theory that could explain an entire set of deviant cases. Eisenhardt and Yin launched influential calls for case study designs employing multiple cases (Eisenhardt, 1989; Yin, 1994), precisely because of their potential to enhance generalizability. As Yin (1994) asserts, multiple cases are like a series of discrete experiments that serve as replications, contrasts, and extensions to the emerging theory. However, unlike in experiments, we cannot “artificially arrange circumstances to suit our purpose” (Mill, 1875: 249) as survey research and case study research by definition preclude manipulation. The solution here is finding (rather than creating) suitable cases. This can be done by applying a theoretical sampling technique. The researchers would collect, code, and analyze their data and decide on what data to collect next and where to find them, in order to develop theory as it emerges (Glaser and Strauss, 1967). Cases are selected because they are particularly suitable for illuminating and extending relationships among

constructs (Eisenhardt and Graebner, 2007). Ultimately, replication of multiple cases in this DCM enables comparisons within and between cases as well as units of analysis within cases that clarify whether an emergent finding is simply idiosyncratic to a single case or consistently replicated across several cases, enhancing external validity (Eisenhardt and Graebner, 2007).

Comparing Deviant Cases with Different Outcomes: Full Range DCM

In contrast to Multiple DCM, the final research design involves deliberately choosing cases from both extremes of the emerging, modified theoretical prediction (e.g., good and bad outcomes, or high/low values on an independent variable, henceforth Full Range DCM). If deviant cases with the right characteristics can be found, Full Range DCM constitutes a particularly sophisticated design from a theory-building perspective. Its sophistication can best be appreciated by comparing it with the two previous methods, Single DCM and Multiple DCM. When a theoretical sampling procedure yields only cases within one of the extreme ends of the (deviant) spectrum (as in the Multiple DCM), the resulting theory is not only necessarily narrower in scope, but more importantly, it is not clear if the focal characteristic (which the selected cases share) does in fact represent the cause of the observed outcome (Dion, 1998; Geddes, 1990).

Given the small number of cases usually available for deviant case analysis, clearly any additional deviant case that can be found may potentially be valuable theoretically (even if it provides a straight replication, “only”). If cases from both ends of a co-variational spectrum of possible outcomes (e.g., “high” as well as “low” values on both dependent and independent variables) can be found and included in a DCM design, then researchers can more effectively build a theory with wide application and strong internal consistency. This is in stark contrast to the Multiple DCM design, where outcomes from only one, “extreme” end of the spectrum of possible outcomes are chosen, precisely because it aims at literal replication with similar cases (Yin, 1994). An even weaker theory would result from a single deviant case, with no replication whatsoever (the single deviant case is well-equipped to falsify a theory, but less well equipped for generating theory that is valid beyond the idiosyncrasies of the studied

“extreme”). Put differently, Full Range DCM constitutes the most sophisticated approach for theory building in organizational research – if cases with the necessary characteristics can be found and included in the analysis, that is.

3.5. Methodology

To identify if and how authors used outliers for theory building in organizational research, we analyzed six top-tier management journals which are commonly used in studies on methodological rigor (e.g., Gibbert et al., 2008; Gomez-Mejia and Balkin, 1992; Podsakoff et al., 2005; Tahai and Meyer, 1999) over a twenty-year period (1993–2012): *Strategic Management Journal* (SMJ), *Academy of Management Journal* (AMJ), *Administrative Science Quarterly* (ASQ), *Journal of Management* (JoM), *Organization Science* (OS) and *Management Science* (MS).

To identify papers analyzing outliers in our target body of papers using quantitative, qualitative, as well as mixed empirical methods, we proceeded in three steps. First, we looked for papers that reported the identification of outliers in their data. To do so, we searched through the full text of all papers published in the target journals in the period 1993–2012 using the same terms as Aguinis et al. (2013), i.e., “outlier”, “outlying”, “influential case”, “influential data”, and “influential observation” in singular and plural forms), as well as adding the term “deviant case” in singular and plural form. Papers based on qualitative or mixed methods (i.e., those that combine quantitative and qualitative analysis) required a different sampling technique. The reason is that papers containing at least some element of qualitative analysis suffer from less well-established reporting conventions (e.g., Gibbert and Ruigrok, 2010), with the result that qualitative studies tend to use different terminology for outlying cases, which might not be captured with a standard keyword search. To give but one example, consider the case of Zott and Huy (2007, discussed in more detail later), where the authors do not employ any of the search terms utilized to identify the quantitative papers. Instead, they circumscribe the identification of deviant cases as follows:

“The general pattern suggested that high impression management activity may be positively associated with high resource acquisition, but we were not fully

satisfied, because a few ventures with high impression management actions experienced moderate to low success, so the impression management findings were imprecise. This dissatisfaction led us to reanalyze our data in a second step and focus on symbolic action as a subset of impression management” (Zott and Huy, 2007: 77).

To obtain all studies containing *some* qualitative analysis published in the six target journals, we conducted a key word search using: qualitative, case study, grounded theory, triangulation, archival data, interview, observation, coding, theoretical sampling, ethnography. We subjected these to a manual search for outliers. As an aside, please note that since papers such as Zott and Huy (2007) did not use the typical terminology used by quantitative papers to describe outlying cases, the successful identification of this paper by the additional search through all qualitative and mixed methods papers provides confirmation of the rigor, reliability, and effectiveness of the sampling technique used here. Third, we compared the results of our search for papers reporting outliers with those of Aguinis et al. (2013) to ensure we did not miss any source they identified.

This process resulted in the identification of 318 papers, which explicitly reported the detection of outliers (less than 5% of all empirical articles published in these six journals during these 20 years). There were 90 papers in SMJ, 62 in AMJ, 16 in ASQ, 28 in JoM, 34 in OS, and 88 in MS. Among the 318 papers identified, we looked for examples actually making use of the theory building potential of influential outliers, based on the definitions above. Figure 3.1. gives a detailed view of the whole process.

Specifically, we first had to identify papers that only identified *construct outliers*, as their theory building potential is not known prior to analysis. We found that 74% of the papers identified involve construct outliers, i.e., 235 papers out of 318. Of these 235 papers, 37%, i.e., 86 papers reported robustness checks or further analysis to ascertain whether the construct outliers were actually influential. Of these 86 papers, only 19 (22 %) reported that the outlier indeed showed large residuals, consistent with our definition of prediction outliers (deviant cases). In addition to these 19 papers, we also found that

83 other papers had reported deviant cases but without prior reporting of construct outliers. We thus ended up with 102 papers reporting deviant cases out of the 318 papers reporting the identification of outliers (32%).

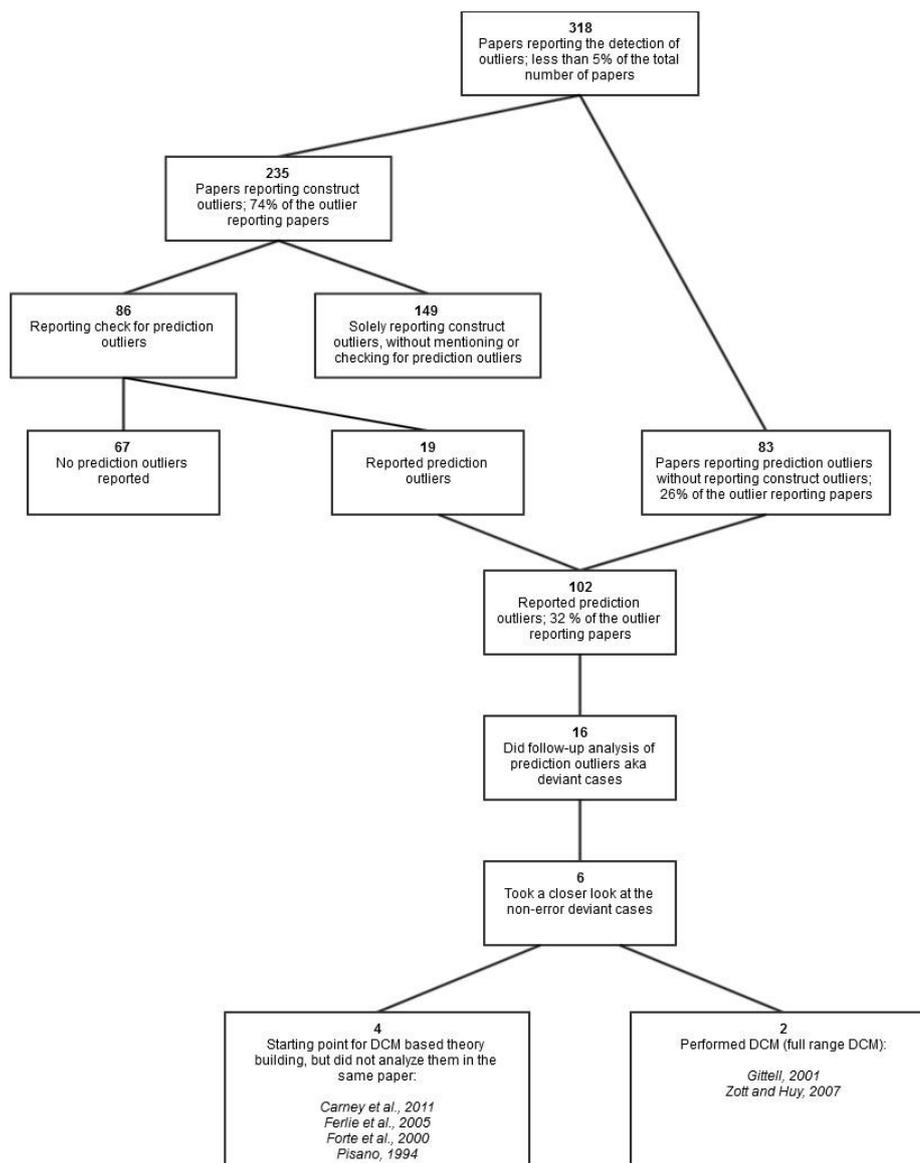


Figure 3.1. Search process for prediction outliers (“deviant cases”)

As a next step, we ascertained which of these papers conducted some kind of follow-up analysis. This was the case in only 16 of the 102 papers. A closer inspection of these 16 papers led to the conclusion that in 10 papers, the reported deviant cases actually represented instances of measurement or data collection error initially not recognized by the authors (e.g., Schoenecker and Cooper, 1998; Worren et al., 2002), and were therefore deleted. For example, Worren et al. (2002: 1132) deleted one outlying case, reporting that in a post-hoc telephone interview they “called up the respondent submitting these data, who said that he had misunderstood some of the questions in that he had considered, e.g., component sharing between product lines rather than within product lines when filling out the questionnaire.” In these papers, the authors typically explained the nature of an outlying case and the underlying reasons in one or two sentences without attempting to use the deviant case for theory building. Following-up a deviant case that came up in the examination of teams’ communication patterns, Cramton (2001: 362), for example, notes that:

“Team 6 is an outlier in that its e-mail volume was the highest of all the teams and its internal relationships appear to have been the most positive. Members of this team realized early on that uneven distribution of information among team members could be a problem and invested considerable effort in avoiding this situation. However, the team's work product was not graded as highly as that of some of the other teams. In analysis papers, members of both Teams 6 and 37 suggested that desire for harmony in the team had interfered with scrutiny of business ideas.”

In the end, this exercise produced six papers, which actually perused deviant cases. Four papers could have provided a starting point for DCM-based theory building, although this opportunity was not taken up (Carney et al., 2011; Ferlie et al., 2005; Forte et al., 2000; Pisano, 1994). Two papers actually did seize the opportunity to develop new theory by using DCM (Gittell, 2001; Zott and Huy, 2007; discussed subsequently).

3.6. Discussion of results: Theory-building potential realized vs. foregone

Theory-building potential fully realized

Only two papers, one qualitative study (Zott and Huy, 2007), and one mixed methods paper (Gittell, 2001) used the potential of deviant cases for theory building in organizational research. Table 3.2. provides a brief summary of these papers' thematic foci, the DCM procedure used, and the contribution the DCM based theory development provided to the literature.

	Zott and Huy (2007)	Gittell (2001)
Research objective	Shedding more light on how entrepreneurs acquire resources, in particular the theoretical rationale behind the actions entrepreneurs engage in when acquiring resources.	Clarifying the role of supervisory span of control by investigating whether broad or narrow spans of control are more efficient in support of group processes.
Theoretical backdrop	Literature on impression management (Aldrich and Fiol, 1994; Lounsbury and Glynn, 2001; Zimmermann and Zeitz, 2002), which suggests that high impression management is associated with high resource acquisition potential in entrepreneurship.	Postbureaucratic theory (Hackman and Oldham, 1980; Walton and Hackman, 1986), suggests broad spans of control, whereas several other theories in organization science call for narrow spans of control (Ancona, 1990; Eisenhardt and Tabrizi, 1995; Ford, 1981; Likert, 1961; Tannenbaum 1968; Udy, 1959).
How was deviant case(s) identified?	Qualitative analysis of seven extreme cases (from an overall sample of 26 cases) left generally confirmed that high impression management was associated with high resource acquisition, however, there were cases where high impression management was unexpectedly associated with moderate to low success.	A regression of supervisory span on group performance suggested that broad supervisory spans are significantly associated with lower levels of group performance (i.e., relational coordination). Narrow supervisory spans are associated with higher levels of performance. Two deviant cases found: narrow spans with (unexpectedly) low levels of performance, and broad spans with high levels of performance.
Which DCM used?	Full-range DCM; re-analysis including new data collection suggested that apart from impression management, symbolic action was a more reliable predictor of resource acquisition (confirmed not only in the seven extreme cases, but in all 26 cases).	Full-range DCM; small spans can also serve as the opportunity for a more directive style of supervision, which in the deviant case appears to have boosted performance by means other than high levels of relational coordination. And larger spans (in the case of the second deviant case) were compatible with high levels of relational coordination.
Analytical	Qualitative, deliberately inductive.	Qualitative, deductive at the outset, then mixed methods: main hypothesis was

<i>approach</i>		examined using quantitative (regression) techniques, qualitative data from same study used to (re-) interpret results.
<i>Theoretical model before DCM</i>	Impression management boosts resource acquisition potential.	Inconsistent results between postbureaucratic theory and several other theories regarding role of supervisory span of control.
<i>Theoretical model after DCM</i>	Symbolic actions (a subset of impression management) are a more reliable predictor of resource acquisition potential.	Deviant cases suggest some countervailing contrary to the postbureaucratic stream in organizational theory considerations.
<i>Theoretical gains</i>	The overall theory of impression management may have confounded symbolic action so far. Disentangling the new factor allowed the authors to not only provide a richer understanding of impression management, but nuance its impact further by pointing to several moderators.	DCM allowed the author to not only reconcile the conflicting predictions in organization theory, but to propose a nuanced, new theory on relational coordination, namely that supervisors with smaller spans may achieve higher levels of relational coordination through working with, and providing intensive coaching and feedback to their peers.

Table 3.2. Examples where deviant cases’ theory-building potential is fully realized

From impression management to symbolic action in resource acquisition: Full-range DCM in a qualitative study - Zott and Huy (2007).

Acquiring resources represents a key challenge for nascent organizations. Especially young, inexperienced and unproven organizations are particularly challenged here since resource holders are often reluctant to commit their resources to new ventures. The starting point for the qualitative study by Zott and Huy (2007) was that it is far from clear just what specific actions entrepreneurs engage in when acquiring resources, and what the theoretical rationale underlying the relatively similar activities might be:

“In many instances, researchers have tended to look at these actions as a kind of checklist but have not really explored how and why performing them would have a differential impact on resource acquisition” (Zott and Huy, 2007: 2).

Zott and Huy (2007) report on the results of a two-year field study of how entrepreneurs in British ventures acquire resources. The study was deliberately inductive

and although there was no predetermined theory that guided initial questions, the authors relied on the impression management literature (Aldrich and Fiol, 1994; Lounsbury and Glynn, 2001; Zimmerman and Zeitz, 2002) to theoretically sample a subset of 26 cases with varying degrees of impression management. Analysis focused on seven extreme cases, which were characterized by particularly high (low) levels of impression management. Impression management theory was then used to code the interview and archival data from both entrepreneurs and resource holders in an attempt to uncover how entrepreneurs acquire resources. This theory predicted that high impression management activity positively influences resource acquisition potential. In the initial analysis, the authors

“first tried a theory-elaboration approach (Lee et al., 1999), in which we analyzed a sample of ventures that displayed high and low levels of impression management actions, according to the taxonomy proposed by Gardner and Avolio (1998)” (Zott and Huy, 2007: 14).

However, the ensuing analysis revealed several cases where high impression management was *unexpectedly* associated with moderate to low success. This, in turn, led to a reanalysis of the data, with a focus specifically on symbolic action (rather than impression management, more generally, Zott and Huy, 2007: 77). To recall the authors’ crucial moment of discovery at this stage:

“The general pattern suggested that high impression management activity may be positively associated with high resource acquisition, but we were not fully satisfied, because a few ventures with high impression management actions experienced moderate to low success, so the impression management findings were imprecise. This dissatisfaction led us to reanalyze our data in a second step and focus on symbolic action as a subset of impression management” (Zott and Huy, 2007: 77).

The authors then clearly relayed how new theoretical categories were used in the re-analysis of the data, now focusing specifically on symbolic action. Zott and Huy thus

used Full Range DCM, which identified deviant cases based on imprecise model fit (due to confounding impression management overall with symbolic action management). Subsequently, the omitted variable improved model fit and a new theory of resource acquisition was established where high (vs. low) symbolic actions were consistently associated with high (vs. low) resource acquisition. The authors then identified seven cases of noticeably high or low levels of symbolic action. Note that the authors also went back to confirm this emergent theory with the original sample of 26 cases. In particular, the authors emphasized that

“We could not find a single case among the 26 venture projects in which a low level of symbolic actions correlated with a high level of success in consistently attracting resources, nor could we find a case in which a high level of symbolic actions correlated with a low level of success” (Zott and Huy, 2007: 76).

Far from sweeping the “untidy” results of the first coding under the model-fit carpet, the authors instead doggedly decided to follow up on those cases that did *not* fit the original assumptions. A dedicated redesign followed and uncovered the theoretically consequential-yet-hitherto-overlooked variable in resource acquisition, which turned into the key feature of the article’s theoretical story line. Ultimately, this approach enabled the authors to contribute to organization theory by demonstrating how variance in symbolic action management led to different resource acquisition outcomes, thereby generating a new theory that is both causally stronger (by introducing a so-far neglected variable), as well more generally applicable (the results could be replicated across all cases in the sample, with no more deviant cases in evidence), overall explaining a larger degree of variance. Put differently, the data-model *mis*-fit provided the impetus to consider alternative explanations of the phenomenon, providing a better fit of the new model with the existing data. In fact, the new focus on symbolic actions enabled the authors to nuance its impact even further by pointing to several moderating factors (structural similarity between resource holder and seeker, intrinsic quality of the venture, and uncertainty in the marketplace, see Zott and Huy, 2007: 36-39).

Highlighting the consequentiality of symbolic action as a key lever in resource acquisition represents a novel contribution to organization and entrepreneurship theory. In particular, prior theorizing on symbolic action had focused on how communication in response to an established organization's image affects its later performance (Elsbach, 1994; Marcus and Goodman, 1991; Sutton and Callahan, 1987), as well as how symbolic action and communication affect power and control relationships within firms (Westphal and Zajac, 1998). Unknown so far were the effects of symbolic action in resource acquisition strategies by new ventures. By disentangling them from impression management, the authors "hope to restore the full richness of symbolic actions in organization studies" (Zott and Huy, 2007: 45).

Towards a theory of relational coordination: Narrow vs. broad supervisory spans in a mixed methods study (Gittell, 2001).

Gittell (2001) points to competing hypotheses from the organizational literature regarding the role spans of control play for group process and performance. Based on postbureaucratic theory (Hackman and Oldham, 1980; Walton and Hackman, 1986), it would appear that supervisors play a potentially important role in support of group process, and this role would require broad spans of control (Gittell, 2001). By contrast, several other theories in organization science can be effectively summarized in that effective leadership is time consuming and relationship intensive, which implies that it may require narrow spans of control (Ancona, 1990; Eisenhardt and Tabrizi, 1995; Ford, 1981; Likert, 1961; Tannenbaum, 1968; Udy, 1959).

The author's approach is to test the two competing hypotheses using quantitative data and then to (re-)interpret the results based on qualitative data from the same study. The quantitative study finds that small supervisory spans improve performance through their positive effects on group processes such as higher levels of relational coordination among their direct reports. Initially, a regression of supervisory span on group performance had suggested that broad supervisory spans are significantly associated with lower levels of group performance (i.e., relational coordination), whereas narrow

supervisory spans are associated with higher levels of performance. In particular, a mediation analysis suggested that:

“Broad supervisory spans reduce objectively measured group performance through their negative effect on group process. In other words, group process mediates the negative effect of broad spans of control on performance” (Gittell, 2001: 476).

There were two deviant cases, which were analyzed qualitatively in a second step. The author introduces the results of the ensuing full-range DCM by suggesting upfront that “there are clearly other factors at work here” (Gittell, 2001: 479). The first deviant case was characterized by narrow spans and (unexpectedly) low levels of performance. Results suggested that although narrow spans of control tend to encourage supervisors to play a facilitative role with respect to group process, they do not require it (Gittell, 2001). In particular, closer inspection of the deviant case revealed that:

“Supervisors were observed to focus more on pressurizing group members and blaming other functions than on providing coaching and feedback to group members. [...] Despite stated attempts to buffer, this punitive approach to performance measurement appeared to encourage front-line supervisors to achieve performance through more expedient means than through building group process” (Gittell, 2001: 479).

The second deviant case suggested that although broad spans of control tend to discourage the development of strong group performance, they do not prevent it (the second deviant case was characterized by broad spans and high levels of performance). Overall, in-depth analysis of the two deviant cases pointed to additional variables, which turned out to be consequential for the model. In the first deviant case, it was found that although spans were relatively small, supervisors were observed to focus on pressuring group members and blaming other functions. In other words, while typically supervisors with small spans improve performance by strengthening group process by positive means such as mutual respect, shared goals and knowledge, and helping more generally,

supervisors can also use small spans of control to achieve performance in negative ways (as observed in the first deviant case). Conversely, broad spans can be associated with strong group process without much supervisory input with the help of supporting practices such as performance measures that focus on cross-functional accountability and the selection of group members for team orientation (the second deviant case). The author concludes by pointing to the theoretical contribution:

“Outliers suggest some countervailing [i.e., contrary to the postbureaucratic stream in organizational theory] considerations. Small spans can also serve as the opportunity for a more directive style of supervision, which at least in one case appears to have boosted performance by means other than high levels of relational coordination [among the team members of the first deviant case]. And larger spans [in the case of the second deviant case] were compatible with high levels of relational coordination” (Gittell, 2001: 480).

The theory building potential of the article hinges directly on the qualitative analysis of the outlying cases. Reconciling the apparent contradiction between competing views on the efficacy of small (vs. broad) supervisory spans would not have been possible without carefully collecting additional data in two teams which deviated markedly from the main body of observations. This allowed the author to not only address the conflicting predictions in organization theory, but also to propose a nuanced, new theory on relational coordination (prominently featured in the abstract of Gittell 2001):

“Qualitative data [from the deviant cases] suggest that supervisors with smaller spans achieved [higher levels of relational coordination] through working with, and providing intensive coaching and feedback to their peers” (Gittell, 2001: 468).

In summary, it should also be noted that the full-range DCM paper analyzed here was part of a larger study, which, beyond providing the basis for a new theory in

organization science, coincidentally also provided the basis for the author’s career in ‘relational coordination’ (see endnote 1 in Gittell, 2001).

Theory-building potential partially forgone

To contrast the theory-building gains and losses from engaging in purposeful analysis of deviant cases (vs. abstaining from such analysis), we now illustrate how the four remaining papers disclosed their existence. These papers sometimes even provided reflections on how the deviant cases might be influential theoretically, but eventually stopped short of following up this potential (Carney et al., 2011; Ferlie et al., 2005; Forte et al., 2000; Pisano, 1994). These articles are still highly noteworthy (and most commendable) for having flagged outlying observations, albeit without fully articulating potential contributions to theory building. Typically, these articles treated deviant cases as a kind of afterthought, acknowledging their existence, even sometimes speculating on factors that accounted for them (typically in the limitations section), but eventually not following up any of the unexpected lines of inquiry. Our intention here is not to chastise authors for failing to leverage the theory-building prowess of deviant cases. In fact, as we have seen from the scanty articles discussing deviant cases at all, the four papers we highlight here must be lauded for at least paving the way for several DCMs, which we summarize in Table 3.3. and discuss briefly below. Moreover, given the primary objectives of any given article, elaborating additional DCM analyses in the same article meet limits of space and other editorial considerations (such as a focused message in one article). As such, flagging such deviant cases provides the fertile ground for any researcher to conceptually pick up from there.

	Pisano (1994)	Carney et al. (2011)	Ferlie et al. (2005)	Forte et al. (2000)
<i>Research objective & initial model</i>	Findings showed that when the underlying scientific knowledge is sufficiently strong, effective learning may take place outside the final use environment in laboratories ('learning-before-doing'), shortening product development lead times, but only in environments of sufficiently developed scientific knowledge and process development experience. In newer, less developed environments, learning-before-doing may fail to shorten lead times.	Broad-based theoretical consent suggests that business groups emerge as a response to underdeveloped institutions or 'institutional voids' in developing economies and that group ties are beneficial in societies in which such voids continue to exist. Meta-analysis of a heterogeneous set of 28 jurisdictions was done to test the institutional voids thesis.	Two qualitative studies in the United Kingdom healthcare sector illustrate how social and cognitive boundaries between different professions retard the spread of innovation in healthcare and elsewhere. In particular, multi-professional organizations should face difficulty in spreading innovation.	Organizational adaptation to major environmental shifts was studied using the strategic types (Prospectors, Defenders, Analyzers, and Reactors). The authors performed quantitative cluster analyses to study the potential constraining effects of organizational form. They focused on average performance to prevent the undue influence of outliers.
<i>How was deviant case(s) identified?</i>	The observations confirmed the general pattern, but three deviant cases in the biotechnology subsample deviated from the model: here, 'learning-by-doing' was associated with shorter lead times.	During the meta-analysis, confirming evidence was found only in nations where group membership compensates for missing institutions, and in nations where the affiliates suffer from conglomerate performance discount observed in developed nations. Six deviant cases were identified: jurisdictions with (unexpectedly) high affiliate	The authors' research approach was qualitative analysis of extreme cases. In the course of this analysis, however, one deviant case was found in their dataset. This case, of laparoscopic hernia surgery, involved only one professional group, with no important boundaries. Yet, it was slow to diffuse, as adoption was	The analyses showed that only one Defender changed its organizational form in view of environmental shifts, and the majority of Defenders unexpectedly did not change. Here, the Defenders which did not change are the 'black swans' or outliers and the Defender which did change is the 'white

		performance and high institutional development; and low affiliate performance and low institutional development.	followed by ‘unadoption’.	swan’.
<i>What was done with them?</i>	Deviant cases were highlighted in the analysis and limitations sections. One characteristic was shared by all three biotech deviant cases: all had significant more process development experience than the others (thus making them quite similar to the chemical-based subsample).	The six deviant cases were flagged in the discussion section, encouraging further research into the reasons for their not confirming with expectations.	The existence of the deviant case was highlighted in the limitations section, suggesting future research avenues.	The existence of the Defender which did change was pointed out in the limitations section, suggesting that closer examination might lead to “critical information into how to better manage the transformation process” (p. 771).
<i>What could have been done with the deviant case(s)/ prediction outlier(s)?</i>	The three cases could have been subjected to a Multiple or Full-Range DCM to explore further the theoretical consequences of this shared characteristic. Further data collection and analysis could have focused on the presence/ absence of this variable in the remaining observations.	The deviant cases occupy two different quadrants, suggesting that a Full-Range DCM could have been performed on the six cases.	Single DCM could have been used to probe the characteristics of the deviant case at hand.	A Single DCM could have been performed to examine why only one case actually behaved as expected, whereas all other Defenders behaved in theoretically inconsistent ways.

<i>Foregone theoretical gains</i>	The critical characteristic which set the three deviant cases apart from the main body of observations (process development experience) seems to be the main explanatory factor here, rather than the environment in which they operate (chemical vs. biotech). Following it up could have completed the theory.	Further analysis of the outlying observations would have broadened the applicability of institutional voids theory to a wider set of jurisdictions.	The relationship between social and cognitive boundaries between professions may retard the spread of innovations. In the outlying case, there were no boundaries since it was developed by one professional group, only, and yet, it was slow to diffuse. Single DCM could have pointed to specific characteristics that moderate the effect of social and cognitive boundaries.	Theoretical expectations regarding the Defender organizational form seem to be the exception, rather than the rule, at least in the sample studied. Closer look at this case might have provided valuable information on how to manage the transformation process.
<i>Unanswered question of theoretical interest</i>	Whether or not learning-before-doing is good or bad for lead times may depend on sufficiently developed scientific knowledge and process development experience, <i>irrespective</i> of the industrial environment.	The authors state this themselves: “Why do business groups members do so well relative to unaffiliated firms in contexts with generally well-functioning institutions such as Malaysia, Singapore, and Sweden? And why do they do so unexpectedly poorly in contexts with severe voids such as Nigeria, Pakistan, and Peru?” (p. 453).	It remains unclear why and under what circumstances innovations developed by only one professional group experience delays in the spread of their innovations.	Under what conditions different types of organizational forms change remains unclear. In particular, the Defender type seems more change resistant than generally assumed – what are the reasons?

Table 3.3. Examples where deviant cases’ theory-building potential is partially realized

Is ‘learning-before-doing’ ultimately positive or negative for development lead times? Pisano (1994).

Pisano (1994) studied 23 process development projects in two environments (traditional chemical-based pharmaceutical environments and newer, biotechnology-based environments). In particular, the model suggested that when the underlying scientific knowledge is sufficiently strong, effective learning may take place outside the final use environment in laboratories (which the author calls ‘learning-before-doing’). Such learning-before-doing may be particularly prevalent in chemical-based pharmaceuticals, an environment characterized by deep theoretical and practical knowledge, and where it is associated with shorter lead times. By contrast, in biotechnology, an environment often considered more an art than a science, a greater emphasis on prior laboratory experimentation may fail to shorten process development lead times. Learning-before-doing was measured as the percentage of total project person-hours expended prior to the first pilot batch of production, and results indicated that (as expected) in the chemical-based environment, learning-by-doing was indeed associated with more rapid development, conversely, in the biotechnology subsample, learning-before-doing did not shorten lead times, thus pointing to a *negative* relationship between the focal variables.

Despite this general pattern, the author noticed three projects that markedly deviated from the other projects in his sample concerning the relationship between the share of project time expended prior to the first pilot batch of production and the projects’ adjusted lead time (Pisano, 1994). In fact, these three deviant cases seem to point to a *positive* relationship between learning-before-doing and lead times. Perusing these cases he notes (in the limitations section):

“One factor appeared to stand out very strongly: all three outlier projects were undertaken by organizations with relatively more biotechnology process development experience than the others. Clearly, a few outliers do not constitute a trend and conclusions cannot be drawn at this time. However, this investigation suggests that the relationships between experience, firm-specific knowledge, and learning strategies may be worthy of further analysis. One plausible hypothesis is

that experienced firms have accumulated deeper technical knowledge that can be tapped through research. A firm with little experience may be forced to 'learn-by-doing' until it accumulates enough understanding of the underlying technical parameters and interactions.” (Pisano, 1994: 98)

Acknowledging the need for further examination of this hypothesis, the author nonetheless decided not to pursue this lead in the 1994 paper, but to follow-up this research question in a separate paper, noting that “subsequent papers from this study will focus on these issues” (Pisano, 1994: 98). Careful search of the literature for this paper remained unsuccessful, though. Ultimately, Pisano’s pattern of outliers would have enabled the author to conduct a full range DCM (his three deviant cases were in both the relevant quadrants of the alternative theoretical prediction), in a way similar to Gittell (above). The necessary new (potentially qualitative) data collection and data (re-) analysis might have confirmed the speculation regarding the intricacies of learning-by-doing vs. learning-before-doing, potentially paving the way for a more nuanced understanding of prior process development experience in the context of biotechnology-based pharmaceuticals.

What are the limits of institutional voids theory? Carney et al. (2011).

The starting point of Carney and colleagues (2011) is to point to broad-based theoretical consent suggesting that business groups emerge as a response to underdeveloped institutions or “institutional voids” (Khanna and Palepu, 1997) in developing economies and that group ties are beneficial in societies in which such voids continue to exist (Carney and Gedajlovic, 2002). In a nutshell, the institutional voids thesis suggests that business groups internalize activities that otherwise fail to materialize due to limitations in a society’s financial, legal, and labor market institutions. The authors explore, via a meta-analysis, the institutional voids thesis, but find confirming evidence only in nations where group membership compensates for missing institutions, and also in nations where the affiliates suffer from conglomerate performance discount that is observed in developed nations.

Six deviant cases were identified after a meta-analysis of a heterogeneous set of 28 jurisdictions. In a plot juxtaposing institutional development and affiliate performance, these cases occupied the top-left and lower-right quadrants. These were jurisdictions with high affiliate performance and high institutional development; and low affiliate performance and low institutional development. The authors suggested

“The nations in the remaining two quadrants [the deviant cases] present some enigmatic questions for institutional voids theorists. Why do business groups’ members do so well relative to unaffiliated firms in contexts with generally well-functioning institutions such as Malaysia, Singapore, and Sweden? And why do they do so unexpectedly poorly in contexts with severe voids such as Nigeria, Pakistan, and Peru? Additional studies are needed to explore why extant institutional voids theory explains these outliers so poorly and to reveal which institutional variables are responsible for their counter-theorized positioning” (Carney et al., 2011: 453).

Beyond incentivizing their readers to address these questions, the authors do not engage in further theory building based on these outliers. Despite this, their study offers an excellent starting point for DCM. A range of outlying cases have been observed. By further analyses of these cases, institutional variables which are responsible for their counter-theorized positioning could be explored. They could provide rigorous answers to questions pertaining to the relatively superior performance of business group members when compared to unaffiliated firms in generally well-functioning contexts. There are enough deviant cases in either of the two counter-theorized quadrants to perform a multiple DCM within their particular groups as well as across the whole range of observed cases (full range DCM). This analysis might also solve the enigma behind the unexpectedly poor performance of some jurisdictions in contexts with severe voids.

Are multi-professional groups as positive or negative in the (non-) spread of innovations? Ferlie et al. (2005).

A qualitative study by Ferlie et al. (2005) focused on extreme performance. In a way similar to Zott and Huy (2007), Ferlie et al. first started by systematically comparing

what they call ‘positive’ and ‘negative’ outlier organizations, i.e., those that fit their theoretical predictions particularly well (in our terms, they compared pure model-fit outliers). In the course of this analysis, however, Ferlie et al. (2005), found a deviant case in their dataset (thus paving the way for a Single DCM), but chose *not* to further investigate this case. Instead, they reported the occurrence of this case in the limitations section, acknowledging its existence, but refraining from further analysis and leaving the original design untouched. In the original design, two qualitative studies in the United Kingdom healthcare sector illustrate how social and cognitive boundaries between different professions retard the spread of innovation in healthcare and elsewhere. The model predicted that multiprofessional organizations face difficulty in spreading innovation. This design was tested in two studies. There was

“one deviant case that we cannot easily explain: the case of laparoscopic hernia surgery. This potential adoption involved only one professional group, with no important boundaries. Yet, it was slow to diffuse, as adoption was followed by ‘unadoption’” (Ferlie et al., 2005: 132).

It should be appreciated that the authors actually reported the existence of the deviant case. However, their reluctance to theoretically engage the unexpected finding may be counter-productive from a theory-building perspective. Had the authors chosen to analyze the deviant case where innovations were slow to diffuse despite there not being a multi-professional organizational structure, they might have found out why it was being unusual with respect to the phenomenon of interest and an even more substantive contribution to theory could perhaps have been made.

Reporting the presence of outlying or deviant observations in their study as grounds for further research clearly is a first step in the right direction. In this case, the analysis of a single deviant case, representing “Y-oriented” research as described above (Blatter and Haverland, 2012), would help in pointing out the boundary conditions (moderating effects) and even process explanations (mediating effects) of this particular theory that, prior to the discovery of the deviant case, remained unchallenged. Ultimately, although representing the most basic DCM research strategy, even a single

DCM would have enabled the authors to peruse more carefully the anatomy of the deviant case which, despite being not multiprofessional, was still slow to diffuse. Insights into the drivers behind this outcome might provide a more nuanced outlook in the direction of the moderating effect of professional groups in the (non-)spread of innovations.

Strategic types: Are defenders likely to change or not? Forte et al. (2000).

The main focus of this quantitative study was on organizational adaptation to major environmental shifts by empirically examining the potential constraining effects of organizational form, which was conceptualized using Miles and Snow’s (1978) strategic types (Prospectors, Defenders, Analyzers, and Reactors). In the introductory pages of Forte et al. (2000), the authors explicitly suggest that their focus was on average performance as a way to “guard against the undue influence of outliers” (Forte et al. 2000: 757). From a theory-building perspective the most noteworthy model misfit was that only one Defender changed its organizational form in view of environmental shifts (it became an Analyzer), and the majority of Defenders unexpectedly did not change. Put differently, all except one Defender did *not* fulfill theoretical expectations (i.e., they did not change their organizational form), resulting in the ironic situation that the “outlying” nature of this case comes from it being the only one behaving as theoretically expected:

“In stark contrast to the apparently effective responses of many of the Reactors, was the apparent inability of most of the Defender hospitals to effectively transform themselves into the forms better matched to the merging industry conditions. In fact, only one Defender hospital in our sample was able to do so. *We did not expect this*, given the ubiquitous transformations of Defender hospitals observed by Zajac and Shortell (1989) in their study. Perhaps methodological and sample differences between the two studies account for this discrepancy. Still, our results are very much in line with theory and anecdotal case evidence (Fox-Wolfgramm et al. 1998; Meyer 1982; Miles and Snow 1978) indicating that Defenders are the least likely of all of the organizational forms to both notice the need to change and be able to overcome the competency and

process constraints of their organizational form. That is, Defender organizations may be the least likely and most challenged to move outside their strategic comfort zones (Shortell et al., 1990)” (Forte et al., 2000: 770, added emphasis).

Analyzing more deeply the majority of Defenders in the sample resisting change might have provided additional vistas into the apparent inconsistency between the two camps’ views on the ability of Defenders to change their organizational form. In particular, the evidence for Defenders’ resistance to change has been anecdotal so far, and Forte et al. (2000) would have been in an excellent position to provide new empirical evidence on the apparent resilience of the Defender form, thereby potentially providing insight into variables that may explain the non-additive results by the two camps so far. Instead, the authors chose to point out that lack of data (only one Defender behaved as expected) as an excuse for not testing one of their hypothesis:

“Since only one Defender changed to the Analyzer form and no Defenders changed to the Prospector form, lack of data precluded testing Hypotheses 4c and 4d that Defenders prior to the environmental shift that change to the Analyzer or Prospector form will have higher performance after the environmental shift than Defenders that do not change their form” (Forte et al., 2000: 767).

In the limitations section the authors then refer to this one deviant case in their sample, which actually complied with initial theoretical expectations, acknowledging very briefly its theoretical potential:

“One Defender did transform itself into an Analyzer. [...] Closer examination of these outliers may provide critical information on how to better manage the transformation process” (Forte et al., 2000: 771).

Overall, Forte et al. (2000) seem to have found not one black swan (refuting the theory that all swans are white) but actually only black swans (none of the Defenders except one changed their organizational form), except for one case of a white swan.

Examining this white swan in greater depth might have provided critical insights into organization theory and into the conditions under which Defenders change their organizational form.

3.7. General discussion: Theoretical gains (and losses) from deviant case analysis

Not all outliers are theoretically consequential. As we have explained, deviant cases, in particular, hold the greatest theory-building potential, while pure model fit outliers are less indicative of the need to refine or newly develop theory, and construct outliers require a minimum of additional analysis to reveal their theoretical consequentiality (Aguinis et al., 2013; Cortina, 2002). Depending on their number and position in the scatterplot, different deviant cases will accommodate different DCM methods, with varying degrees of theoretical gains. We hope that our discussion of the different types of outliers and the different types of analysis strategies for deviant cases has outlined the boundary conditions in terms of theoretical gains and losses in organization science. When the “oops-I’ve-got-an-outlier” moment comes, we encourage authors to consider the potential for significant advancement of (parsimonious) theory gained from the exploratory investigation of deviant cases. Our call for a more constructive treatment of deviant cases is not to be misunderstood as a manifesto for overly complex theories that account for every data point on the scatter plot. The message is that any deviant case can hold theoretical potential, and unless we analyze it, we can never be sure of its theory-building potential.

The empirical reality, however, is that very few authors publishing in high-profile outlets seem to heed the theory-building prowess of deviant cases. Four authors provided some kind of reflection on the theoretical significance of the encountered outlying cases. Although relaying to the reader (usually in the limitations section) the discovery of a potentially interesting deviant case, these papers eventually abstain from following up the theoretical implications. Only two papers in 20 years of research published in top tier organization and management science outlets systematically disclosed and used model-data *misfit* as a starting point for theory building in

organization science. This practice wants rectifying since it comes at the cost of foregone opportunities for developing greater theoretical purchase on empirical realities.

What might be the reasons for the rarely consummated theory-building potential of deviant cases? To begin with, editorial policies about article length-to-contribution ratio come to mind. For example, the considerable extra space needed to report a sufficiently developed Single DCM might deter authors (and editors) of top-tier journal articles, because the value added by examining only a single case might not be justifying enough for further expanding an otherwise fully developed and “coherent” paper. For instance, in a 40-page manuscript analyzing hundreds of quantitative observations, should scholars devote four pages analyzing the seven deviant cases excluded from the regression analysis? From a theory development (rather than testing) standpoint, the answer would often be yes. At least, though, authors should flag such cases and hence provide the opportunity for other scholars to pick up on such deviant cases in their research.

Note that although the number of outlying observations is typically small, the new or refined theory does *not* only apply to these few cases. Rather, these are the cases where the “hidden” phenomenon is most blatantly visible, and it is for this reason that they point researchers in the right direction in the various avenues for theory refinement. As in the example on intelligence and job performance we used above, emotional intelligence is likely to compensate for a lack of (cognitive) intelligence for the average employee as well, potentially improving model fit with the additional variable included. However, in the stage when this additional explanatory variable is not yet identified and included in the statistical models, the lack of this predictor will become clearly visible *only* for those cases that represent an extreme coincidence of attribute specification (in this case, e.g., very low IQ together with very high job performance). For other, more average (less deviant), cases the missing predictor simply disappears in the noise of unexplained variance. Thus, at first glance, the disproportionately small numbers of deviant cases (compared to the complete sample) might lead to a kind of theoretical complacency particularly if authors perceive deviant cases as gratuitous, mistaking the

relevance of the new theory as limited only to those specific (in our example, seven) deviant cases. The point is that deviant cases may wield their theoretical heft not only over themselves but also over the remaining observations in the sample, more fully explaining the focal relationship for the entire sample. The potential of deviant cases rests in their capacity to outsoar their idiosyncratic characteristics, their paraphernalia, and the apparent (but misleading) empirical provinciality in which they were first established.

We also need to ask ourselves the question whether scholars should spend more time in a paper on topic A, discussing something other than topic A (as in the DCM). If authors laser-focus a research design to explain a particular range of phenomena, it is generally not to spend time and energies on discussing what does *not* fall within that range. Of course this point is well taken: it surely would not be deemed appropriate by the recipients of journal articles if authors deviate from the initially focused topic after having identified interesting outliers in their sample. It should be appreciated, however, that deviant cases are spotted due to large residuals in statistical analyses on the focal relationships (i.e., topic A). As such, applying a DCM on these cases enables a much more in-depth and constructive discussion of just that particular topic A.

Rather than a switch in topic, DCM often requires a switch in methods when examining the deviant cases. For example, in a quantitative study the DCM strategies are likely to involve qualitative methods of analysis (unless there are a large enough number of deviant cases to be examined using quantitative methods again), even though the outliers may appear in quantitative studies. In this sense, our call for applying DCM for theory building joins the calls for more mixed methods research in management and organizational research (e.g., Aguinis et al., 2010; Daft and Lewin, 1990; Edmondson and McManus, 2007; Hitt et al., 1998). The combination of qualitative and quantitative methodologies is suggested to “provide a better understanding of research problems than either approach alone” (Molina-Azorin, 2012: 34). They are also highly appreciated by the research community, as evident, for instance, in the larger numbers of citations these articles tend to attract when compared to mono-method articles (Molina-Azorin, 2012).

We realize that there may be instances precluding the meaningful incorporation of DCM strategies in a paper, even when deviant cases of high theory building potential are discovered. Such instances can occur, for example, when there are constraints in article length that prevent adding further analyses, when researchers are not familiar with the (qualitative) methods necessary for carrying out the DCM strategies, or when there is not sufficient data available to probe the deviant case in detail. In these cases, however, the identified cases need not to disappear in a drawer, but can be explored in subsequent publications or research projects. Thus, when pointing to the foregone theory-building potential of the four papers above (Carney et al., 2011; Ferlie et al., 2005; Forte et al., 2000; Pisano, 1994), we do not wish to reprimand the authors. Quite the opposite: our objective is to applaud those papers that transparently reported the existence of deviant cases as well as provide a frank acknowledgement regarding their potential theoretical significance, whether in the results section (Forte et al., 2000, being perhaps the most illustrative example here) or in the limitations section as a kind of disclaimer (e.g., in the case of Ferlie et al., 2005). The majority of authors who detected (and reported) deviant cases did not in fact trouble themselves with perusing their theoretical significance at all. Against this peer group, the four papers above stand out by virtue of allowing us to at least speculate about the theory-building potential of the detected deviant cases.

The four papers also serve as a means for putting into even higher relief the methodological sophistication of the two papers that actually did perform DCM (Gittell, 2001; Zott and Huy, 2007). They serve to highlight just which theoretical aims can be hit if authors are prepared (and given the space to, by editors and reviewers) to go the extra mile and perform one of the three DCM strategies. Specifically, Single DCM and Multiple DCM represent productive methods for testing the boundaries of existing theories, and may hold the potential to point to additional, and so far neglected, variables that might be consequential for theory building. Full Range DCM represents the most sophisticated theory-*building* instrument at our disposal. Methodologically, it picks up the additional variables and manifests how they behave along a range of possible outcomes, thereby going beyond testing the boundary conditions of an established theory and actually moving towards advancing extant theory or building entirely new theory. If

outliers do occur, the researcher may indeed need to go back to the study’s original theoretical propositions and reconsider the causalities underlying the theory in the light of the case or cases deviating from the norm. As a result, the existing theory may have to be changed and potentially also the research design, including data collection methods, analysis techniques, and sampling criteria for additional cases. Ultimately, the cogency of the research results depend directly on how compelling the claim is that observed outcomes are caused by the independent variables of interest, rather than other potentially plausible (but omitted) factors. Yin admonishes his readers that

“Without such redesign you risk being accused of distorting or ignoring the discovery, just to accommodate the original design. This condition leads quickly to a further accusation: that you have been selective in reporting your data to suit your preconceived ideas” (Yin, 2003: 51).

Redesign may involve theoretical sampling of additional cases for the purpose of replicating (for the use of multiple DCM) or complementing (for full range DCM) the deviant case, as well as making changes in the case study data collection protocol more generally, in particular with regard to control variables that would, if included in the design, influence the results.

Note that we also do not wish to denigrate the importance of serendipitous finds of outliers; that is, when outliers deviate not from the general pattern of observations in that same article, but from a more generally accepted theory. The rife aversion to qualitative “single cases” seems misplaced in the context of single DCM, where the single case does not in fact stand on its own but is compared with the main body of results from which it deviates, making it powerful theoretically (e.g., Gerring, 2007). In sum, the fundamental message for organizational scholars is to not (per default) ignore deviant cases as gratuitous oddities, but to deliberately go after them in our quest to explore possibilities off the beaten theoretical tracks. On this note, consider Lewin (1992: 14):

“A common admonition in statistical methodology texts and in doctoral research methodology seminars involves the importance of examining and understanding

outlier data. In practice, however, because of a research tradition that places great value on fitting models that account for the greatest amount of variance, researchers have honed their arguments for discarding outliers.”

A final caveat is very much in order at this stage. The essay here is not a call for conjuring up (new) theory at all cost. Instead, it aims at suggesting “how” to use deviant cases constructively, “why” using them is beneficial theoretically, and that (most likely) many opportunities for theory building have been foregone over the last 20 years.

3.8. Conclusion

In conclusion, the three DCM strategies provide a widely applicable approach to perform constructive in-depth deviant case analysis for theory building. We hope that the DCM strategies bring deviant cases (back) in the focus of management and organizational research. In many ways, a hallmark of solid theory building is that it is based firmly on a dedicated engagement with cases that deviated from an initially plausible theory. If transparently performed, theory building from deviant cases improves the likelihood of producing theoretical contributions (i.e., contributions that go beyond mainstream conceptual development, e.g., Shapira, 2011), which often decide the fate of a manuscript above and beyond methodological sophistication (how many times have we read rejection letters which were based on lacking theoretical contribution despite perfect methodological craftsmanship?). Pointing to deviant cases is part of a larger call to researchers that a systematic assessment of model fit can be a powerful tool for theory building (apart from, say, unusual residual distributions or a low model R-squared, artifacts in these distributions can plausibly point to interesting things going on in the data that are not captured by the model). The methods we outline in this essay are intended to enable and encourage a more exploratory approach alongside the paths of confirmatory (hypotheses-testing) research. In fact, another merit of confirmatory research is to provide exactly that opportunity of empirically detecting deviant cases, while the three methods proposed here provide the means to explore them, expanding the range of questions that organization researchers can answer.

Chapter IV*

Analyzing inconsistent cases in management fsQCA studies: A methodological manifesto

Lakshmi Balachandran Nair

Prof. Michael Gibbert

Journal

Nair, L. B., & Gibbert, M. (2015). Analyzing inconsistent cases in Management fsQCA studies: A methodological manifesto. Journal of Business Research, 69 (4), 1464-1470.

Conferences

Nair, L.B., Gibbert, M. (2016, July). EGOS Conference, Naples, Italy.

Nair, L.B., Gibbert, M. (2015, July). 5th GIKA-JBR Conference, Valencia, Spain.

Nair, L.B., Gibbert, M. (2015, June). European Academy of Management Conference, Warsaw, Poland.

Nair, L.B., Gibbert, M. (2015, January). QCA - Social Science Applications and Methodological Challenges Conference, Tilburg, Netherlands.

Nair, L.B., Gibbert, M. (2014, September). Industrial Marketing & Purchasing Conference, Bordeaux, France.

Nair, L.B., Gibbert, M. (2014, August). 74th Academy of Management Annual Meeting QCA PDW, Philadelphia, USA.

Nair, L.B., Gibbert, M. (2013, October). International QCA Expert Workshop, Zurich, Switzerland.

Journal

4.1. Abstract

Cases inconsistent with theoretical expectations are by default indicators for a lack of theory-data fit, and as such are prime candidates for theory building. However, there is a tendency to ignore inconsistent cases in management research. The current article focuses on the theory-building prowess of inconsistent or deviant cases which turn up during an fsQCA study. The study looks at some of the key tenets of QCA: A cross-tabulation of cause and effect can demonstrate superior explanatory completeness only if one can account for all cases (be they deviant or not). To improve the neat theory-data fit characteristic of QCA, the paper proposes two new strategies for analyzing inconsistent cases of necessity and sufficiency in fuzzy set QCA studies and discusses their contributions to methodological sophistication.

Key words: QCA, inconsistent cases, theory building, multi method research, case oriented approach

4.2. Introduction

An important arena for using multi method research is in the analysis of inconsistent cases or outliers. The inconsistent cases can occur both in quantitative survey-based research designs as well as in qualitative case study research, and are often clear indicators for a lack of theory-data fit, and as such are prime candidates for theory building (see Katz, 1988; Sullivan, 2011). A research design allowing for outlier analysis would therefore help re-establish the cases' significance in the theorizing process (Locke, Golden-Biddle, & Feldman, 2008). In management, despite their theory-building competence, researchers tend to ignore, and sweep the outliers under the proverbial carpet of model fit (Aguinis, Gottfredson, & Joo, 2013; Gibbert, Nair, & Weiss, 2014). Typically, outliers are small in number, calling for a qualitative, and often comparative, analysis of the cases that deviate from the main body of observations (Kendall & Wolf, 1949; Pearce, 2002).

Fundamentally, Qualitative Comparative Analysis (QCA) is herein a particularly promising method. One of QCA's hallmarks, the cross-tabulation of cause and effect, can demonstrate superior explanatory completeness only if one accounts for all cases (be they deviant or not). To improve the neat theory-data fit characteristic of QCA, the current study proposes two new strategies for analyzing inconsistent cases of necessity and sufficiency in fuzzy set QCA studies. Specifically, the paper conceptually details two post-QCA research strategies, namely "Comparative Outlier Analysis" (COA), and discusses their relative merits in terms of building theory more rigorously. The paper deliberates examples for both strategies, from prior studies on and applying QCA. With QCA progressively acknowledged as a viable research method in management and organization research (Greckhamer, Misangyi, Elms, & Lacey, 2008; Fiss, 2007; Fiss, 2011), a research design integrating QCA and COA will be not only timely, but also very relevant for theoretical advancement in management research as well as methodological progress in QCA.

4.3. Case oriented strategies in QCA

Prior research addresses the importance of case-oriented strategies for supplementing QCA studies. Case study research serves as a pre-QCA step (Schneider & Rohlfing, 2013) providing help in ascertaining the specific cases (Ragin, 2000), establishing and calibrating the relevant conditions included in the analysis (Berg-Schlusser, De Meur, Rihoux, & Ragin, 2009), and addressing contradictory configurations and outcomes (Ragin, 1987). Case oriented strategies are also applicable simultaneously during the course of QCA studies for refining causal arguments and addressing “empirical refutation of initial arguments” (Ragin & Schneider, 2012).

Finally, case studies are also suitable for identifying post-QCA deviant cases. Process tracing a deviant case is helpful in further improving the theory and the QCA model (Schneider & Rohlfing, 2013). The present paper moves beyond process tracing of a single deviant case and offers further steps for comparatively analyzing a series of post-QCA deviant cases. The focus here is predominantly on fsQCA cases, so as to present a clear application scenario for the application of the COA technique. Clearly, the underlying strategy is also suitable for analyzing csQCA deviant cases. But for expository purposes and space constraints, the current study does not discuss csQCA deviant case scenarios.

4.4. Why analyze the inconsistent cases in QCA?

The QCA technique bases itself on Mill’s “canons,” especially the Method of Difference and the Method of Agreement. The Method of Difference states that if an instance where a phenomenon occurs and an instance where a phenomenon does not occur have every circumstance in common except one, which occurs in the former instance and not in the latter, the particular circumstance is the cause, or an indispensable part of the cause, of the phenomenon. The Method of Agreement instead deals with similarities in observed cases. The logic states that if two or more instances of a phenomenon show only one circumstance in common, that particular circumstance then constitutes the cause or effect of the given phenomenon (Mill, 1875). Both methods have to do with establishing common causal relationships by eliminating all other possible alternative explanations

(Berg-Schlosser et al., 2009). Nonetheless, rigid positivistic assumptions of cause and effect relationships may not always work so neatly in the case of social sciences, where a multitude of causes and conditions intersect in time and space to produce a particular outcome.

In fact, non-conforming cases may even be more important than typical cases as they can be of great help in the understanding of causal complexity. Unlike in other research methods, where researchers neglect deviant cases or outliers as unavoidable nuisance, QCA “tends to give explanations without dismissing exceptions or outliers” (Berg-Schlosser et al., 2009). The tendency occurs because most studies undertaking a regression or similar methods focus on averaging out the large numbers of cases under study. QCA, meanwhile, takes into consideration even a combination of conditions that explains only a single case (Berg-Schlosser et al., 2009).

However, sometimes even in QCA, researchers tend to reject cases inconsistent with expectations (Rihoux & Ragin, 2009). The rejection significantly reduces the concerned study’s explanatory power (i.e. internal validity). The rationale behind the above rife tendency is the belief that the exceptional circumstances are somehow idiosyncratic, unlikely to be repeated elsewhere, and are therefore of little interest theoretically. From the perspective of the philosophy of science, clearly one would be unwise to take the same stance. Instead of trying to resolve or ignore deviant cases, authors should rather accept paradoxical cases, and use them constructively for theory development (Gibbert, 2006). In addition, one of the hallmarks of QCA is that a cross-tabulation of cause and effect can demonstrate superior explanatory completeness (Western, 2001) only if there is the possibility of accounting for all cases, deviant or not (Ragin, 1987). The present study hones in on the very essence of QCA by providing a road map for analyzing the deviant cases identified during a QCA study in terms of two concrete strategies.

Naturally, since deviant cases are assessable only in relation to an explicitly framed theoretical model, the relative deviance of a case would also change when altering the general model (Seawright & Gerring, 2008). In a similar vein, what

constitutes a QCA deviant case is very much in the eye of the beholder, and so depends upon the specifications made by the researcher regarding just which causal conditions apply, and which individual cases to include or exclude from the initial study. Without any reflection and systematic application to the available data, there would be a serious impairment to both the internal as well as the external validity of the emergent theory.

4.5. COA- How to analyze outliers in a QCA?

The current section details the methodological sophistication of the two COA designs and their potential in terms of providing stronger (i.e. more internally and externally valid) theory. Outlier analysis techniques have their basis in Yin's (2003) "Replication Logic." Replication Logic is a tool for selecting further cases for comparison with an initial case, in order to enhance a study's potential for making causal claims. Two basic approaches exist; Literal Replication and Theoretical Replication (Yin, 2003).

"Literal Replication" involves comparing cases that are similar to each other, are from the same end of the theoretical spectrum, and could predict similar results (Yin, 2003; Gibbert and Nair, 2013). The Literal Replication procedure is similar to the Replication Deviant Case Method (Gibbert et al., 2014) or the Most similar system design- Method of agreement (Faure, 1994; Levi-Faur, 2006), both of which involve a comparison of cases which are very similar to each other in terms of the phenomena of interest. The cases are similar not only in terms of the causal condition, but also in terms of the outcome. Outlier Literal Replication (OLR), the first COA strategy, has its basis herein.

"Theoretical Replication," on the other hand, involves the comparison of cases from different ends of a theoretical model and the prediction of dissimilar results, but due to foreseeable reasons (Yin, 2003; Gibbert and Nair, 2013). The strategy is comparable to the "Full Range Deviant Case Method" (Gibbert et al., 2014) in qualitative and quantitative studies. As per the Method of Difference and Method of Agreement, two possible types of Outlier Theoretical Replication (OTR) could exist (Faure, 1994; Levi-Faur, 2006). The first design is the Most similar system (Method of Difference). In the design, the comparison is between cases which show the same causal

condition as typical cases, but with a different (though theoretically predictable) outcome. The underlying notion is that, the more comparable the cases are with respect to the causal conditions, the greater the feasibility to segregate factors which cause the disparity in the outcome. The second type of OTR is the Most different system (Method of Agreement). The design considers cases which are similar in the outcome shown, but differ from each other in terms of the causal conditions. Comparing cases which differ with regard to the causal conditions supposedly causing the observed effects (Gibbert and Nair, 2013) helps in exploring plausible alternative explanations, the development of new causal claims (Gibbert and Ruigrok, 2010) and, ultimately, building stronger theory (i.e. theory which is both internally and externally valid).

For emphasis, one can recall that the outlier analysis techniques would differ in csQCA and fsQCA, and also with respect to the necessity and sufficiency of conditions. Here, the focus is specifically on the fsQCA inconsistent cases. As fsQCA handles varying degrees of membership in the causal condition and outcome (Ragin, 2008), a wide range of typical and outlying cases are available for COA, making fsQCA particularly attractive and instructive as an application context.

4.6. Analysis of multiple outliers in fsQCA

Analyzing and comparing multiple outliers is of great importance in theory building because doing so would help establish the internal validity and external validity of the modified proposition (Lijphart, 1971). COA permits the researcher to do within-case and between-case comparisons that would help determine whether the deviance noticed is just a one-off case, or is observable across several cases (Eisenhardt & Graebner, 2007). Unlike in prior studies on outlier analysis techniques, the general purpose of COA is not to find the necessary and sufficient conditions for an outcome and explain how each term relates to the phenomenon of interest exclusively (Berg-Schlosser et al., 2009; George & Bennett, 2005). Rather, the focus is on ascertaining the plausibility of causal relationships between variables (Mill, 1875; Cook & Campbell, 1979), expanding the scope of the theory in hand, or even laying the foundations for a new theory. For the same reason the focus here is not solely on positive outcomes, but on different

combinations of memberships and outcomes. The first step here would also be conducting the analysis of a single deviant case. The specification of the causal recipe herein could in turn function as a hypothesis about other cases (Ragin & Schneider, 2012). The OLR and OTR would follow consecutively. Tables 4.1 and 4.2 adapt elements from Levi-Faur’s (2006) inferential strategies and Yin’s (2003) Replication Logic, and show all the COA techniques in fsQCA.

Comparative Outlier Analysis: fsQCA (Necessity)

Outlier Literal Replication	<p><i>Most similar system (Method of Agreement)</i></p> <p><i>Similar causal condition (X), similar outcome (Y)</i></p> <ol style="list-style-type: none"> 1. Deviant cases of consistency (degree) and deviant cases of consistency (degree) 2. Deviant cases of consistency (kind) and deviant cases of consistency (kind) 	
Outlier Theoretical Replication	<p><i>Most similar system (Method of Difference)</i></p> <p><i>Similar causal condition (X), different outcome (Y)</i></p> <ol style="list-style-type: none"> 1. Typical cases and deviant cases of consistency (degree) 	<p><i>Most different system (Method of Agreement)</i></p> <p><i>Different causal condition (X), similar outcome (Y)</i></p> <ol style="list-style-type: none"> 1. Typical cases and deviant cases of consistency (kind)
Irrelevant for Comparative Outlier Analysis	<p><i>Most different system (Method of Difference)</i></p> <p><i>Different causal condition (X), different outcome (Y)</i></p>	

Typical cases: high X, high Y; Deviant cases of consistency (degree): high X, high Y, inconsistent with statement of necessity; Deviant cases of consistency (kind): low X, high Y

Note: In Most similar system (Method of Difference), the typical cases and deviant cases of consistency (degree) differ only with respect to the statement of necessity.

Table 4.1. COA in fsQCA Necessity

Comparative Outlier Analysis: fsQCA (Sufficiency)

Outlier Literal Replication	<p><i>Most similar system (Method of Agreement)</i></p> <p><i>Similar causal condition (X), similar outcome (Y)</i></p> <ol style="list-style-type: none"> 1. Deviant cases of consistency (degree) and deviant cases of consistency (degree) 2. Deviant cases of consistency (kind) and deviant cases of consistency (kind) 3. Deviant cases for coverage and deviant cases for coverage 	
Outlier Theoretical Replication	<p><i>Most similar system (Method of Difference)</i></p> <p><i>Similar causal condition (X), different outcome (Y)</i></p> <ol style="list-style-type: none"> 1. Typical cases and deviant cases of consistency (degree) 2. Typical cases and deviant cases of consistency (kind) 	<p><i>Most different system (Method of Agreement)</i></p> <p><i>Different causal condition (X), similar outcome (Y)</i></p> <ol style="list-style-type: none"> 1. Typical cases and deviant cases for coverage
Irrelevant for Comparative Outlier Analysis	<p><i>Most different system (Method of Difference)</i></p> <p><i>Different causal condition (X), different outcome (Y)</i></p>	

Typical cases: high X, high Y; Deviant cases of consistency (degree): high X, high Y, inconsistent with statement of sufficiency; Deviant cases of consistency (kind): high X, low Y; Deviant cases for coverage: low X, high Y

Note: In Most similar system (Method of Difference), the typical cases and deviant cases of consistency (degree) differ only with respect to the statement of sufficiency.

Table 4.2. COA in fsQCA Sufficiency

COA in fsQCA cases of Necessity

In necessity, the instances of outcome constitute a subset of the instances of the causal condition (Rihoux & Ragin, 2009). The subset relation signals the connection between the causal conditions and the outcome. Accordingly, instances inconsistent with the subset relation are not typical cases. Based on the presence and absence of an outcome in the presence and absence of a necessary condition, cases can be classified.

In fsQCA of Necessity, suitable cases for COA would be the most and least likely typical cases (Beach & Pederson, 2013) and deviant cases of consistency, subdivided into cases of degree and kind (Schneider & Rohlfing, 2013). Figure 4.1. shows the cases in fsQCA of Necessity, which could be suitable candidates for COA.

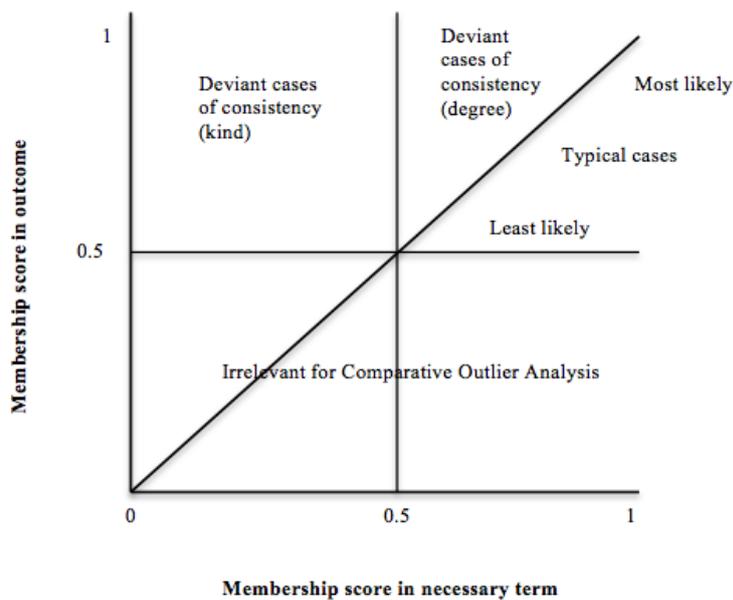


Figure 4.1. Type of cases in fsQCA (Necessity) Adapted from Schneider & Rohlfing, 2013

For comparative analysis, the cases where the necessary condition is absent but the outcome is present, that is, the deviant cases of consistency (kind), can be considered first. The cases' occurrence hints at the possibility that another condition or alternative explanation could have been present for the outcome in question to happen. By applying COA, the researcher could suggest a modification of the theory in hand that is externally valid beyond the case in hand. The second possibility involves deviant cases of consistency (degree). Deviant cases (degree) are qualitatively quite similar to typical cases, except that they do not confirm to the statement of necessity (Schneider & Rohlfing, 2013).

Typical cases are cases which have membership in both the necessary term and outcome. The typical cases can again be subdivided into “most-likely” and “least-likely” cases depending on their respective positions within the zone (Mahoney & Goertz, 2004) of typical cases. Most-likely cases have high values on both the necessary term and outcome, while least-likely cases have low values. The remaining fsQCA Necessity case-types are not anomalous by reference to the general model of causal relations (Maggetti, Radelli, & Gilardi, 2012), and hence are irrelevant for COA.

OLR in fsQCA cases of Necessity

The generalizability and theory building potential of analyzing a single deviant case is relatively limited (relative to analyzing several cases in a replication mode). If the purpose of the research is greater generalizability and theory enhancement, there is the possibility of a comparison with other similar outcome deviant cases. The researcher embarks on within-case, and between-case comparisons to understand and establish the applicability of the causal factors across a broader spectrum of cases (whether consistent or inconsistent with theoretical expectations). In the case of deviant cases of consistency occurring while testing necessity, one can carry out an OLR, a.k.a. comparison of the deviant cases of consistency with other similar deviant cases. The focus is on the causal conditions that the positive cases have in common, while there is an elimination of conditions that have nothing in common as they have no explanatory power here (Blatter & Haverland, 2012).

Analysis of fsQCA deviant cases of necessity would require a comparison of deviant cases of consistency (degree and kind). The deviant cases of consistency (kind) are situated in the top left-hand corner of the plot. Literal Replication of deviant cases of consistency (kind) with other deviant cases of consistency from the same location, is the most sophisticated OLR technique. Another possible category, deviant cases of consistency (degree), is in the top right-hand corner of the plot. Comparing the deviant cases of consistency (degree), with each other would be another interesting OLR technique.

OTR in fsQCA cases of Necessity

An even more sophisticated analysis technique compares cases with different outcomes with an eye on reconciling any differences theoretically. OTR suggests that the researcher deliberately chooses outcomes that are on the extreme ends of an emerging, theoretical continuum (e.g. high/low values on the causal recipe). When compared with analysis of a single deviant case or OLR, the technique is particularly advanced methodologically, as well as being more promising theoretically. Since there is an examination of cases from extreme ends of a theoretical spectrum, the theoretical purchase on the empirical data (internal validity) increases, and the scope of the theory becomes broader (enhancing the external validity).

Recall that the focus here is on cases with similar necessary conditions, but different results; a comparison of typical cases (most and least likely) with the deviant cases of consistency (degree). Here, the outlying cases are quite similar to the typical cases in terms of the outcome. As discussed earlier, this OTR design is quite similar to the Most similar system (Method of Difference). A comparison would thus lead to a refined hypothesis, the identification of further necessary conditions, or clarification of the scope conditions (Blatter & Haverland, 2012). To sum up, OTR goes beyond refuting an existing theory and moves towards formulating a potentially new theory. OTR thus represents a kind of recipe for creating a theoretical contribution.

The second type of OTR of fsQCA cases of Necessity involves the deviant cases of consistency (kind) with typical cases. Here the deviant cases have a causal mechanism different from typical cases, and yet they show the outcome. Comparing the typical cases (most and least likely) with the deviant cases of consistency (kind) might point to new causal conditions or alternate mechanisms in play which could have caused the occurrence of the outcome. The OTR design under discussion here is similar to the Most different system (Method of Agreement).

COA in fsQCA cases of Sufficiency

FsQCA Sufficiency provides a myriad of cases (Schneider & Rohlfing, 2013) for COA (see Figure 4.2.).

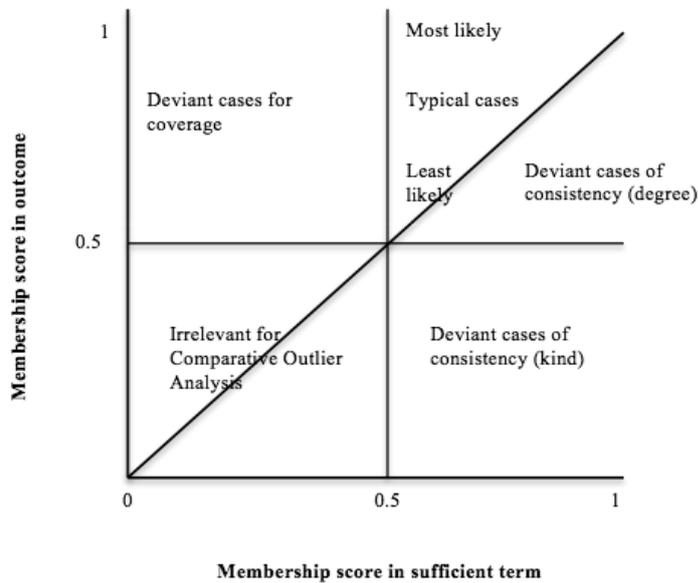


Figure 4.2. Type of cases in fsQCA (Sufficiency) Adapted from Schneider & Rohlfing, 2013

Deviant cases for consistency contradict the statement of sufficiency by being a member of the QCA solution, but not being a member of the outcome. The typical cases are the ones which lie on the upper right-hand side of the spectrum (above the diagonal as they are consistent with the statement of sufficiency). The cases in the lower half of the diagonal are inconsistent with the statement of sufficiency and are thus deviant. As in the case of the fsQCA cases of Necessity, one can also subdivide the cases into deviant cases of consistency (degree and kind). The former occupy the location below the diagonal adjacent to the typical cases, whereas the latter occupy the lower right-hand side of the theoretical prediction.

Deviant cases for coverage, on the other hand, are the ones that are not members of the QCA solution, but still show the outcomes. Such deviant cases occupy the upper left-hand side of the spectrum. They have no membership score in the sufficient term, a high membership score in the outcome, and are consistent with the statement of sufficiency. Deviant cases for coverage do not exactly contradict the statement of

sufficiency, as the sufficient cause is not always essential for the outcome to occur. However, analyzing such deviant cases would still be fruitful, as the analysis could lead to the discovery of an overlooked causal recipe or condition which could have been the cause. There is a list of the different COA techniques in the cases of sufficiency in the subsequent sections. As in the case of fsQCA Necessity, the remaining cases are irrelevant for COA.

OLR in fsQCA cases of Sufficiency

In cases of sufficiency, one can perform an OLR (Most similar system - Method of Agreement) by comparing the deviant cases of consistency with other similar deviant cases. Here, one first process traces the cases which are members of the QCA solution, but are not members of the outcome, and then compares the cases with each other. As in the instance of fsQCA Necessity cases, the technique would involve comparing deviant cases of consistency (degree) which are similar to typical cases, with other deviant cases of consistency (degree). An alternative (and even more promising) possibility is to compare deviant cases of consistency (kind) with a high membership score on the sufficient term and a low membership score on the outcome, with other deviant cases of consistency (kind). An analysis of the latter could help identify the reasons why the supposedly sufficient causal condition/ recipe does not produce the expected outcome.

Yet, another possibility would be to compare the different deviant cases for coverage with each other. Here, there is a comparison of the cases which are not members of the QCA solution, but which still display the outcome. The comparison is useful when the focus of the analysis is on identifying the reasons behind the high membership in the outcome despite having no membership in the sufficient term. Something to be cautious about is that the deviance can be due to any reason other than the sufficient condition, which could make the comparison difficult. The initial process-tracing step would be able to curb the difficulty.

OTR in fsQCA cases of Sufficiency

OTR in cases of Sufficiency can be in two forms. The first comparison could be between typical cases and deviant cases for coverage, where the outcome is present in both

instances, with the sufficient term absent in the deviant cases (Most different system - Method of Agreement). The second comparison would be between typical cases and deviant cases of consistency, where both cases would be members of the QCA solution of sufficiency, with the deviant cases not members of the outcome, or being a member but also showing inconsistency with the set theoretic relationship of sufficiency (Most similar system - Method of Difference).

OTR of fsQCA Sufficiency cases thus provide an extensive range of possible comparisons. The “most likely” cases are compared with the deviant cases of consistency (degree and kind) and the deviant cases for coverage, providing three possible ways of comparison. Likewise, the least likely cases are also compared with the three types of deviant cases, providing yet another three opportunities for theory enhancement.

4.7. Concrete examples of analysis of deviant cases using COA

So far, the present study has illustrated the methodological sophistication of the two COA methods in terms of their underscoring the key strength of QCA, with only the occasional hint at the theoretical gains associated with the proposed methodological innovation. To illustrate the theoretical gains more graphically, Table 4.3. includes some cases identified as deviant cases during the course of studies using a configurational approach. As can be seen from the examples, not all authors undertake a further analysis of identified deviant cases. In principle, there is no problem if the authors intend to focus only on the phenomenon which interested them initially. However, if they are keen on understanding the reasons behind the observed deviance, and eventually in expanding the theory building potential of their study beyond the scope of the initial study, COA would indeed be a blueprint for moving forward.

Article	Brief summary	Causal conditions	Outcome as in typical cases	Deviant cases	COA technique involved
<i>“The push and pull of ministerial resignations in Germany, 1969-2005”</i> by Fischer, Kaiser, and Rohlfing (2006), as mentioned in the paper by Rihoux and Ragin (2009).	The authors study the reasons leading to ministerial resignations. They argue that ministers are forced to resign when the political costs of a minister staying in office is higher than the benefits of keeping the status quo. They test this by doing a QCA study involving resignation decisions in Germany (1969 to 2005)	<i>“date of offense,” “criminal relevance,” “minister’s position,” “federal chancellor’s position,” “position of the minister’s party,” “position of the coalition partner,” “position of the parliamentary opposition,” “position of the media and public,” “political or extra political kind of resignation issue,” and “relationship of offense to office”</i> (Sufficient Conditions)	<i>“ministerial resignation”</i>	<i>“non-resignation”</i> (Two cases)	Outlier Literal Replication (Most similar system- Method of Agreement) By comparing the two similar deviant cases with each other, the authors identify a potential omitted variable <i>“the intensity of a crisis in terms of aggravating events within a period of time”</i> (Rihoux and Ragin, 2009)
<i>“Achieving accuracy, generalization-to-contexts, and complexity in theories of business-to-business decision processes”</i> by Woodside and Baxter (2013)	The authors elaborate on a business-to-business process study of marketing and purchasing chemicals conducted by Woodside and Wilson (2000). The study looks upon the multiple contingency paths of the ways in which the buyers and marketers think, and act before making a buying decision.	<i>“customer has large annual purchase requirements,” “customer willingness to single source requirements,” and “customer objectiveness/aggressiveness with respect to price.”</i> (Sufficient conditions)	<i>“customer share of business awarded to firm X”</i>	<i>“Case 11”</i> , has a high membership in the causal recipe and yet a low membership in the outcome (One case)	Outlier Theoretical Replication (Most similar system- Method of Difference) The authors conduct further analysis of this outlier to find out that this customer (Case 11) filed complaints with his manufacturer-sales manager and tried to renegotiate prices during the annual contract several times. Clearly, Case 11 is an unsatisfied

<p><i>“Qualitative work and the testing and development of theory: Lessons from a study combining cross-case and within-case analysis via Ragin’s QCA” by Cooper and Glaesser (2012).</i></p>	<p>The authors use the SOEP (representative panel study) from Germany to ascertain the types of school young people attend at the age of 17 (Gymnasium: offers the qualification ‘Abitur’ for university entry, Realschule, and Hauptschule).</p>	<p><i>“having at least one parent with the highest school qualification ‘Abitur’,” “having at least one parent in the service class,” “whether the student is male or female,” and “whether the student had a recommendation for Gymnasium.”</i> (The recommendation for Gymnasium is a quasi-necessary condition for the outcome.)</p>	<p><i>“the student was at gymnasium by age 17”</i></p>	<p><i>“the student was at gymnasium by age 17”</i> despite having no recommendation to attend the gymnasium (Six cases)</p>	<p>customer. Probing for the reasons behind his dissatisfaction may require further follow-up analysis.</p> <p>Outlier Theoretical Replication (Most different system-Method of Agreement) Further analysis of two of these deviant cases in comparison with the typical cases pointed out that when it comes to the outcome; some factors like a good knowledge of the system, family environment, level of cognitive ability etc. might be able to substitute for the quasi-necessary condition of gymnasium recommendation.</p>
---	---	---	--	---	---

Table 4.3. Examples of COA from the prior literature

4.8. Discussion

COA and process tracing

Process tracing a post-QCA deviant case is valuable in serving as a test for a theory in hand, finding out new properties of the identified deviant case, and is also the first step in every COA study. However, with a single case, generating a theory with much depth would be difficult and its theoretical grounding would not be credible enough (Eisenhardt, 1989). Process tracing a deviant case can establish the causal mechanism existing between causal conditions and outcomes, thus advancing the internal validity of a study. COA, on the other hand, replicates the findings on other similar (OLR) and dissimilar (OTR) cases to see how the causal conditions exist in a range of cases, thus ensuring the present study's external validity. Together, the two methods further the rigor of the study.

COA and contributions to QCA

COA, like QCA, has a fundamental anchor in Mill's Method of Difference and Method of Agreement. For the same reason, COA techniques are quite complementary to QCA. At the same time, the causality and complexity of social causation and the heterogeneity of cases lead to the impossibility of a QCA model covering all cases or providing precise, focused explanations to all cases (Glaesser & Cooper, 2011). COA focuses on analyzing cases which a particular QCA model cannot cover, thus contributing to theory enhancement. QCA provides configurations which comprise skeletal types of cases (Glaesser & Cooper, 2011), as they exist in the social world, thus eliminating the need for further theoretical sampling. Thus, COA and QCA are mutually supplementary. An OLR design compares similar deviant cases with one another, thus helping to figure out the common circumstance which leads to the phenomenon of interest. OTR design (Most similar system- Method of Difference) looks into cases which notify the omitted conditions not incorporated into the initial QCA model, and hence cause the expected outcome not to occur despite the presence of other conditions. In contrast, the Most different system- Method of Agreement OTR considers cases which point to alternative conditions which lead the deviant cases to show the same outcome as typical cases despite the absence of relevant conditions.

COA for achieving generality, accuracy, and coverage

According to Thorngate's (1976) "postulate of commensurate complexity," a researcher always has to make compromises between the three meta-theoretical virtues of generality, accuracy, and coverage in research. Conversely, Woodside (2010) puts forth the "postulate of disproportionate achievement," whereby tradeoffs between the three objectives are not mandatory. Through property space analysis, the possibility arises of arriving at theoretically possible combinations of conditions for achieving all three objectives (Woodside, 2010). Fuzzy set social sciences (like fsQCA) have high generality and accuracy, but low coverage. Multiple case studies (as in COA) meanwhile have high accuracy and coverage, but low generality. A combination of QCA and COA would make achieving all three objectives possible.

4.9. Limitations

One of the main observations of the present study is that there is not always a comparative analysis of post-QCA inconsistent cases in management, thus presenting a difficulty in demonstrating the various replication designs that the study is suggesting. Even some of the examples presented are not demonstrative enough as not all of them actually carry out an outlier analysis per se. However, the current paper includes the best instances from the study sample and the literature, which could illustrate the idea.

COA as a technique is not without its flaws either. OLR, where one compares outlying cases with each other, may be prone to an external validity problem as there is a between-case comparison of similar cases only. OTR does not face the problem. However, conducting an OTR could be taxing in terms of the time, effort, and resources involved. A good way to avoid any squandering of resources would be to first make sure that the inconsistent cases under study are not due to any error, but due to some phenomenon which may be "interesting" to the researcher (Aguinis et al., 2013). Prior knowledge about the cases and a sound theoretical base could help in addressing the problems to some extent. As QCA is a method already requiring familiarity with the relevant theories, the relevant literature, and sensitivity to the cases in hand, obtaining in-depth knowledge on the cases for COA would be feasible.

4.10. Conclusion

The COA technique is perfectly complementary and supplementary at the same time to the QCA approach. QCA studies normally require the number of cases to be more than eight. COA, in contrast, takes into account cases which are very small-N. In a case where incorporating inconsistent cases into a QCA study is impossible due to their inadequate numbers, COA is a very good alternative. Furthermore, the technique perfectly fits in with QCA, as both methods implicate “thick data,” making the individual QCA “units of observation” very suitable for undergoing COA.

Today a lot of work is taking place out with regard to analyzing outliers identified during qualitative and quantitative studies. Deviant cases are gaining more and more importance in management research, as is the configurational approach. Yet, the current paper is the first one to encourage a comparative analysis of inconsistent cases in QCA using a qualitative method. The present study is also the first one to advocate a comparison of deviant cases with other deviant and typical cases in QCA. By combining QCA and COA together, the current study contributes to management research methodology.

PART TWO

RIGOR AND WRITING

Chapter V*

What passes as a transparent field study in management? An analysis of qualitative and quantitative research in top management journals

Dr. Matthias Weiss

Lakshmi Balachandran Nair

Prof. Michael Gibbert

Prof. Martin Hoegl

Journals

Weiss, M., Nair, L.B., Gibbert, M., Hoegl, M., & Koeplin, H. Journal of Management (Submitted)

Weiss, M., Nair, L.B., Gibbert, M., & Koeplin, H. (2015). Academy of Management Proceedings (Published)

Conferences

Nair, L.B., Weiss, M., Gibbert, M., Hoegl, M., & Koeplin, H. (2016, March). Western Academy of Management Conference, Portland, USA.

Weiss, M., Nair, L.B., Gibbert, M., & Koeplin, H. (2015, August). 75th Academy of Management Annual Meeting, Vancouver, Canada.

Nair, L.B., Weiss, M., Gibbert, M., & Hoegl, M. (2015, March). Western Academy of Management Conference, Hawaii, USA.

5.1. Abstract

There is agreement in the management community that research methodology needs to be transparently reported. Unlike in other disciplines, however, it is far from clear what the label ‘transparent research procedures’ constitutes in management field studies. This may entail adverse effects during write-up, revision, and even after publication. To help rectify this, we review 365 field studies across seven major management journals (1997-2006) in order to develop a transparency index, and link it to article impact (i.e. citation count). Overall, transparency boosts impact across methods; quantitative studies tend to be more transparent than qualitative studies, although key transparency indicators were underutilized in both methods.

Key words: *Transparency, citations, qualitative, quantitative, reporting standards*

5.2. Introduction

Transparency in research methodology matters. In order to appreciate a study's results, and to successfully integrate them into the body of scholarly knowledge, they need to be transparently relayed (Aguinis et al., 2010; Aytug, Rothstein, Zhou, & Kern, 2012; Bergh, Perry, & Hanke, 2006; Judge, Cable, Colbert, & Rynes, 2007; McGrath, Martin, & Kulka, 1982; Miguel et al., 2014). Moreover, transparency constitutes a prerequisite for rigor: without clarity about characteristics of the sample, indicating the presence/absence of missing data (including outliers), and a clear rationale for control variables, key rigor aspects such as validity and replicability, are difficult to assess (e.g., Cook & Campbell, 1979; Yin, 1994). Most recently, in the wake of several high-profile scandals in management and in the social sciences more generally, transparency has even been proposed as a safeguard against the thin line dividing best practices and scientific misconduct, called 'questionable research practices'. While there is no clarity in the research community just where this line is, transparency acts as an antidote to questionable research practices and enables replication (Banks et al., 2016; Miguel et al., 2014; Nosek et al., 2015).

Fields other than management (in particular psychology and medicine), have made much progress when it comes to transparency, leading to greater cohesion in the conventions relating to reporting research methods and data analyses (Fuchs, Jenny, & Fiedler, 2012; Miguel et al., 2014; Wilkinson, 1999), with widely-used manuals and detailed 'checklists' (e.g., APA, 2008; Hancock & Mueller, 2010; Kilkeny, Browne, Cuthill, Emerson, & Altman, 2010; Schulz, Altman, & Moher, 2010). In management, however, while the rigor of research procedures has been studied for decades, its logical prerequisite, transparency, is much less well understood. This has dire consequences during write-up, while a paper is under revision as well as after acceptance and even after publication. During the review process, for instance, many times lack of transparency causes additional revision rounds. Even after successful completion of the review process, lacking transparency in field research can lead to severe consequences, such as retraction of published articles (Atwater, Mumford, Schriesheim, & Yammarino, 2014). A systematic appraisal to what extent published authors adhere to such explicit

transparency standards (or even go beyond them) is so far lacking in management studies, however. Part of the reason for this is that management studies as a field has not yet developed a common understanding of what ‘transparent’ research procedures entail, i.e. has not yet developed a standard transparency index for gauging transparency. In addition, it is not clear to what extent transparency ‘pays off’ in terms of its impact on the field. In particular, it would appear that articles with transparent methods would be more readily appreciated and consequently gain more impact in the academic community.

The primary aim of this study is therefore to theoretically develop and empirically identify a transparency index, following the approach of Aytug et al. (2012) in the domain of meta-analyses, capturing which methodological choices and data properties are reported in qualitative and quantitative field research, as well as link transparency to its recognition and appreciation in the academic community (in terms of the impact of the published article). To this end, we reviewed literature on empirical research methods and interviewed ten editors from top tier journals in the field of management research in order to identify those aspects which are most likely to be expected for reporting in management field studies. Building on this transparency index, we investigated papers published in seven leading management journals over an extended period of time (1997-2006). Finally, we performed an exploratory analysis to identify antecedents of transparency, that is, identifying paper and author-related aspects that drive or inhibit transparency. In these respects, we consider both qualitative and quantitative approaches of field studies, thereby not only providing a more complete picture of field research reporting standards, but also enabling a comparison of the different rationales behind reporting standards underlying these two fundamental approaches.

The study reported here sets out to contribute to the literature in three ways. First, we provide a long-missing assessment of what can be considered the *de facto* reporting standards (i.e., as observable in articles published in particularly impactful and rigorous journals – rather than in normative textbooks) of quantitative and qualitative field studies

and which features of the paper and its authors are related to transparency. We focus on top-tier management journals, specifically, since they represent the showcase of management research and thus the landmark for those who conduct impactful field research in management. Second, while the descriptive results of our study show what reviewers and editors valued in terms of reporting, and thus establish *de facto* reporting standards, the regression analyses on article impact investigate if variances in reporting transparency across individual articles are related to how scholars in the field (i.e., the journals' readership) reference articles over time. Several antecedents of article impact were examined, including methodological rigor, clarity, coherence (Bergh, Perry, & Hanke, 2006; Judge, Cable, Colbert, & Rynes, 2007), article structure, quality (Bergh et al., 2006; Flickinger, Tuschke, Gruber-Muecke, & Fiedler, 2013; Haslam et al., 2008; Molina-Azorin, 2012; Tams & Grover, 2010), as well as the impact and citation count of the journal or type of issue in which it has been published (Conlon, Morgeson, McNamara, Wiseman, & Skilton, 2006; Judge et al., 2007). However, this has not been done for transparency so far. Therefore, in this paper, we specifically focus on the variance in article citation explained through a paper's transparency in reporting the methodology and results of field studies. Third, the comparison between practices and norms for establishing transparency in qualitative and quantitative methods of field research sheds light on different strategies and logics applied in these two methodological fields.

5.3. Operationalizing transparency

Important forays into greater transparency were made in scientific disciplines other than management. With the rise of experimental methods and a greater reliance of meta-analyses in the field of psychology, scholars have started to pay increasing attention towards "evidence-based" decision-making and compliance with detailed reporting standards. Particularly, sufficient transparency in reporting methods and results is essential for the use of research synthesis and experiential study designs (Aytug et al., 2012). Often practices are borrowed from medical trials to endorse guidelines for studies in the field of psychology (e.g., CONSORT, TREND, JARS) (Altman et al., 2001). These practices require researchers to include information in the general areas of

measurements, subsets of experimental conditions, manipulations, data exclusions, as well as how the final sample size has been achieved.

To make a first step into this direction in our discipline, management, we offer a straightforward definition of transparency as the degree to which authors share relevant information with stakeholders (Pirson & Malhotra, 2011). In our case, stakeholders are those involved in the decision to publish a paper (i.e., reviewers, editors), as well as the readers of the published article (Aguinis et al., 2010). These stakeholders usually depend on the author to report the relevant information regarding the procedures of empirical field and laboratory studies. Naturally, only what gets reported can be assessed, so transparency constitutes the fundamental prerequisite for assessing the quality (i.e., rigor) of a paper. Consider also the power distance between the two groups of stakeholders: those involved in the decision to publish a manuscript may at least exert some influence on authors to provide (more) relevant information. The second group of stakeholders, the actual consumers of the published article, lack this option (Aguinis et al., 2010). Ultimately, though, both groups of stakeholders depend upon the transparency of a given article to appropriately interpret, appreciate, and (perhaps) approve its results, and integrate them into the existing body of scholarly research via citations.

Note that transparency as defined here refers to the breadth of reporting, as opposed to the depth of reporting, which is part of rigor. This is best illustrated by contrasting whether certain methodological procedures are reported in a paper (the realm of transparency), with what these procedures look like (the realm of rigor). An adequate level of transparency in reporting methods and data properties constitutes a necessary prerequisite for allowing readers to evaluate the rigor of a specific field study, and thus the validity of its findings (Cook & Campbell, 1979; Gephart, 2004). To illustrate, consider a paper presenting the results of a quantitative field study: to operationalize transparency, we need to know *whether* the authors of this paper report reliability coefficients for the variables used. Assessing rigor would then entail examining *what* these coefficients point out (the level of variable reliability), which is beyond the scope of a study on transparency.

In an attempt to construct an operational transparency index, in the following section we will outline important features to be reported in a paper in order to provide an adequate level of transparency, derived from existing guidelines or reporting standards (e.g., APA, 2008; Gibbert, Ruigrok, & Wicki, 2008; Hinkin, 1995; Yin, 1994). We will do so separately for quantitative and qualitative field studies, as these two methodological approaches require different features to be reported in a paper.

Transparency of quantitative field studies

Regarding quantitative field studies, features to be reported as mentioned by manuals and guidelines on reporting issues can be grouped in three major categories (APA, 2008; Hancock & Mueller, 2010; Sterba, Christ, Prinstein, & Nock, 2011): the setting as well as practices of data collection, properties of the collected data, as well as data analysis and its results. For many features within these categories we find certain quantifiable indicators, as well as predefined thresholds or rules of thumb. These thresholds are usually derived from statistical theory, or experience (or both). Unfortunately, their use and specific values are not always consistent or well-founded even within the same field of research (Lance & Vandenberg, 2009; LeBreton & Senter, 2008), and their very origins may be obscure (Lance, Butts, & Michels, 2006). Regardless of such problems, the idea behind most of the features to be reported in quantitative field studies is clear. That is, providing quantified information about the data collection, the dataset itself, and the data analysis, which can be compared to any predefined critical value in order to provide 'hard' evidence for the results' validity.

Reporting certain key features of the data collection process in quantitative field studies first of all serves the purpose of providing readers with information about the setting of the study (Wilkinson, 1999). This includes details on the business context in which the field study was conducted, such as the specific industry, geographical area, and time in which data were collected. Moreover, detailed information on the organizations in which data were collected as well as on the specific sample entities and respondents targeted in data collection, is recommended to be reported here. Ideally, this information on the setting of a field study is accompanied by a rationale explaining the

decisions regarding the study's setting based on theoretical considerations. Other features in this category correspond to the actual data collection process, that is, the report of details concerning circumstances (e.g., the response rate) and the method used to collect the data (Church, 2001; Weigold, Weigold, & Russell, 2013).

Reporting features related to data properties allows evaluating the realized sample (Wilkinson, 1999). Although the setting of the study should have been reported before, additional information should be provided to show the attributes of the cases included in the realized sample like, for example, the size and type of firms or teams, and the demographics of the respondents. This information is necessary to assess the representativeness of the sample and to allow a better interpretation of the results given the attributes of the cases actually included in the study. Beyond this, features that provide information about the general properties of the data are expected to be reported here (Wilkinson, 1999), such as correlation tables showing interrelations between variables included in the study, descriptive data, information on the presence of outliers and missing data (Aguinis, Gottfredson, & Joo, 2013; Newman, 2014), as well as properties of the instruments used to measure the variables that allow the assessment of reliability and construct validity of the used measures (Hinkin, 1995).

The third category of features that are recommended for reporting in papers concerns information on data analysis and its results (APA, 2008; Sterba et al., 2011). This category includes all information on the statistical analyses used to test hypotheses and that constitutes the foundation of the conclusions drawn in the paper (Wilkinson, 1999). This information is important for the assessment of the given interpretation of study results. Thus, it is particularly important to provide confidence that the analyses have been executed rigorously so that appropriate conclusions can be drawn from these results (Cook & Campbell, 1979). This includes features such as reporting information on the use of control variables, significance levels, or error terms (Aguinis et al., 2010; APA, 2008).

Transparency of qualitative field studies

Traditionally, qualitative research procedures have lagged behind their quantitative counterparts regarding the degree of codification in reporting conventions (Denzin & Lincoln, 2005). However, recent literature (Gibbert & Ruigrok, 2010; Gibbert et al., 2008) has recognized the significant role of rigor and has surveyed the actual use of different methods for enhancing the rigor of qualitative research as published in major management journals. In qualitative research, the main features of a transparent article are full revelation and unambiguousness regarding the procedures at every stage of the research process (Hiles, 2008). In quantitative research this means authors have to provide easily codified information about the collection, analysis and synthesis of data. By contrast, in qualitative research, reporting conventions being much less codified, transparency is about guiding the reader through various stages of the argument by ‘walking the talk’ (Gibbert & Ruigrok, 2010). To achieve this, measures such as careful documentation and clarification of the research procedures should be taken. A report that specifies how the entire case study has been conducted (Dubois & Gibbert, 2010) reduces vagueness (easier to determine the strengths & weaknesses) and complexity, walks the reader through the various stages of the argument, enhances replicability, makes the logic, reasoning and causalities evident and may demonstrate the overall rigor of a study. As Glass pointed out, “a full and true report is the hallmark of the scientist, a report as accurate and faithful as he can make it in every detail. The process of verification depends upon the ability of another scientist who wishes to repeat a procedure and to confirm an observation” (Glass, 1965: 83).

Bearing in mind the flexibility of qualitative research and that various formulations and types of qualitative research have different stages, steps and decision points (Gephart, 2004; Van Maanen, 1998), our operationalization of transparency in qualitative studies builds on generalizable stages in the qualitative research process (Eisenhardt, 1989; Yin, 1994). This enables that the transparency features are equally applicable and thus adequate for looking into the breadth of reporting in a variety of qualitative articles. This ideal type four-stage qualitative research process is depicted in

Figure 5.1. and consists of stages corresponding to data selection, data collection, data analysis/management, and the presentation of results.

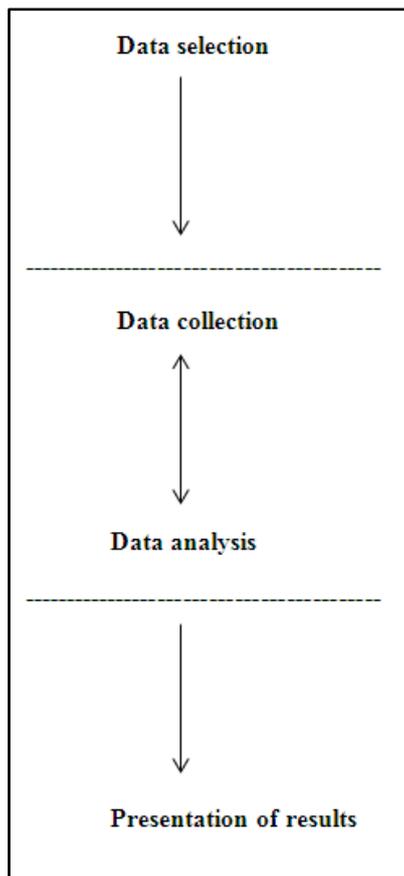


Figure 5.1. Stages of the qualitative research process

It reflects the rationale underlying transparency in qualitative field research to provide a clear chain of evidence which would allow any stakeholder to reconstruct how the researcher went from the initial research questions to the final conclusions (Yin, 1994: 102). Of course, this process might take on quite diverse forms in research practice; some stages might be carried out substantially more or less extensively and some stages might take place simultaneously. Establishing a clear chain of evidence also indicates a satisfactory correspondence between an observation of reality (Denzin &

Lincoln, 2005) and the operational procedure designed to measure it (Lee, Mitchell, & Sablinski, 1999; Yin, 1994). Typically, the qualitative research process is not linear: often stages of data collection alternate with data analysis, with authors travelling back and forth between data and emerging theoretical predictions. Similarly, the specific approaches or processes that individual qualitative studies take may differ from each other. Nonetheless, even though the nomenclature and position of the stages may differ, the stages involved remain more or less the same throughout all these alternatives (Berg & Lune, 2004; Miles & Huberman, 1984).

In qualitative studies, data selection is a deliberate process (Morse, 2004) which is conceptually driven either by the theoretical framework or by a developing theory which is derived inductively from the data as the research progresses – in both instances, cases are sampled theoretically, not randomly (Curtis, Gesler, Smith, & Washburn, 2000). Accordingly, a reflexive and explicit rationale for data selection (Curtis et al., 2000) would be indispensable in providing the audience with relevant information about the rationale behind choosing the data in the first place (Flick, 2014), the number of data points, and how well they fit in (or do not fit in) with the theory in hand or other data points. Moreover, researchers may provide details on the case study context, in order to allow the reader to appreciate the researchers' sampling choices (Cook & Campbell, 1979: 83).

As qualitative research is highly heterogeneous, data collection processes require 'methodological versatility' (Yin, 1994) and hinge heavily on the epistemological framework chosen. Hence, reporting features related to qualitative data collection in general should take account of data sources, methods, or investigators involved in the study (Denzin, 1970), as well as data collection circumstances (Yin, 1994). Basically, the authors should mention whether they have or have not looked at the same phenomenon from different angles by using different data collection strategies and data sources, as a means for ruling out alternative explanations (Denzin & Lincoln, 2005; Pettigrew, 1990; Stake, 1995). As collecting data in naturalistic settings requires compromises and adjustments to procedures (Hiles, 2008), maintaining a chain of

evidence allows to follow the derivation of the collected evidence, that is, to trace the evidentiary process backward (Yin, 1994).

Providing transparency concerning features of data analysis aims at offering precise and interpretive information about the corresponding (data interpretation or data management) stage of the qualitative research process. These features include providing a clear definition of the research question (whether it is inductive, deductive, neither or both). A sufficiently transparent paper should also have an explicit mentioning of the unit of analysis, controls, and dependent and independent variables. Additionally, discussion of the coding principles and explanation of data analysis would help in ensuring a clear chain of evidence. Principally, as Moravcsik (2014) suggests, authors should provide readers access to this information in order to support specific descriptive, interpretive, or causal claims.

The final category of features recommended for ensuring transparency concerns the presentation of results. Essentially, during this stage a detailed description of features should be provided, to help others in understanding the generalizability of the research findings (Fossey, Harvey, McDermott, & Davidson, 2002). For instance, a relevant feature for presenting results is confirming whether the focal organization of a study is mentioned by its own name. Providing raw data as illustrations or examples is also considered important for providing transparency when reporting results of qualitative studies. As a final point, whether the rigor parameters are explicitly mentioned is part of this category. For the same reason, in this study we use the Natural Science model (Eisenhardt & Graebner, 2007; Piekkari, Welch, & Paavilainen, 2008), which groups together various research actions under four rigor criteria: construct validity, internal validity, external validity and reliability (Behling, 1980; Cook & Campbell, 1979), and has been accepted by researchers from both positivist and interpretivist traditions.

5.4. Methods

Sample

To examine our research questions and to assess the predictors of article transparency as well as the influence of article transparency on scientific impact, we needed to select a

representative set of management journals from which we could draw our sample of articles. In this study, we focused on a set of seven top tier management journals that publish articles reporting field studies. This selection is based on previous work rating the impact of management journals (Gomez-Mejia & Balkin, 1992; Podsakoff, Mackenzie, Bachrach, & Podsakoff, 2005; Tahai & Meyer, 1999), as well as on recent statistics of journal impact ratings from the SSCI Journal Citation Reports (ISI Web of Knowledge, 2013). The journals were *Academy of Management Journal*, *Administrative Science Quarterly*, *Journal of Applied Psychology*, *Journal of Management*, *Organizational Behavior and Human Decision Processes*, *Organization Science*, and *Strategic Management Journal*.

In a next step, we manually searched through these journals for qualitative and quantitative field studies. This search covered the period from 1997 to 2006, which, on the one hand, represents a period long enough to provide the opportunity to control for or examine any time effects regarding the analyzed relationships. On the other hand, this period allows for meaningful analyses of article citation counts, given that a comparably long time is needed after publication of an article in the social sciences for citations to accumulate (Walters, 2011).

Following related prior research (Judge et al., 2007) and to achieve sufficient power for our statistical analyses, our original aim was to extract 30 articles per journal for qualitative and quantitative field studies each. However, given the relatively small number of qualitative field studies, we decided to include the entire population of qualitative field studies published in these journals during the relevant time period which amounted to 146 articles, an average of 20.8 per journal. Given the much higher number of articles reporting quantitative field studies, we randomly extracted three to four articles per year per journal, although the number of articles slightly varied depending on how many quantitative field studies were published in each journal. In total, we arrived at 219 articles reporting quantitative field studies, an average of 31.3 per journal, making it 365 articles in total. This sample size provided sufficient statistical power for our design (Ferguson & Ketchen, 1999). For transparency of the research procedures

employed here, an even more detailed description of the sampled articles can be found in Table 5.1.

	Quantitative	Qualitative
AMJ	32	36
ASQ	27	31
JAP	34	3
JoM	35	2
OBHDP	22	0
OS	34	58
SMJ	35	16
1997	22	8
1998	21	12
1999	25	11
2000	22	20
2001	20	19
2002	21	21
2003	22	11
2004	21	15
2005	23	9
2006	22	20

Table 5.1. Sample description

Interview study

To complement our quantitative empirical analyses and findings we conducted an interview study with ten editors (associate editors, senior editors, and co-editors) from the selection of top tier management journals presented above. We took care to include experts of quantitative and qualitative methods in our sample of editors. six out of these ten interviews were conducted by one member of the author team on-site, while four interviews were conducted via telephone. The interviews typically lasted between 20 and 35 minutes and have been voice-recorded and transcribed.

In the interview study, we followed a semi-structured approach. First, we asked the editors for the aspects they considered most important to be reported in quantitative and qualitative field studies. Second, we asked them for the most important aspects to be reported with regard to each stage in the research process. We used this input for developing our selection of transparency features and the transparency index as explained below. Finally, we gave the associate editors the opportunity to express their general thoughts and opinions regarding the topic of reporting standards and transparency in the field of management. We used these statements for a better interpretation of our empirical findings. We have integrated quotes of the editors' statements in the discussion section.

Measures and data extraction

Transparency.

To identify a meaningful set of aspects that tend to be expected for reporting in quantitative or qualitative field studies in management, we first looked through manuals and textbooks on quantitative and qualitative research methodology (e.g., APA, 2008; Cook & Campbell, 1979; Denzin & Lincoln, 2005; Gibbert et al., 2008; Hancock & Mueller, 2010; Miles & Huberman, 1984; Nosek et al., 2015; Stake, 1995; Sterba et al., 2011; Yin, 1994) and created a list of features that are recommended for reporting from these sources. For a reality and relevance check, we then matched these features with those mentioned by the editors in our interview study and deleted those features that have not been mentioned as an important feature by at least one editor. The complete set of features relevant for transparent reporting of quantitative and qualitative field studies that resulted from this two-stage approach is shown in Table 5.2. (quantitative) and Table 5.3. (qualitative).

Category	Transparency codes	Relevant literature	Coding
Data collection	Description of access to data provided	(AERA, 2006; APA, 2008;	Y/N
	Rationale for data selection provided	Church, 2001; Gephart, 2004;	Y/N
	Description of data collection approach provided	Miller, Washburn, & Glick, 2013; Sterba et al., 2011; Stone & Shiffman, 2002; Weigold et al., 2013; Wilkinson, 1999)	Y/N
	Details on study context provided		Y/N
	Voluntariness of participation mentioned		Y/N
	Details on respondents provided		Y/N
	Response rate given		Y/N
	Formal definitions for all focal variables provided		Y/N
	All items used in the study shown / clearly referenced		Y/N
	Source / development of scales and items mentioned		Y/N
Data properties	Details on sample provided	(AERA, 2006; Aguinis et al., 2013; APA, 2008; Atinc, Simmering, & Kroll, 2011; Bliese, 2000; Hinkin, 1995; Newman, 2014; Wilkinson, 1999)	Y/N
	Presence/absence of missing data indicated		Y/N
	Rationale for all control variables provided		Y/N
	Information on interrater reliability / agreement given		Y/N
	Descriptive statistics provided		Y/N
	Correlations between study variables provided		Y/N
	Variable reliability indicated		Y/N
Presence/absence of outliers indicated		Y/N	
Data analysis and results	Unit of analysis explicitly indicated	(AERA, 2006; APA, 2008;	Y/N
	Standard errors or equivalent values given	Cook & Campbell, 1979; Wilkinson, 1999)	Y/N
	One vs. two-tailed significance testing indicated		Y/N
	Validity aspects explicitly mentioned		Y/N

Table 5.2. Codes for transparency of quantitative field studies

Category	Transparency codes	Relevant literature	Coding
Data selection	Rationale for data selection provided	(Cook & Campbell, 1979; Curtis et al., 2000; Flick, 2014)	Y/N
	Details of case study context provided		Y/N
Data collection	Review by peers indicated	(Cook & Campbell, 1979; Denzin, 1970; Denzin & Lincoln, 2005; Gibbert & Ruigrok, 2010; Gibbert et al., 2008; Pettigrew, 1990; Stake, 1995; Yin, 1994)	Y/N
	Review by key informants indicated		Y/N
	Circumstances of data collection indicated		Y/N
	Data sources; Triangulation of data sources, investigators, (or) methods specified		Y/N
	Multiple case studies indicated		Y/N
	Use of deviant cases mentioned		Y/N
Data analysis	Mentioned inductive/deductive nature of research question	(Cook & Campbell, 1979; Gibbert et al., 2008; Glaser & Strauss, 1967)	Y/N
	Explanation of data analysis provided		Y/N
	Coding principles indicated		Y/N
	Unit of analysis explicitly indicated		Y/N
	Discussion of control variables provided		Y/N
	Dependent and independent variables identified		Y/N
Presentation of results	‘Validity’ explicitly mentioned	(Eisenhardt & Graebner, 2007; Gibbert & Ruigrok, 2010; Gibbert et al., 2008; Locke & Golden-Biddle, 1997; Piekkari et al., 2008)	Y/N
	Construct validity explicitly referred to		Y/N
	Internal validity explicitly referred to		Y/N
	External validity explicitly referred to		Y/N
	Reliability explicitly referred to		Y/N
	Raw data provided as examples		Y/N
	Organization mentioned by own name		Y/N

Table 5.3. Codes for transparency of qualitative field studies

To obtain the data on transparency, we extracted information about the features listed in Table 5.1 and 5.2. In line with our approach and definition of transparency, our primary concern in this regard was *whether* each of these features that could affect the results and conclusions of the field studies was reported fully and transparently. In contrast, we did not code *what* was reported (e.g., the specific values or rationales).

Thus, all the codes relating to transparency represented dichotomous variables (yes/no), indicating whether the respective features has been reported or not. For example, the item “Was the presence/absence of outliers explicitly mentioned?” was assigned with a score of 1 if this information was reported, and 0 if the authors did not report it.

In addition to coding the single transparency items, we created an overall transparency index for each article in our sample. It is important to note that the individual transparency items are not universally applicable to all field studies; some are conditional on the studies’ specific research design. We therefore calculated the transparency index by computing the percentage of applicable features that were actually reported, given each study’s specific research design. For example, in quantitative studies using multiple respondents from the same unit of analysis, such as teams, it is relevant to report the level of agreement between these representatives of the same entity, while this does not apply for studies using individuals as the focal unit of analysis. Thus, we computed the proportion of applicable features by summing the scores for all applicable items and dividing them by the number of applicable items.

Article impact

Impact of articles was operationalized by article citation counts, which represents, the conventional and most frequently method used (Adam, 2002; Leung, 2007; Stremersch, Verniers, & Verhoef, 2007). We used the number of citations that accumulated for each article until December 2013. In order to achieve more robust findings, we utilized three different sources for obtaining data on citation counts: ISI Web of Knowledge, Scopus, and Google Scholar. Further, we specified each of these citation counts as absolute number of citations and citations per year (Molina-Azorin, 2012), to arrive at six impact variables as outcome variables for our analyses. In order to minimize complexity and size of our results tables, we generally refer to ISI citation counts (total and per year), and only mention results for the other two sources in case there are noteworthy differences. All citation data had been collected on the same day to avoid any potential distortions resulting from the steady growth of citation counts.

Antecedents of transparency

To look for potential antecedents of article transparency, we examined core author and article attributes as specified in previous studies on related topics (e.g., Bergh et al., 2006; Jones, Wuchty, & Uzzi, 2008; Judge et al., 2007). Specifically, we included the following author attributes that might have an influence on how transparent an article is: Number of authors, gender of first author (coded 0 for male and 1 for female), gender ratio within the author team, and geographical location of first author's affiliation (dummy coded for continents: North America, Europe, Asia, Oceania). With regard to the article attributes that might influence its transparency we incorporated the following variables in our models: Article age (number of years since publication), article length (number of pages), publication in special issue versus publication in a regular issue (coded 0 for regular issue and 1 for special issue), article type: research note versus regular article (coded 0 for regular article and 1 for research note), and article position in the issue (coded 1 for first article in the issue and 0 for other positions).

Control variables

We controlled for the journal in which articles have been published, applying six dummy variables representing the journals in our sample (with *Administrative Science Quarterly* as reference group). Moreover, in our analyses regarding the relationship between article transparency and impact, we included the variables used to examine antecedents of transparency in all equations, since these attributes of authors and articles have been shown to be related to article impact in prior studies (Bergh et al., 2006; Conlon et al., 2006; Haslam et al., 2008; Mingers & Xu, 2010; Stremersch et al., 2007).

Coding

The articles in our sample were content analyzed and coded by multiple coders. For articles reporting the results of qualitative and quantitative field studies, respectively, four trained coders content analyzed the articles, two focusing on qualitative papers, the other two on quantitative papers. First, a standardized coding scheme was developed. To facilitate intercoder agreement we ensured that all coders agreed on the coding of each item and discussed and clarified eventual disagreements. To further ensure

standardization and reliability of the coding approach, each pair of coders checked and calibrated the coding after each coder coded 10 articles. Thereafter, the two pairs of coders completed coding the whole sample of articles representing their focused methodology. To estimate reliability of the coding process, we checked the level of agreement between coders in each pair. Initial agreements were high, each over 80 percent. Resulting disagreements were discussed among coders and resolved until 100 percent agreement existed.

5.5. Results

De facto reporting standards

Table 5.4. shows the descriptive statistics for the mean frequency of each of the coded items in our sample. Identifying which features have been reported regularly versus which features are reported only rarely in field study articles illustrate the actual reporting standards in top management journals. The least frequently reported feature in papers based on quantitative studies is the presence/absence of outliers (4.1%). In papers on qualitative studies, features that have been rarely reported are direct references to construct (7.5%) and internal validity (8.9%), discussion of control variables (6.2%), and whether deviant cases have been used for theory building (4.1%). In contrast, a number of our codes have been mentioned in most papers (by more than 80%) and thus indeed seem to reflect standard features for reporting. Specifically, these are for papers using a quantitative methodology: description of data collection approach, details on the study context, the study's response rate, the provision of descriptive statistics and correlations, and variable reliability indicators. For papers based on qualitative methods, these are: rationale for data selection, indicating review of peers, provision of raw data, data collection circumstances, explanation of data analysis, and details on the case study context.

Item reported	Yes	No
<i>Quantitative</i>		
Description of access to data provided	50.7	49.3
Rationale for data selection provided	36.1	63.9
Description of data collection approach	82.2	17.8
Details on study context provided	81.3	18.7
Voluntariness of participation mentioned	21.9	78.1
Details on respondents provided	29.0	71.0
Details on sample provided	72.1	27.9
Response rate given	85.4	14.6
Formal definitions for all focal variables provided	56.6	43.4
All items used in the study shown / clearly referenced	63.3	36.7
Source / development of scales and items mentioned	66.8	33.2
Rationale for all control variables provided	58.2	41.8
Presence/absence of missing data indicated	30.1	69.9
Information on interrater reliability / agreement given	76.5	23.5
Descriptive statistics provided	92.7	7.3
Correlations between study variables provided	94.1	5.9
Variable reliability indicator given	98.1	1.9
Presence/absence of outliers indicated	4.1	95.9
Unit of analysis explicitly indicated	25.6	74.4
Standard errors or equivalent values given	50.5	49.5
One vs. two-tailed significance testing indicated	26.9	73.1
Validity aspects explicitly mentioned	74.4	25.6
<i>Qualitative</i>		
Validity aspects explicitly mentioned	28.1	71.9
Construct validity explicitly referred to	7.5	92.5
Internal validity explicitly referred to	8.9	91.1
External validity explicitly referred to	36.3	63.7
Reliability explicitly referred to	30.8	69.2
Rationale for data selection provided	86.3	13.7
Review of peers indicated	84.2	15.8
Review of key informants indicated	27.4	72.6
Raw data provided	86.3	13.7
Circumstances of data collection indicated	93.2	6.8
Explanation of data analysis provided	83.6	16.4
Coding principles indicated	72.6	27.4
Unit of analysis explicitly indicated	60.3	39.7
Discussion of control variables provided	6.2	93.8
Dependent and independent variables identified	19.2	80.8
Details on case study context provided	91.1	8.9
Organization mentioned by own name	45.2	54.8
Mentioned inductive/deductive nature of research question	40.4	59.6
Use of deviant case for theory building mentioned	4.1	95.9
Multiple case studies indicated	49.3	50.7

Table 5.4. Reporting frequency of individual codes (percentages)

We calculated the mean transparency index of the papers in our sample, which is 53.6% (s.d. = 12.5), with 57.4% for quantitative papers (s.d. = 10.6) and 48.1% for qualitative papers (s.d. = 13.1). Thus, we note that papers based on quantitative methods

tend to be more transparent than papers using a qualitative methodology (the difference is significant, $p < .01$).

Transparency and article impact

Finally, we examined the effect of article transparency on impact. As mentioned above, the results of these analyses refer to citation counts in ISI Web of Knowledge. We specified two dependent variables, an article citation count per year and an article's total citation count (descriptive statistics and intercorrelations of variables included in the regression models are provided in Table 5.5.). Results of analyzing article citation count per year are presented in Table 5.6., and those examining total citation counts in Table 5.7. As some control variables showed high correlations among each other, we checked the variance inflation factors (VIF) and tolerance statistics for the coefficients in the models. VIFs are smaller than 5 and tolerances are above .20 for all coefficients, which suggest no substantial problems with multicollinearity.

The results for both specifications of article impact showed significant linear positive effects of transparency on article impact in the merged data set. Moreover, we also found a significant curvilinear effect of transparency on article impact that shows a U-shaped relationship between these variables. These effects appear to be driven by the qualitative articles in the sample, since the separate analysis for qualitative and quantitative articles revealed bigger coefficients for the qualitative subsample. Moreover, we screened for interaction effects with the third variables that might systematically influence the transparency-impact relationship. Testing potential interactions with the included control variables revealed one significant interaction effect for article age that materialized in similar strength for both qualitative and quantitative articles. Apart from minor differences in the magnitude of the revealed effects, analyses based on citation counts in SCOPUS or Google Scholar yielded similar results.

To further explore the nature of the detected interaction effect, we probed the simple slopes, revealing a significant positive relationship between article transparency and article impact for older articles ($p < .01$), while there is no significant relationship for articles of younger age ($p > .10$). The plots of these simple slopes are shown for the

merged data set and citations per year as outcome variable in Figure 5.2, depicting the described pattern. We also probed the simple slopes for the separate subsamples and total citations as outcome variable. As these yielded qualitatively identical results, we ended up not showing them in the paper.

To test the robustness of our findings to alternative specifications of the transparency indices for qualitative and quantitative papers, we repeated our analyses on the transparency-impact relationship several times, each time omitting two randomly chosen items included in the transparency index. The results of these analyses were consistent with those presented in Table 5.5. and 5.6., yielding no differences regarding the significance and direction of relationships, which indicates that the results were unlikely subject to the distribution of single items within the transparency indices.

Variable	Mean	s.d.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1) Transparency	0.53	0.12	-	-.01	-.01	.11	.08	-.15	-.02	.01	.16*	-.09	.20*	-.01	-.25**	.06
2) Citations/year (ISI)	11.07	11.42	.09	-	.96**	.03	-.05	-.08	-.05	-.20*	-.03	.01	.23**	-.05	-.03	-.07
3) Total citations (ISI)	127.87	149.68	.06	.97**	-	.07	-.06	-.08	-.05	-.20*	.01	.23**	.26**	-.07	-.01	-.04
4) Gender first author	1.38	0.49	-.14*	-.02	-.02	-	-.15	.03	-.07	.03	.85**	.06	.08	.16*	-.07	-.11
5) First author: Europe	0.16	0.37	.04	-.04	-.06	-.14*	-	-.12	-.04	-.13	-.13	-.09	-.14	.00	-.03	.09
6) First author: Asia	0.06	0.24	-.07	-.03	-.03	-.08	-.11	-	-.02	-.03	.02	.06	-.03	-.03	-.07	-.03
7) First author: Oceania	0.03	0.16	.03	-.09	-.08	.10	-.08	-.06	-	.14	-.03	-.04	-.05	-.04	-.03	-.01
8) Number of authors	2.25	1.10	-.09	-.09	-.08	-.07	-.06	.04	-.10	-	.02	-.08	-.14	.03	.24**	.09
9) Female ratio	0.36	0.39	-.14*	-.03	-.03	.80**	-.19**	-.00	.05	-.03	-	.09	.05	.14	-.11	-.12
10) Article age	11.39	2.82	-.19**	.09	.26**	-.04	-.07	-.05	-.05	.06	-.03	-	.24**	-.15	-.02	.12
11) Article length	20.67	7.79	.03	.17*	.18**	-.04	-.08	-.03	.05	-.12	-.01	.13	-	.04	-.06	-.00
12) Article position	0.15	0.35	-.02	-.01	.01	-.02	.00	.05	-.01	-.03	-.02	.07	.30**	-	-.07	-.05
13) Special issue	0.09	0.28	.08	.04	.05	.01	.15*	.04	-.07	-.04	-.03	.13	.04	-.12	-	.18*
14) Research note	0.06	0.24	-.04	-.06	-.05	.10	.01	-.03	-.07	-.03	.01	-.05	-.29**	-.13	-.05	-

*Correlations below the diagonal represent quantitative field studies ($n = 219$), correlations above the diagonal represent qualitative field studies ($n = 146$).

** $p \leq .01$

* $p \leq .05$

Two-tailed tests.

Table 5.5. Descriptive statistics and variable correlations

Variable	Article Impact (ISI citations/year)		
	Merged	Qualitative	Quantitative
Controls			
Dummies for journals	included	included	included
Gender of first author	.04 (.18)	.04 (.32)	.04 (.24)
First author: Europe	-.06 (.14)	-.02 (.21)	-.04 (.20)
First author: Asia	-.03 (.21)	-.09 (.36)	-.01 (.26)
First author: Oceania	-.06 (.31)	-.02 (.96)	-.08 (.33)
Number of authors	-.12* (.05)	-.16 (.08)	-.11 (.07)
Female ratio in author team	-.08 (.23)	-.04 (.38)	-.09 (.30)
Article age	.05 (.02)	.06 (.04)	.04 (.03)
Article length (pages)	.12 (.01)	.48** (.02)	.02 (.01)
Article position	-.05 (.15)	-.11 (.23)	-.05 (.21)
Issue type (special issue)	-.01 (.18)	.04 (.32)	.01 (.24)
Article type (research note)	-.05 (.23)	-.05 (.70)	-.06 (.25)
Main effects			
Transparency	.09 ⁺ (.06)	.12 (.11)	.02 (.08)
Transparency squared	.11* (.06)	.22 ⁺ (.09)	.11 (.11)
Interaction effects			
Transparency*article age	.17** (.16)	.21 ⁺ (.29)	.19** (.24)
N	365	146	219
R ²	.16	.18	.22
F	3.28**	1.42	2.72**

^aStandardized regression coefficients reported with standard errors in parentheses.

** $p \leq .01$;

* $p \leq .05$;

⁺ $p \leq .10$

Two-tailed tests.

Table 5.6. Results of regression analyses on consequences of article transparency (article citations per year)^a

Variable	Article Impact (total ISI citations)		
	<i>Merged</i>	<i>Qualitative</i>	<i>Quantitative</i>
Controls			
Dummies for journals	included	included	included
Gender of first author	.05 (.18)	.04 (.29)	.04 (.24)
First author: Europe	-.05 (.14)	-.01 (.19)	-.04 (.20)
First author: Asia	-.02 (.20)	-.08 (.33)	-.01 (.26)
First author: Oceania	-.05 (.30)	-.01 (.89)	-.06 (.33)
Number of authors	-.11* (.05)	-.15 (.07)	-.11 (.07)
Female ratio in author team	-.07 (.22)	-.03 (.35)	-.09 (.30)
Article age	.25** (.02)	.31** (.03)	.23** (.03)
Article length (pages)	.06 (.01)	.41** (.02)	-.01 (.01)
Article position	-.03 (.14)	-.07 (.21)	-.05 (.21)
Issue type (special issue)	.01 (.18)	.07 (.29)	.01 (.24)
Article type (research note)	-.04 (.22)	-.07 (.65)	-.04 (.26)
Main effects			
Transparency	.11* (.05)	.16 (.10)	.04 (.08)
Transparency squared	.12* (.06)	.24* (.09)	.11 (.11)
Interaction effects			
Transparency*article age	.16** (.16)	.25* (.26)	.18* (.25)
N	365	146	219
R ²	.21	.22	.26
F	4.53**	1.87*	3.49**

^aStandardized regression coefficients reported with standard errors in parentheses.

** $p \leq .01$;

* $p \leq .05$;

⁺ $p \leq .10$

Two-tailed tests.

Table 5.7. Results of regression analyses on consequences of article transparency (total article citations)^a

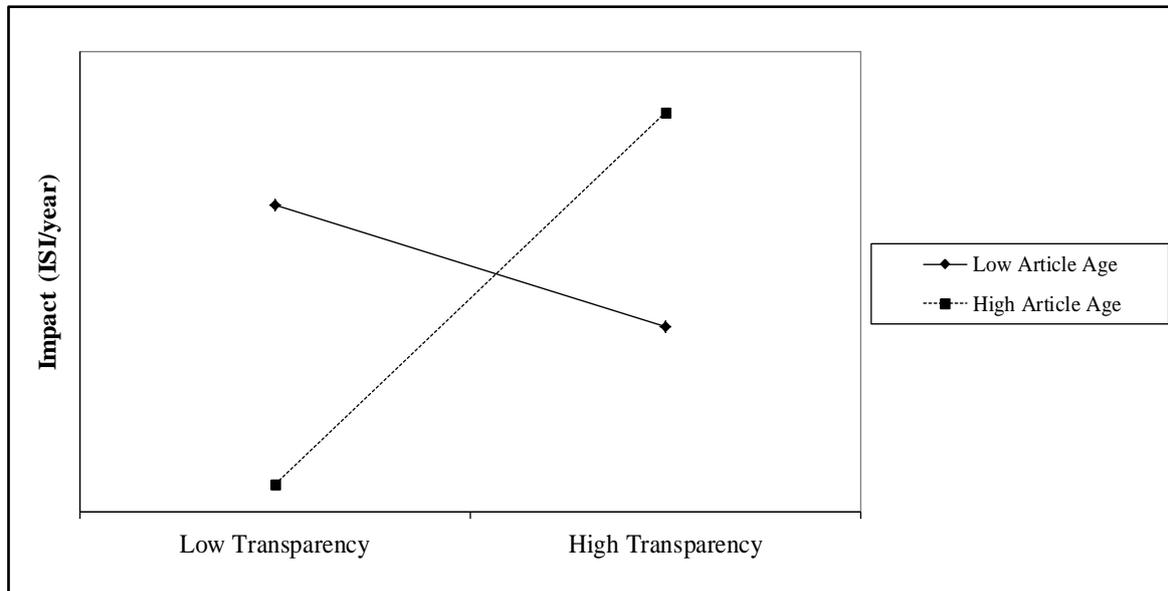


Figure 5.2. Plot of interaction effect

Antecedents of article transparency

Given that transparency turned out to be significantly related to article impact, both directly and indirectly, our next step was to examine whether some general features of journal articles and / or their authors systematically relate to the level of transparency in reporting. To do so, we regressed the articles’ transparency index on such general features. The results of this analysis are shown in Table 5.8. In these analyses we also used dummy variables to control for the journals the articles have been published in.

Only a few of the potential antecedents of article transparency turned out to be statistically significant. Among these significant antecedents, only one was consistently found to relate to article transparency for both quantitative and qualitative studies: article age. The older an article is, the less transparent it tends to be. To put it in other words, the negative relationship between article age and article transparency suggests actual reporting standards are increasingly requiring more article features to be reported, pointing to a trend of increasing transparency in the management literature.

Variable	Transparency Index		
	<i>Merged</i>	<i>Qualitative</i>	<i>Quantitative</i>
Dummies for journals	included	included	included
Gender of first author	-.06 (.02)	.03 (.04)	-.15 (.03)
First author: Europe	.01 (.02)	.07 (.03)	-.03 (.02)
First author: Asia	-.09 ⁺ (.03)	-.17* (.05)	-.09 (.03)
First author: Oceania	.03 (.04)	-.02 (.12)	.01 (.04)
Number of authors	-.01 (.01)	.03 (.01)	-.07 (.01)
Female ratio in author team	.02 (.03)	.14 (.05)	-.02 (.03)
Article age	-.14* (.01)	-.18* (.01)	-.22** (.01)
Article length (pages)	-.01 (.01)	.27* (.01)	.02 (.01)
Article position	-.05 (.02)	-.12 (.03)	.00 (.02)
Issue type (special issue)	-.05 (.02)	-.28** (.04)	.12 ⁺ (.03)
Article type (research note)	-.01 (.03)	.14 (.09)	-.08 (.03)
N	365	146	219
R ²	.08	.22	.12
F	1.79*	2.21**	1.68*

^aStandardized regression coefficients reported with standard errors in parentheses.

** $p \leq .01$;

* $p \leq .05$;

⁺ $p \leq .10$

Two-tailed tests.

Table 5.8. Results of regression analyses on antecedents of article transparency

In contrast, the other significant antecedents identified in our analysis are only applicable to either quantitative or qualitative papers. In this regard, qualitative papers tend to be less transparent if the first author is from Asia and if the article is shorter. The latter finding might thereby point to the more article space-consuming nature of transparency in qualitative studies compared to quantitative studies. For quantitative papers, these features did not turn out to be significant. One particularly eye-catching result was the revealed relationship between publication in a special issue and article transparency. While being significant for both quantitative and qualitative studies, the

direction of this relationship is in the opposite direction for the two types of field studies. Quantitative articles published in special issues tend to be more transparent (on a marginally significant level), while, in contrast, qualitative articles published in special issues tend to be less transparent.

5.6. Discussion

De facto reporting standards and their implications

In this study, we scrutinized how transparent the research procedures reported in 365 qualitative and quantitative field studies published in leading management journals are. Our results point to considerable heterogeneity in qualitative vs. quantitative papers. For example, in papers on qualitative field studies, reporting about the rationale for data selection and providing excerpts of raw data can be considered a standard procedure. Validity aspects, in contrast, are only seldom referred to and the variables under study are usually not clearly identified when it comes to qualitative papers. A similar view emerged in papers on quantitative field studies, where several of the coded aspects are clearly reported nearly by default, such as variable reliability metrics, response rates, correlations between study variables, or descriptive statistics. However, other coded aspects that are potentially important for assessing rigor of a quantitative field study and its analysis tend to be lacking, such as the indication of presence/absence of missing data or outliers in the data. That being said, a particularly striking finding might just reside in the most frequently mentioned features. This is because these features are so basic that one actually would expect them to be reported in any paper published in top tier journals, which, however, is not the case.

Questionable research practices or 'best' practices?

It is important to underscore that we do not intent to instrumentalize the inconsistent reporting practices and the strikingly varying degrees of transparency provided in the papers as a foundation to pass criticism to the parties involved in the publication process, i.e., authors, reviewers, and editors. As Aytug et al. (2012) noted, unless there are established and explicitly outlined reporting standards, as is the case in some other disciplines (e.g., Kilkenny et al., 2010; Lepage et al., 2001; Schulz et al., 2010) such as

bioscience, medicine, or methodological approaches such as meta-analysis (e.g., APA, 2008; Aytug et al., 2012; Kepes, McDaniel, Brannick, & Banks, 2013), one can hardly expect anyone to behave in accordance to such (absent) standards. This was also stressed by several of the associate editors in the interviews, who mentioned that there are no common standards of reporting transparency, neither a general one nor specific ones for the respective journals (with some exceptions for particular indicators in some journals). At the same time, precisely because the field of management studies at present suffers from underdefined reporting standards when it comes to transparency, the dividing line between “best” research practices and questionable research practices is not clear, leaving authors, editors, and ultimately readers unsure about how to best showcase their work in the short term, while fending off potential criticism in the longer term (e.g.; Banks et al., 2016).

In that sense, the present study marks a step ahead towards the establishment of reporting standards for field studies in management by portraying the *de facto* reporting standards in this domain. Whether the current standard and the corresponding level of transparency are functional and desirable with regard to enabling the detection of methodological rigor can now be discussed. In sum, when thinking about where we want to go and what to change, however, one first needs to know where we actually are at the moment. It is this status quo of reporting practices that our article detailed out. Our results therefore provide an evidence-based foundation for such discussions that might ultimately lead to taking action in shaping the actual field study reporting standards in a way that more closely resembles the standards desired in our field. These discussions should also cover whether the desired level of transparency actually should approach complete transparency, or as one editor put it: “It is possible for people to overdo things as well. One can be ‘scientific’, using methods for the sake of methods. This sort of overelaboration is not necessary.” Moreover, journal and manuscript space is scarce and valuable. This might put ‘natural’ limits to an article’s transparency unless electronic appendices gain more acceptance and prevalence in the field of management. After all, most journals explicitly mention to judge the contribution of submitted papers in relation to the page count. Generally, most editors stressed the importance of offering a degree of

transparency that allows for replicating a field study, echoing recent calls in this direction (e.g., Banks et al., 2016; Nosek et al., 2015; O'Boyle, Banks, & Gonzalez-Mulé, 2014).

Transparency “pays off”: transparent papers get cited more

Reporting standards matter not only from a methodological perspective. Our analysis showed that the level of article transparency is related to article citation and therefore, could be included as an additional meaningful explanatory variable in studies on article impact. Our study contributes to literature by suggesting this feature, which has been missed out in management article impact literature so far (Bergh et al., 2006; Flickinger et al., 2013; Judge et al., 2007). We found significant direct effects of transparency on citation counts that point to the importance of transparency for the decision to refer to field studies' results. Even more interesting is the interaction effect between article age and transparency on article citation, suggesting that transparency is an important facilitator of article half-life (Walters, 2011). The relationship between transparency and citation counts was more strongly positive for older articles while there was no significant effect for younger articles. In other words, transparently reporting methodological aspects and results directly leads to more sustainable article impact. An explanation of this finding might be that a higher degree of transparency allows today's scholars to better evaluate whether the applied methods in older papers still comply with current standards of rigor and thereby facilitate citation. An elevated level of transparency might cause higher trust in the findings (Bråten, Strømsø, & Salmerón, 2011; Nicolaou & McKnight, 2006), especially when one is not sure about past standards of rigor. This is mirrored by statements of several editors in our interview study, who emphasized the pronounced role of transparency for creating trust with regard to a field study's findings (e.g., “the main criterion is to trust the results”). An alternative interpretation of this finding, however, could be that authors able to design field studies that have a high impact in the literature are not only able to offer important contributions that advance the theoretical state of the art, but also tend to go beyond current methodological standards and offer an elevated degree of transparency in reporting their studies. This might matter even more, given that our data suggest a trend

of increasing transparency over the examined period of time that materialized in a significant negative effect of article age on article transparency, which implies that papers being more transparent in times characterized by lower levels of transparency might stand out more visibly in the crowd of publications. Irrespective of the specific reason underlying the observed interaction effect, we can sum up that article transparency benefits citations, particularly in the long run. This knowledge might also provide an incentive to journal editors and reviewers to ensure high levels of transparency of papers reporting the results of field studies and help in implementing appropriate procedures to guarantee for these high levels of transparency, for example, as outlined by Nosek et al. (2015). Not only does it help readers to better understand and evaluate the studies' findings, and help other researchers in replicating the findings, but also it actually pays off in terms of bigger and prolonged impact of articles published in their journals.

Transparency of qualitative vs. quantitative field studies

Our study also uncovered commonalities as well as important differences between qualitative and quantitative field studies in these issues. First of all, in both methodological camps we see a clear trend towards increasing transparency over time. One could interpret this as an indicator that management scholars, irrespective of the specific methodological approach, have recognized and increasingly internalized the value of transparency as necessary for evaluating a study's rigor (Moravcsik, 2014), as has been repeatedly called for (e.g., Daft & Lewin, 2008; Edwards, 2008; Gibbert et al., 2008; Nosek et al., 2015). Moreover, the interaction effect (older papers showing a stronger positive relationship between transparency and impact) was replicated in both qualitative and quantitative field studies, pointing to the universal value of transparency.

In contrast to these converging results for qualitative and quantitative field studies, we also found notable differences between papers using the two different approaches. First, the result that quantitative papers tended to be more transparent and reported a significantly higher portion of the coded aspects than qualitative papers, stands out. Several reasons might account for this finding. One might say it is rooted in

the fact that it is simply more convenient to achieve a higher level of transparency in quantitative papers, since most of the aspects are expressed by numbers or coefficients and thus do not need space consuming explanations or descriptions and can even nicely be integrated in (already included) tables. In qualitative research, in contrast, most aspects to be reported require elaborated and voluminous explanations and descriptions.

For example, reliability of variables in quantitative studies can be established by presenting specific coefficients that need to comply with certain (more or less formally established) criteria or thresholds such as Cronbach's alpha (Hinkin, 1995). In qualitative research, however, establishing reliability requires a convincing and therefore usually extensive explanation to show that the study's design and analyses precluded an undue influence of random error on the conclusions and that these conclusions are consistent to a potential replication of the steps taken in the study (Denzin & Lincoln, 2005; Gibbert et al., 2008). This was also underlined by several editors in our interviews.

For example, one editor emphasized that qualitative reporting requires a sophisticated process and another one stated: "Since you [the researcher] are the instrument in a qualitative study, you should show how you made sure the study is not about your biased views". Therefore, it does not come as a surprise that article length (in pages) is significantly positively related to transparency for qualitative papers but not for quantitative papers. This might result in the reluctance of authors to be more transparent, or in removing initially reported aspects from the paper after reviewers in the review process could convince themselves of these aspects' appropriateness (although this practice appears highly problematic since if this information was included in the published version of an article not only reviewers but also readers of the paper could have benefitted). Having said that, there might be further aspects that might influence the observed differences in the level of transparency between quantitative and qualitative papers. One is likely to stem from the huge discrepancy in the number of papers published using these two methodologies, only a small fraction of papers in top journals build on qualitative methodology (Gibbert et al., 2008). Our dataset, for example, comprises the entire set of qualitative studies (146 papers) published in the journals and

time period under study. In turn, there were several thousand quantitative papers published in the same journals and time period. This higher number of publications using quantitative field methods might result in a quicker establishment of implicit reporting standards and reduce variation in reporting due to a bigger number of study exemplars. Irrespective of the reasons for the differences, however, we see that papers publishing results of qualitative studies tend to be less transparent than their quantitative counterparts, which further highlights the need for transparency particularly in qualitative studies (e.g., Doz, 2011; Dubois & Gibbert, 2010; Moravcsik, 2014). This appears particularly important, given the frequently observed underspecification of applied methods, which Gephart (2004) pointed to as a major hindrance of qualitative research getting published.

Regarding the influence of article transparency on article impact, the results strikingly showed that there is only a direct curvilinear effect for qualitative field studies and that the significant positive linear effect of transparency on article impact in the complete sample is also clearly driven by papers using a qualitative methodology. In contrast, papers using a quantitative methodology did not show significant direct effects between these variables, with particularly small coefficients for the linear effect. Gibbert et al. (2008) point to one potential reason for this observation. While acknowledging the core role of rigor for every kind of study methodology, they argue that rigor in qualitative research is particularly important due to its closeness to managerial application and its early position in the theory development process, which might entail severe consequences for later stages of theory development when starting with insufficiently based assumptions (Gibbert et al., 2008). Unfortunately, qualitative research has sometimes struggled with methodological rigor (e.g., Daft & Lewin, 1990; Denzin & Lincoln, 2005; Eisenhardt & Graebner, 2007; Yin, 1981), precisely because reporting standards in qualitative research have traditionally been less codified, and because there is a larger share of scholars who lack proficiency in qualitative methods than there is for quantitative methods (Doz, 2011). Thus, relaying rigorous procedures in qualitative work cogently, i.e. ‘walking the talk’ (Gibbert & Ruigrok, 2010) simply takes time (and article space).

Another eye-catching difference between qualitative and quantitative papers, which our results showed, is the role that special issues play in predicting article transparency. Specifically, we found substantial differences relating to the transparency of papers published in special versus regular issues in that quantitative papers published in special issues tended to be more transparent, while the opposite was the case for qualitative papers. Prior research on the role and attributes of special issues in management research (Conlon et al., 2006; Olk & Griffith, 2004) has presented some arguments on how the peculiar publication process in special issues might have an impact on article properties. In this regard, Olk and Griffith (2004) report top management journal editors' concerns that quality of papers in special issues could be lower than in regular issues due to higher acceptance rates (Conlon et al., 2006; Smith, 1999; Tsui, 1999), which might hint to less transparent reporting practices in special issues. Moreover, Conlon et al. (2006) showed that authors of papers published in special issues tend to have less prior publications than authors in regular issue papers. This authorship pattern reflecting a higher variety of authors and less established authors in special issues, is also likely to result in a higher variety and a more pronounced deviance of reporting practices commonly applied in regular issues, which might also suggest lower transparency levels. However, only the qualitative portion of the sample behaves in line with these expectations based on previous considerations and findings. Why quantitative field studies do not follow the same pattern and even tend to be more transparent when published in special issues is unclear and opens up an avenue for more detailed future research. For instance, further research could have a more detailed look at whether the specific type of special issue might play a role here (Conlon et al., 2006). Thus, our results also bear implications for a better understanding of how the growing number of special issues influences and contributes to the advancement of the field of management (Conlon et al., 2006; McKinley, 2007), by showing significant influences on reporting practices in field studies that, however, diverge for qualitative and quantitative studies.

5.7. Limitations

The analyses in this paper bear several limitations that might stimulate future research. First, we examined papers from seven top management journals published during a complete decade (1997-2006). While this represents a substantial time period to study developments over time which is in line with prior research on related topics (e.g., Bergh et al., 2006; Judge et al., 2007), still longer time periods might be desirable to test for longitudinal effects and to get a more complete picture of transparency in field studies. Similarly, other sets of top-journals in the field of management might be chosen, and our results might change depending on the set of specific journals included in the sample. However, we focused on those management journals that have been consistently mentioned among top tier journals in management, particularly for the period under study (Gomez-Mejia & Balkin, 1992; Podsakoff et al., 2005; Tahai & Meyer, 1999), and also regularly publish results of field studies. Moreover, to minimize journal-specific biases in our results we controlled for this aspect in all our regression models, recognizing that we cannot completely rule out any sample-specific findings, particularly in the descriptive analyses. Despite these limitations we believe that our analyses provide an empirically grounded starting point for developing appropriate reporting standards for field research in management and stimulate a discussion on how they should look like for qualitative and quantitative methodologies.

Chapter VI*

Hot on the audit trail: How to assess methodological transparency of grounded theory in management?

Lakshmi Balachandran Nair

Prof. Michael Gibbert

Journal

Nair, L.B., & Gibbert, M. International Journal of Management Reviews (work in progress)

Conferences

Nair, L.B. (2016, August). 75th Academy of Management Annual Meeting, Anaheim, USA (Awaiting decision).

Nair, L.B. (2016, June). European Academy of Management Conference, Paris, France (Awaiting decision).

Nair, L.B. (2015, August). 75th Academy of Management Annual Meeting Content Analysis PDW, Vancouver, Canada.

6.1. Abstract

Iterative methods such as grounded theory rely on several cycles of comparing empirical phenomena with theoretical inclinations until saturation is reached. We propose a new method for assessing the methodological rigor of these procedures ex-post using an audit trail perspective. Analyzing all grounded theory articles published in ten major management journals 1970-2010, we find that key parameters are underutilized. In particular, negative cases which are the starting points for constant comparisons, are largely not considered. Likewise, majority of articles do not problematize theoretical sampling, leaving it unclear whether theoretical saturation is reached or not.

Key words: *Grounded theory, audit trail, transparency, qualitative content analysis, rigor*

6.2. Introduction

‘‘A full and true report is the hallmark of the scientist, a report as accurate and faithful as he can make it in every detail. The process of verification depends upon the ability of another scientist who wishes to repeat a procedure and to confirm an observation.’’ Glass (1965: 83)

For rigorous theory building, there should be a strong but flexible link between the empirical and the conceptual worlds, and this link relies on iterations between (emerging) theory and empirical data leading to an interpretive understanding (Glaser, 1978; Kaufmann and Denk, 2011; Strauss and Corbin, 1998; Tsang and Ellsaesser, 2011; Van Maanen, Sorensen, and Mitchell, 2007;). One such iterative approach is grounded theory (Glaser and Strauss, 1967). It includes a set of systematic procedures for gathering, synthesizing, analyzing and conceptualizing qualitative data for theory building - a central activity for management researchers (Eisenhardt, 1989). By definition, the iterative nature of grounded theory requires the researcher to constantly improvise his research techniques and methods, which can make them seem ‘‘messy’’ (Eisenhardt, 1989). The characteristic messiness of iterative research can often make it difficult for the ‘customers’ (i.e. readers, reviewers, and editors) to *audit* the research procedures, i.e. to perceive and appreciate the chain of evidence, and the boundaries between various stages of the study.

Auditing apparently messy research procedures ex-post is not straightforward. To help it along, it was suggested that model mis-fit due to negative cases (cases which disconfirm one’s theoretical inclinations) should actually be problematized in the research procedures, taken as an explicit impetus for further data collection, and relayed in the final write-up. This practice of exploring negative cases would make it easier to illustrate where each iterative cycle starts and stops, and when theoretical saturation is reached (Eisenhardt, 1989). This is in stark contrast to other empirical methods, in particular quantitative field and laboratory research, where the final write-up typically reads like a neat account of a linear process which produces results that confirm or disconfirm one’s hypotheses. In iterative research procedures in particular, then, a clear

audit trail, the chain of evidence of how the researcher moved from data to theory, the stumbling blocks encountered on the way, and the confidence with which theoretical saturation was reached facilitates subsequent replications, evaluations, and citations of the study (Judge, Cable, Colbert and Rynes, 2007; Tams and Grover, 2010).

Additionally, since grounded theory is intimately linked to practical evidence, a clear audit trail would underscore its relevance for managerial practice (Eisenhardt, 1989; Lawrence and Tar, 2013; Orlikowski, 1993).

Unfortunately, unlike deductive studies, including quantitative field and laboratory research, iterative research such as grounded theory suffers from less-well established reporting conventions for the final write-up, and there is less agreement regarding the structure, content, and format of audit trails (Gibbert, Ruigrok, and Wicki, 2008; Kauffman and Denk, 2011; Orton, 1997; Weiss et al., 2015). For instance, where does data collection stop? And where does data analysis start in these iterative cycles? How could negative cases be reconciled with emerging theory so as to reach theoretical saturation? As a result, the rigor of audit trails remains very much in the eye of the beholder (e.g. Gibbert and Ruigrok, 2010), making clear benchmarks difficult to establish and implement (Yin, 1994). This lack of understanding what an audit trail entails is highly problematic at various stages of the consumption of research results. For instance, gatekeepers such as reviewers and editors often lack generalized procedures for assessing the rigor of qualitative research procedures, and instead require authors to “walk the talk” when it comes to methodological clarity (Gibbert and Ruigrok, 2010). Similarly, readers of an eventually published article often struggle to reconstruct individual cycles of iterative sensemaking (Aguinis et al., 2010), which can cause ripple-effects throughout the later stages of development and subsequent testing of that theory in deductive studies (Eisenhardt and Graebner, 2007; Kauffman and Denk, 2011), and, in case of doubt, can even lead to the retraction of a published article (e.g. www.retractionwatch.com for recent updates).

While there are several normative guidelines on how to practice grounded theory (Charmaz, 2003; Corbin and Strauss, 1990; Glaser and Strauss, 1967; Strauss and

Corbin, 1998), there is a lack of empirical research on how these recommendations play out in practice, i.e. how to assess an audit trail in a grounded theory study. The few notable exceptions typically focus on auditability from the perspective of the authors of the audit trail (i.e. the researchers themselves), rather than from a customer perspective (Bowen, 2009; Halpern, 1983; Lincoln and Guba, 1985). Such a ‘customer-centric’ perspective on the transparency of research procedures (Aguinis et al, 2010; Weiss et al., 2015) places the burden of the most accurate conclusions on the reader (reviewers, editors, as well as readers) and ensures complete transparency of the individual cycles in the iterative approach, so that the main conclusions do not appear to arbitrarily support one’s theoretical inclinations (Aguinis et al, 2013: 292).

The purpose of this research is threefold. First we examine the degree of sophistication and transparency of audit trails in grounded theory i.e., the extent to which they transparently report research procedures, for reviewers, editors (‘intermediate customers’ of the article) and readers (‘end-customers’). We find that this auditability is represented poorly in management grounded theory research. We then build on literature in auditing to propose a new perspective to facilitate the appreciation of iterative research procedures, which we call ‘second party auditability’. We use this second-party auditability to specifically investigate how often core grounded theory parameters, such as constant comparison, theoretical sampling, triangulation, the analysis of negative cases, and theoretical saturation are reported (Glaser and Strauss, 1967; Strauss and Corbin, 1998; Suddaby, 2006) in top management journal grounded theory articles (1970-2010). We find that these parameters are grossly underrepresented. To improve second-party auditing, we provide an exemplar of an article with excellent (> 81%) second-party auditability.

6.3. Auditability in qualitative research

Audit trails

One way to ensure the rigor and transparency of a research study is to preserve an ‘audit trail’ (Bowen, 2009), otherwise known as a ‘decision trail’ (Koch, 1994). An audit trail is a ‘residue of records stemming from the inquiry’ (Lincoln and Guba, 1985). It is a

thorough collection of documentation (Rodgers, 2008) of the research procedures undertaken and decisions made during the course of a study, from the pre-data collection stage to the presentation of results.

Previous research in qualitative research has underscored the importance of audit trails (Halpern, 1983; Lincoln and Guba, 1985) in the verification of the main theoretical, methodological, and analytical choices made during a study (Koch, 1994). Audit trails provide a systematic recording and help in establishing, increasing, or ascertaining the rigor of a qualitative study (Lincoln and Guba, 1985).

Audits: Internal and external

To address the issue of low auditability and rigor in qualitative inquiry, Halpern (1983) proposed a prototype of an audit trail, along with an algorithm for the auditing process. Particularly relevant for our purposes of constructing an audit trail from a customer-centric perspective, other previous studies have listed two types of audits which are applicable to qualitative research: internal and external. An internal auditor is a colleague or a team member who provides balances and checks for each other. Internal auditing thus decreases investigator bias and make sure the research process is rigorous and replicable (Rodgers, 2008). External auditing, on the other hand, involves the expertise of a ‘proficient but independent colleague’ who is not involved with the concerned research. The researcher would seek the assistance of an external auditor in evaluating the research process through examination of research records, peer debriefs etc. (Rodgers, 2008).

In both these scenarios the readers and reviewers aka. the intermediate and end-customers of the research are not directly involved, since the construction of the audit trail involves a ‘residue of records’ (Lincoln and Guba, 1985) which is typically not available to the article’s customers, especially to the readers or end-customers. This is unfortunate, as it tends to leave a knowledge gap between the producers (authors) of an audit trail and its intermediate and end-consumers. Against the characteristic messiness of grounded theory’s methodological procedures, this knowledge gap is counter-productive, as apparently messy research procedures easily come across as non-rigorous.

Drawing on the literature on audits allows us to bridge this gap. We put forth a new classification of audit trails, which involves the readers and reviewers of a research article in appreciating and verifying the process of theory development (Bowen, 2009; Koch, 1994).

Audits: First, second, and third parties

Adapting from the audit literature, we propose to classify audits into first-party, second-party, and third-party audits. The idea behind the first-party auditing is similar to the existing internal auditing. It is the informal checking of procedures which happen during the course of a qualitative inquiry. It is performed within a research team or organization, to measure a study's strength and weaknesses against its own procedures or methods or external standards, and ensures the rigor and credibility of a given study (Russell, 2012). The existing external auditing procedures would concur with our definition of third-party audits. A third-party audit is independent, with no conflict of interest (Russell, 2012). The auditors are not involved in the researcher-reader relationship. They could be blind peers or uninvolved experts, who are not interested in the study.

To these two audits, we add on one more type of external audit which has not yet been discussed in the context of grounded theory research: the second-party audit. Here, auditing is done by journal reviewers, editors, and end-readers (the 'customers') of the article (Weiss et al., 2015). This auditing happens once the whole study is completed and presented in the form of a manuscript. Naturally, auditability at this stage is important in decisions regarding the acceptance of the article for publication, integration of the article into the academic body of knowledge, and its replications, and even the eventual scholarly impact. Table 6.1. shows the new classification of audits.

	First-party audits	Third-party audits	Second-party audits
Audit category	Internal	External	External
Purpose of audit	Self-evaluation	Compliance to standards and good research practice vouched by an external member. High quality of subsequent manuscript.	Better understanding of the study and its trustworthiness by customers, especially readers. Facilitation of further replications and extensions of the study. High quality of academic exchange.
Selection of Auditor	Decided by the researchers involved	Decided by the researchers involved	Not decided by the involved researchers
Auditors	Team members, Involved colleagues	Uninvolved experts in methodology (with sufficient expertise in the area of research)	Journal editors, Reviewers, Readers
Audit period	During the course of the concerned study	From the early stages of the concerned study or After the concerned study, before the final manuscript is prepared. (Lincoln and Guba, 1985)	After the manuscript is written. Usually when the manuscript is submitted to a journal or after it gets published.
Problems associated	Over-identification between the inquirer and auditor (Rodwell and Byers, 1997) Expectations and common value orientations	Early cooption (early stage auditor involvement) Paucity of information (late stage auditor involvement) (Lincoln and Guba, 1985)	Have no or limited access to raw data and field procedures (especially the readers or end customers of the article)
Documents/procedures reviewed	Internal reports, Archival records, Interview transcripts, Memos, Methodological log (Patton, 1990), Reflexive journal (Lincoln and Guba, 1985)	Internal reports, Archival records, Interview transcripts, Memos, Methodological log (Patton, 1990), Reflexive journal (Lincoln and Guba, 1985)	Completed manuscript (for editors and reviewers), Published article (for authors)
Documents/procedures not reviewed	Final published manuscript	Final published manuscript	Internal reports, Archival records, Interview transcripts, Memos, Methodological log, Reflexive journal
Available audit procedures	Halpern, 1993	'Halpern Algorithm' (Halpern, 1993; Lincoln and Guba, 1985)	--

Table 6.1. New classifications of audits

6.4. Auditability of research procedures: In the eye of the second-party?

Due to a lack of cohesion in reporting conventions when it comes to qualitative studies, it is currently far from clear when a given grounded theory study is transparent (enough) for second-parties, making the assessment of rigor and results far from accessible, convenient, and ultimately reliable. This is in contrast to quantitative studies which are characterized by far more codified reporting conventions, with veritable checklists and widely-accepted manuals in particular for laboratory-based research (APA, 2008; Beyer, Chanove, and Fox, 1995; Gephart, 2004; Gibbert and Ruigrok, 2010; Scandura and Williams, 2000).

While in quantitative research auditability can be ensured by providing transparent, quantified information about the collection, analysis and reporting of data, in grounded theory it refers to ‘talking the walk’ or guiding the customer through various stages of the argument, convincing her of the cogency of the argument and its conclusions. By disclosing the underlying logic, reasoning, and causalities behind a study, the knowledge gap is reduced, which makes it easier for readers to decipher its results better and even replicate them. Since grounded theory in management is characterized by a ‘state of confusion’ (Jones and Noble, 2007; Parkhe, 1993), auditability regarding the different stages of the research process would also help decrease the perception that the messiness of grounded theory research is indicative of inadequate validity and subjectivity (Denk, Kaufmann, and Karter; 2012; George and Bennett, 2005).

However, an audit trail presenting grounded theory in its pure form, as a jumble of iterative steps of literature consultation, data collection and analysis (Suddaby, 2006); would make it overly complex and difficult to understand for researchers, in particular those who work in other, more linear and deductive paradigms. To reduce this complexity, we are focusing on ‘second-party auditors’, who might not always be expert qualitative researchers (Rodgers and Cowles, 1993). As such, and to facilitate better perception by the article’s intermediate and end-customers we categorize the audit actions according to the individual stages of research process pertaining to the trustworthiness model put forth by Lincoln and Guba (1985).

As a first step in the auditing process, we examine the auditability parameters involved in the ‘pre-data collection’ stage (for example, the nature of the study, how the research question was defined, and how the instrument was developed). Since prior research has suggested that the criteria for judgment of grounded theory research should be based on the detailed elements of the actual strategies used for collecting, coding, analyzing and presenting data (Strauss and Corbin, 1998), further audit trail parameters are categorized along these stages (see table 6.2.)

Stages of the study	Audit trail reports	Literature
<i>Pre-data collection</i>	1 Definition of the research question	Weiss et al., 2015; Halpern, 1983; Kaufman & Denk, 2011
	2 Instrument development (why GT was chosen)	
	3 Nature of the study <ul style="list-style-type: none"> - Verifying (theory testing)/ deductive - Generating (theory building)/ inductive - Abductive 	
<i>Data collection</i>	4 Rationale for research setting selection	Corbin & Strauss, 1990; Cook & Campbell, 1979; Denzin, 1970; Denzin & Lincoln, 2005; Gibbert et al., 2008; Pettigrew, 1990; Stake, 1995; Weiss et al., 2015; Welch et al., 2010; Yin, 1994
	5 Sampling technique: Theoretical sampling	
	6 Context of data collection	
	7 Sources of data used	
	8 Triangulation (data, researcher, theory)	
	9 Data collection process	
10 Review by key informants		
<i>Data coding and analysis</i>	11 Memo/ list of indicators (events, incidents, actions)	Cook & Campbell, 1979; ; Corbin & Strauss, 1990; Gibbert et al., 2008; Gibbert, Nair, & Weiss, 2014; Glaser & Strauss, 1967; Halpern, 1983; Kaufman & Denk, 2011
	12 Coding principles	
	13 List of categories derived from data and literature	
	14 Identification of the core category in emergent theory	
	15 Description of how the categories are linked to the core category	
	16 Diversity in data considered: each incident is compared with other incidents, properties of a category (similarities and dissimilarities)	
	17 Constant comparison	
	18 Theoretical saturation	
	19 Analysis of negative cases	
	20 Discussion of theoretical notions with team mates/peers	
<i>Presentation of results</i>	21 Using illustrations for properties	Eisenhardt & Graebner, 2007; Gibbert, et al., 2008; Gibbert & Ruigrok, 2010; Halpern, 1983; Kaufman & Denk, 2011; Lincoln & Guba, 1985)
	22 Explanation of new theory	
	23 Theoretical implications	
	24 Rigor parameters (credibility, dependability, conformability, transferability)	
	25 Managerial implications	
	26 Limitations of study	
27 Future directions		

Table 6.2. Second-party audit trail parameters for grounded theory (GT) articles

Pre data collection stage

The pre data collection stage commences the steps in the four stage design proposed by Strauss and Corbin (1998). The researcher should explicate how he or she *formulated the preliminary research question* to identify the phenomenon of interest. The question should ideally direct the researcher towards the involved actions and processes (How did the phenomenon of interest occur?) rather than the conditions involved (What caused the phenomenon of interest? Why did the phenomenon of interest occur?). Evidently, as the study progresses the research question would also change accordingly (Strauss and Corbin, 1998).

The second step in this stage is the development of the instrument. The audit trail at this stage would involve providing details behind the *rationale for selecting the instruments*. For instance, in the paper by Soklaridis (2009), in-depth interviews were conducted since,

“It was possible to gain an understanding of how the stakeholders at the clinic interact with each other, how they interpret those interactions, and how these meanings are informed by the wider socio-political context in which IHC (Integrative Health Care) takes place.”

The third step in the pre data collection stage involves discussions about the *nature of the concerned study*, whether it is purely inductive (Glaser, 1978), deductive (Strauss and Corbin, 1998) or abductive (Charmaz, 2001; Locke, 2007). An effective grounded theory study would ideally be inductive, abductive, or a combination of induction and deduction (Strauss and Corbin, 1998).

Data collection stage

Audit trail pertaining to the data collection stage would involve providing the *rationale for selecting the research setting* (Cook and Campbell, 1979), and the *sampling technique used* i.e., theoretical sampling. Theoretical sampling by definition is directed by the emerging theory, and focuses on data collection to refine and integrate the same theory.

“It is the process of data collection for generating theory whereby the analyst jointly collects, codes and analyses his data and decides what data to collect next and where to find them, in order to develop his theory as it emerges.” (Glaser and Strauss, 1967: 45).

Like many other steps in grounded theory building, theoretical sampling would involve going back and forth between the stages of data collection and data analysis, and hence could be part of both these stages simultaneously. We include such steps only once, to avoid double coding and bias.

Other steps in the data collection stage includes the *context of data collection* (Cook and Campbell, 1979), the *sources of data* used: interviews, observation, focus group discussions, informal chats etc., mentioning of *triangulation* (Denzin and Lincoln, 2005; Yin, 1994), *details of the process of data collection*, and *review of collected data by key informants* (Gibbert et al., 2008; Gibbert and Ruigrok, 2010; Weiss et al., 2015).

Data coding and analysis stage

Data coding stage involves explication of *creating a memo*, or a list of indicators of events, incidents, and actions (Kauffman and Denk, 2011) that depict the relationship among concepts (Lawrence and Tar, 2013). Memos help to tie together different pieces of data into a recognizable cluster of instances of a concept (Lawrence and Tar, 2013). A similar step in this stage is *explicit mentioning of coding principles* involved. Coding principles facilitate clearer understanding of the logic behind fracturing the data and conceptually grouping them into codes which in turn leads to a theory which explains the same data (Glaser, 1978). Other codes pertaining to data coding are *list of categories derived from data and literature*, *identification of the core category*, and *description of how other categories are linked to core category* (Kauffman and Denk, 2011).

Audit trail parameters concerning data analysis stage include ‘chain of evidence’ which allows the reader to follow the derivation of evidence from initial to later or final theoretical statements, and trace the chain in either direction (Yin, 1994). An auditable data analysis section would thus give clear details of all the steps involved in the data coding and analysis stage. The first parameter here is to be transparent about *how the*

diversity in data was considered i.e., whether each incident was compared with other similar and dissimilar categories and incidents.

Constant comparison is another parameter which is important for ensuring the richness and rigor of a grounded theory study, and hence another parameter which should be reported transparently. Constant comparison involves building and expounding a category by constantly comparing the emerging category with the concerned data and new data (Glaser, 1978). Constant comparison along with theoretical sampling are stopped when the researcher achieves *theoretical saturation* or the point of minimal incremental learning from further comparison or sampling (Eisenhardt, 1989; Glaser, 1992; Lawrence and Tar, 2013). *Analysis of negative cases* (Gibbert et al., 2014) which do not confirm to the observed patterns is yet another prerequisite for an audit trail. In fact, the analysis of negative cases constitutes a kind of starting point for iterative cycles: once we perceive model *mis-fit*, we need to engage in further data collection and reconsider our initial theoretical inclinations. Without negative cases, there is much less of an impetus to start another iteration before reaching saturation. Finally, the *discussion of one's theoretical notions with team members and unaffiliated peers* would also be beneficial from auditable theory building point of view.

Presentation of results stage

In this stage, the trail include *illustrations or excerpts of data indicative of the involved properties* (for example, interview excerpts) (Gibbert et al., 2008), and *explanation of the new theory* (Kauffman and Denk, 2011). The researcher should also be transparent about the *theoretical and managerial implications* of the study, as well as the *limitations and future directions the study could take*.

Finally, an auditable grounded theory study should also include explicit mentioning of the rigor parameters. In this study, we consider the naturalistic criteria of *credibility, dependability, conformability, and transferability* for ascertaining rigor (Lincoln and Guba, 1985). Credibility (roughly similar to internal validity) examines the congruence of a grounded theory study's findings with the reality (Merriam, 1998). Dependability is similar to the positivistic notion of reliability and focuses on the

replicability of the study. Conformability addresses the objectivity of the study i.e., how to reduce the influence of investigator bias (Patton, 1990). Transferability, similar to (positivistic) external validity, is concerned with how generalizable the study results are in a similar setting (Shenton, 2004).

6.5. Methods

We conducted two studies, a pre-study to test the second-party audit trail, as well as a main study where we use it to assess the methodological sophistication of all grounded theory papers in major management journals. We selected these journals based on their quality. The quality of scholarly journals is usually judged based on their relative impact (Aguinis et al., 2012). An article in a top journal is highly likely to represent the best practices in the discipline. Recently, Weiss and colleagues (2015) have proposed parameters to measure the transparency of management research articles and their relationship with the citation count. The current study extends such efforts by suggesting a customer-centric approach focusing specifically on in-depth reporting of published grounded theory articles. With the new criteria developed, we propose to review the grounded theory articles in top management journals to evaluate their level of auditability (main study). Before doing so, to ensure the fit of our coding parameters with practices observed in management grounded theory studies, we do a preliminary review of all such articles which appear in management journals during the time period 2003-2005.

For the pilot study, we used directed content analysis (Hsieh and Shannon, 2005; Potter and Levine-Donnerstein, 1999) and scrutinized all grounded theory articles from management journals, from the time period (2002-2005), using the sample by Jones and Noble (2007). Additionally, to make sure we did not exclude any papers during this time period, we do a key word search using the term ‘grounded theory’. The preliminary descriptive statistics pointed out that all the codes which we have included in our coding sheet are in fact reported in management grounded theory research. This validates our coding sheet. The details of the preliminary study are included in appendix A.

The pilot study also served as a training session for the coders to get familiarized with the management grounded theory articles. Since each grounded theory article could report the parameters in different ways, we did not limit our coding to the methodology section only. We read each article completely (Gibbert et al., 2008) and looked for implicit and explicit mentioning of relevant audit trail criteria.

6.6. Main study

The chief purpose of the preliminary study was to validate the coding parameters. Using the coding criteria we developed from existing literature (For example, Kaufman and Denk, 2011; Lincoln and Guba, 1985; Weiss, Nair, Gibbert, and Koepplin, 2015) and refined through the pilot study, we looked into the grounded theory articles in top ten management journals from 1970 to 2010. We select the journals using Thomson Reuters Journal Citation Reports (2014). The journals we considered are *Academy of Management Journal*, *Journal of Management*, *MIS Quarterly*, *Journal of Applied Psychology*, *Journal of Operations Management*, *Administrative Science Quarterly*, *Organization Science*, *Personnel Psychology*, *Journal of International Business Studies*, and *Strategic Management Journal*.

Altogether 27 grounded theory articles were published in the top management journals during this period. We included them all into our study. We also use a longer time period of four decades (1970-2010). This time period was chosen because grounded theory started getting popular in management research as early as in the 1970s (Locke, 2001). The extended time frame helps us comparatively analyze the trends in the field.

Since the positivistic rigor and transparency criteria are incompatible with the audit trail in grounded theory method, we use an interpretivistic coding criteria, drafted specifically for assessing grounded theory articles (based on the audit parameters in Table 6.2). We believe that despite the fact grounded theory has branched out to form different variants, all of them are still similar in terms of how they collect, analyze, and build theory from data i.e. from processes, to abstract categories, to theory (Charmaz and Mitchell, 2001).

The codes are dichotomous: a paper which meets a particular code is marked with '1' and a paper which does not, is marked with '0'. Since grounded theory involves continuous, systematic and sequential stages of research processes, a parameter might not always be confined to one stage only. Nonetheless, to avoid duplicity of parameters, we include each of them only once, in the stage where they appear to fit the most.

Apart from these factors, similar to our approach in the pilot study, we also examine whether the studies explicitly mention the four interpretivistic rigor criteria. For these reports on rigor, we adapt the interpretivistic model that is, credibility, dependability, transferability, and confirmability (Lincoln and Guba, 1985). At the same time, we also coded articles which explicitly report the positivistic parameters of rigor. These articles were coded '1' for their respective interpretivist criterion (e.g.: a paper mentioning reliability was coded given a value of '1' for the dependability code), since irrespective of their nomenclature, both the positivistic and interpretivistic models underlying the audit trail parameters roughly correspond to each other even though they are not exact substitutes (Morrow, 2005; Rolfe, 2006).

Research questions

The quality of scholarly journals is usually judged based on their relative impact (Aguinis et al., 2012). An article in a top journal is highly likely to represent the best practices in the discipline. With the new criteria developed from the literature and pilot study, we propose to review the grounded theory articles in top management journals to evaluate their level of second-party auditability.

Our review intends to address the following research questions:

Research Question 1: How often does grounded theory articles in top management journals report second party audit parameters pertaining to pre data collection stage?

Research Question 2: How often does grounded theory articles in top management journals report second party audit parameters pertaining to data collection stage?

Research Question 3: How often does grounded theory articles in top management journals report second party audit parameters pertaining to data coding and analysis stage?

Research Question 4: How often does grounded theory articles in top management journals report second party audit parameters pertaining to presentation of results stage?

Research Question 5: How often does grounded theory articles in top management journals report second party audit parameters?

Coding and Data analysis

In a directed content analysis, the codes or initial coding categories are identified from prior literature or theory. The articles were content analyzed and consensually coded by multiple coders. The whole article, and not just the methodology part, was read, since methodological considerations in qualitative studies tend to not be limited to the method section, only (Gibbert et al., 2008). To ensure further standardization and reliability, after coding 10 articles the coders checked and calibrated the coding. To ensure a rigorous coding process, a consensus coding approach was followed: the coders discussed and agreed upon the most correct answer to a specific code. During instances of differences in the initial coding, the coders went back to the concerned papers and discuss until they reach consensus.

The coding sessions (of both the pilot and the main study) helped us to refine our coding sheet through successive iterations between the coding parameters and the data (Ryan and Bernard, 2000). Any text which could not be categorized using the preliminary coding sheet were identified, analyzed afterwards (Hsieh and Shannon, 2005), and represented as a new category or a subcategory of an existing code.

6.7. Results

Since our study sample was quite small, it precluded us from doing large scale quantitative analysis. Table 6.3. includes details of the frequency of distribution of parameters pertaining to each stage of the research process. The total number of

auditability parameters in our coding sheet is 30 (pre data collection stage: 3, data collection stage: 7, data coding and analysis: 10, and presentation of results: 10). Altogether, only 14.8% of the articles mentioned more than 75% of the overall second party audit trail parameters. Regarding individual stages, 40.7% articles discussed more than 75% of the pre data collection stage codes. 33.3% of the sample discussed 75% or more of the data collection parameters. 37% of articles analyzed discussed more than 75% of the data coding and analysis parameters. The lowest frequency of reporting was evident in the presentation of results stage. In this stage, only 7.4% of the articles reported 75% or more of the parameters.

Frequency of reporting					
Percentage of second party audit trail parameters met	<i>Pre data collection stage</i>	<i>Data collection stage</i>	<i>Data coding and analysis stage</i>	<i>Presentation of results stage</i>	<i>Overall second party auditability</i>
0-25%	3.7	3.7	14.8	0	0
26-50%	22.2	3.7	18.5	63	25.9
51-75%	33.3	59.3	29.6	29.6	59.3
>75%	40.7	33.3	37.0	7.4	14.8

Table 6.3. Reporting frequency of audit trail parameters pertaining to stages of research process

Table 6.4 shows the reporting frequency of individual audit trail parameters in our sample. As can be seen, the most frequently reported parameters are ‘explanation of new theory’ (100%), ‘sources of data used’ (96.3%), ‘data collection process’ (96.3%), and ‘theoretical implications’ (96.3%). The least reported ones are the ‘limitations’ (29.6%), ‘rigor parameters’: ‘credibility’ (14.8%), ‘dependability’ (18.5%), and ‘conformability’ (3.7%), ‘constant comparison’ (37%), and ‘theoretical saturation’ (33.3%).

	Percentage of articles
<i>Pre-data collection</i>	
1 Definition of research question	85.2
2 Rationale for choosing Grounded Theory	48.1
3 Nature of the study (Inductive/theory building, deductive/theory testing, abductive)	77.8
<i>Data collection</i>	
4 Rationale for research setting selection	88.9
5 Sampling technique (Theoretical sampling)	44.4
6 Context of data collection	85.2
7 Sources of data used	96.3
8 Triangulation (Data, researcher, theory)	40.7
9 Data collection process	96.3
10 Review by key informants	51.9
<i>Data coding and analysis</i>	
11 Memo/list of indicators (Events, incidents, actions)	44.4
12 Coding principles	74.1
13 List of categories derived from data and literature	81.5
14 Identification of core category in the emergent theory	85.2
15 Description of how the categories are linked to the core category	77.8
16 Diversity of data considered	70.4
17 Constant comparison	37.0
18 Theoretical saturation	33.3
19 Analysis of negative cases	40.7
20 Discussion of theoretical notion with team mates/peers	74.1
<i>Presentation of results</i>	
21 Using illustrations for explaining properties	88.9
22 Explanation of new theory	100
23 Theoretical implications	96.3
24 Rigor parameters	
*Credibility	14.8
*Dependability	18.5
*Conformability	3.7
*Transferability	44.4
25 Managerial implications	40.7
26 Limitations of study	29.6
27 Future directions	77.8

Table 6.4. Second-party audit trail in management research

6.8. Discussion

The iterative nature of a grounded theory study often makes it challenging to explicate the various steps involved and decisions made in a grounded theory study for the

‘customers’ of the article, and thereby make it difficult to ascertain the study’s overall rigor. One way to handle this problem would be to transparently report what steps were undertaken by the researcher at each stage of the research process, by leaving a clear second-party audit trail about how the researcher constantly compared the theoretically sampled incidents until reaching saturation, i.e. until no new, surprising, or negative incidents or category properties were found (Locke, 2001). In the present study, we examined the methodological sophistication of these procedures in management research. While existing research on the methodological sophistication of grounded theory was typically done from the perspective of the author or producer of the research, our perspective is customer-centric, both in terms of the end-customer (i.e. the reader or other author), as well as the intermediate customer (i.e. reviewers and editors). The former has less access to methodological residues than the latter, who can demand further clarifications and detail regarding the study prior to publication of an article. In particular, we drew on the audit literature to propose a perspective of ‘second-party audits’. This perspective is helpful in that it focuses attention on the concrete research actions which need to be transparently relayed in the final manuscript, so that an article’s chain of evidence can be conveniently assessed *ex post*.

Our results show that grounded theory articles in top management journals exhibit low second-party auditability: as Table 6.3 shows, majority of the grounded theory articles in management discuss less than 3/4th of the audit trail parameters considered, except in pre data collection stage. If we look at the individual codes, the least reported ones are the ones pertaining to explicit mentioning of certain rigor parameters such as ‘credibility’(14.8%), ‘dependability’(18.5%), ‘conformability’(3.7%), and ‘transferability’ (44.4%). This seems to mirror the trend in other qualitative research, where authors were found not to explicitly mention key terms such as validity and reliability (Gibbert and Ruigrok, 2010). It might be that authors are reluctant to subscribe to the interpretivistic ideology explicitly, to avoid debates in the review process with reviewers who are more familiar with the positivistic model.

The most important tendency we observed is that key parameters of grounded theory are completely underrepresented. For instance, ‘triangulation’ or using more than one strategy to collect, analyze, or interpret data is completely underrepresented. Triangulation, however, constitutes one of the most important criteria for ensuring the trustworthiness of a study. Only 40.7% of the articles reported triangulation. Another key parameter of grounded theory, ‘theoretical sampling’ was reported in only 44.4% of the articles. Theoretically sampling individual cases or units of analysis and constantly comparing them with other incidents or properties of a category, in terms of as many similarities and differences as possible until one reaches theoretical saturation (Glaser and Strauss, 1967) constitutes one of the cornerstones of grounded theory. The fact that almost half of the studies in top management journals do not transparently relay how they identified the very basis of their emerging theory, i.e. the cases on which the grounded theory is grounded is disquieting, since it might be indicative of sampling bias. While sampling bias is generally a problem, even for large-N studies which typically sample randomly, small-N studies such as grounded theory in particular are exposed to sampling bias stemming from selecting cases which might precipitate incorrect conclusions, since they inadequately over represent certain characteristics. Even more problematically, the percentage of articles discussing ‘negative cases’ is even lower: 40.7%. This again is quite contradictory to the core tenants of grounded theory. Iteration by definition involves comparison of theoretically sampled incidents (similar and dissimilar) with theory and other incidents. An iterative cycle starts with model mis-fit. That is, a case that does not fit the emerging theory prompts another cycle of data collection and theoretical refinement via constant comparison. Unless there are negative or dissimilar cases, iterations remain limited to one cycle only, defying the main procedural characteristics of grounded theory.

In this vein, as was to be expected given the low reporting of ‘theoretical sampling’ and ‘negative cases’, only 37% of the articles report ‘constant comparison’ explicitly, and ‘theoretical saturation’ was reported in only 33.3% of the articles. ‘Constant comparison’ and ‘theoretical saturation’ are the crux of a valid grounded theory study (Strauss and Corbin, 1998). A rigorous grounded theory study, through

iterative cycles of constant comparison, should reach a point where theorizing the events being studied reaches completion (Sandelowski, 2008), i.e. when the researchers are fully satisfied that the relationships and properties they are interested in are described in their full complexity and variation. It signals the end of the data analysis part of the study. Based on our results, it would appear that only in about a third of grounded theory articles, the emerging theory was actually saturated empirically. Put differently, in two-thirds of grounded theory articles published in top management journals, a second-party audit cannot establish to what extent the reported findings actually are conclusive.

6.9. Theoretical implications and future directions

Grounded theory studies, due to the lack of a common structure, often fail to sufficiently detail their research process. This in turn makes it difficult for ‘customers’ of grounded theory to fully understand how the results were empirically derived and what their theoretical contributions are (Kauffman and Denk, 2011). We used a second-party audit perspective to reveal a number of serious shortcomings in existing grounded theory research, including underrepresented key actions such as theoretical sampling, constant comparison, the integration of negative cases, as well as theoretical saturation. Using this second-party audit, we propose key parameters on an audit trail which grounded theory authors, as well as journals could use as a template or a boilerplate (Pratt, 2009) for enhancing the rigor and transparency. Indeed, our audit trail is not intended to make grounded theory more formulaic or overburdened with rules (Urquhart et al., 2010). Rather the intention is to create an agreeable and acceptable basis for management researchers to report grounded theory studies, based on how second-party audits would perceive the rigor of the underlying research procedures.

Our study shows that top management journals during the four decades (1970-2010) reported second party auditability parameters scarcely. On the other hand, grounded theory is becoming more and more popular in management research of late. In the past five years, the top management journals in our sample have published 20 grounded theory studies compared to the 27 they published during 1970 to 2010.

Examining the grounded theory articles during the past five years (2010-2015) would be the logical next step.

Since we are analyzing published grounded theory studies and not the original manuscripts, we realize that it might not be able to understand the level of auditability originally intended by the authors or the modifications that were made to the manuscript during the review process. Further ahead, we intend to interview journal editors and grounded theory authors to better understand these changes. Another major limitation of the study is the small sample size, which curbed the possibility of examining the influence of auditability on the reception of an article by the academic community (via citation count). Further studies could examine this relationship between second party auditability and scholarly impact, by using a bigger sample (all grounded theory articles in management during a particular time period).

6.10. Conclusion

As pointed out in an editorial by Suddaby (2006) in *Academy of Management Journal*, there is a dire need for transparent reporting of grounded theory research in management. Our study confirms the same by observing that reporting patterns in management grounded theory studies are to be further developed. However, all is not bad news. Some of the articles we examined were found to be very good at second party auditability. To conclude this study on a positive note, we would like to point to a best practice or exemplar grounded theory article. Table 6.5 gives an example of such an exemplar (Browning et al., 1995) which reported more than 81% of our auditability parameters.

While space constraints at this stage prevent us from a full discussion, Table 6.5 indicates in particular, that key parameters such as the inclusion of negative cases as a starting point for iterations was explicitly problematized. The selection of cases (both negative as well as positive) was theoretically motivated (i.e. they were sampled in a theoretically-informed way so as to avoid sampling bias), enabling constant comparison between empirical phenomena and their theoretical relevance until theoretical saturation was reached and the grounded theory was in fact grounded firmly in empirical realities.

	Y/N	Excerpt from the article	Page
<i>Pre-data collection</i>			
1	Y	<i>The current research is focused on rather different issues: how the SEMATECH organization emerged out of...semiconductor industry.</i>	115
2	Y	<i>...primarily because we aspired to derive new theoretical insights...unprecedented and unique effort in building co-operation.</i>	120
3	Y	<i>...the reported research was carried out as a qualitative inductive study...</i>	114
<i>Data collection</i>			
4	Y	<i>What is novel and theoretically interesting about SEMATECH...how cooperation can arise and persist in a competitive industry.</i>	114
5	Y	<i>As is appropriate in qualitative research, theoretical sampling was used....</i>	119
6	Y	<i>Research Setting: Background information on SEMATECH</i>	115-119
7	Y	<i>This article presents the results of a grounded theory analysis of observations, interview, and archival data....</i>	113
8	Y	<i>Our methods also permitted some within-method and between-methods triangulation...member companies.</i>	121-122
9	Y	<i>Interviews, which were private and.in 1992 and 1993.</i>	120
10	Y	<i>The few apparent discrepancies of fact that arouse were reconciled through additional interviews with the original informants involved. To further verify the accuracy of our statements and interpretations, we submitted this article to the standard document review process at SEMATECH.</i>	121
<i>Data coding and analysis</i>			
11	Y	<i>Appendix: Chronology of Important Events at SEMATECH</i>	149-151
12	Y	<i>We coded the transcripts using....described in the Results section.</i>	120-121
13	Y	<i>Table 3: Coding Theme and Core Categories</i>	124
14	Y	<i>Three core categories of events and behaviors are described...foster cooperation.</i>	113
15	Y	<i>Table 3: Coding Theme and Core Categories</i>	124
16	Y	<i>When new data yielded new or inconsistent information, conceptual categories and the emerging</i>	

17	Constant comparison	Y	<i>theory are modified to take them into account.</i>	121
18	Theoretical saturation	Y	<i>The process involved in the constant comparative method we used....</i>	121
19	Analysis of negative cases	Y	<i>This process was repeated until theoretical saturation is reached...tentative hypothesis is being generated.</i>	121
20	Discussion of theoretical notion with team mates/peers	Y	<i>When new data yielded new or inconsistent information, conceptual categories and the emerging theory are modified to take them into account.</i>	121
			<i>The remaining author participated in the selective coding stage, playing the role of questioner and devil's advocate.</i>	121
	Presentation of results			
21	Using illustrations for explaining properties	Y	<i>As Anna Bowers Noyce...commented "Bob just assumed a person would act on what needed to be done."</i>	128
22	Explanation of new theory	Y	<i>Our results suggest that the interdependencies involved in interorganizational arrangements can pay off...individual honor and pride.</i>	144
23	Theoretical implications	Y	<i>Section: Other implications on research and practice</i>	143-145
24	Rigor parameters (Credibility, Dependability, Conformability, Transferability)	N	None mentioned explicitly (But validity concerns addressed. Subsection: Validity. p.121)	NA
25	Managerial implications	Y	<i>Section: Other implications on research and practice</i>	143-145
26	Limitations of study	N	Not mentioned	NA
27	Future directions	N	Not mentioned	NA

Table 6.5. Second-party audit trails parameters on an exemplar article (Browning et al., 1995)

6.11. Appendix A: Pilot study

We used directed content analysis (Potter and Levine-Donnerstein, 1999; Hsieh and Shannon, 2005) and scrutinized all grounded theory articles from management journals, from the time period (2002-2005), using the sample by Jones and Noble (2007). The journals we consider are: *Organization Science*, *Journal of Management Studies*, *Long Range Planning*, *British Journal of Management* (2.70), and *Journal of Small Business Management*, and *Education+Training*, *Human Resource Development International*, *International Journal of Entrepreneurial Behaviour and Research*, *International Journal of Production Research*, *Journal of European Industrial Training*, *Journal of Health Organization and Management*, *Journal of Knowledge Management*, *Journal of Management Development*, *Journal of Organizational Change Management*, *Journal of Workplace Learning*, *Leadership & Organization Development Journal*, *Management Decision*, *Organizational Analysis*, *Project Management Journal*, *Qualitative Market Research: An international journal*, *Team Performance Management: An international journal*, *The Journal of Applied Behavioral Science*, and *Women in Management Review*. Additionally, we did a key word search using the term ‘grounded theory’. The preliminary descriptive statistics pointed out that all the codes which we have included in our coding sheet are in fact reported in management grounded theory research. This pilot study involved a total of 32 articles. The pilot study was mainly intended to validate the coding sheet and see the frequency of reporting of our second-party audit trail parameters. For the same reason, we did not formulate any specific research questions at this stage.

Analysis

Table A1 shows the reporting frequency of different codes in our preliminary coding sheet. In particular, the most widely reported parameters are definition of the research question (90.6%), data sources used (100%), list of categories derived from literature and data (90.6%), explanation of new theory (90.6%), and theoretical implications (90.6%). Likewise, when we consider the parameters, rigor factors (<15% altogether) were reported less than expected. We also found that the specific parameters and core analytic

tenants pertaining to grounded theory research were less reported than the general parameters.

	% of articles)
<i>Pre-data collection</i>	
Definition of research question	90.6
Rationale for choosing Grounded Theory	62.5
Nature of the study (Inductive/theory building, deductive/theory testing, abductive)	62.5
<i>Data collection</i>	
Rationale for research setting selection	78.1
Sampling technique (Theoretical sampling)	43.8
Context of data collection	62.5
Sources of data used	100
Triangulation (Data, researcher, theory)	25
Data collection process	71.9
Review by key informants	18.8
<i>Data coding and analysis</i>	
Memo/list of indicators (Events, incidents, actions)	34.4
Coding principles	62.5
List of categories derived from data and literature	90.6
Identification of core category in the emergent theory	78.1
Description of how the categories are linked to the core category	65.6
Diversity of data considered	65.6
Constant comparison	31.3
Theoretical saturation	25
Analysis of negative cases	15.6
Discussion of theoretical notion with team mates/peers	28.1
<i>Presentation of results</i>	
Using illustrations for explaining properties	87.5
Explanation of new theory	87.5
Theoretical implications	90.6
<i>Rigor parameters</i>	
*Credibility	6.3
*Dependability	12.5
*Conformability	3.1
*Transferability	31.3
Managerial implications	65.6
Limitations of study	53.1
Future directions	46.9

Table A1. Pilot study- Descriptives

Some articles in our pilot study were found to be not exactly grounded theory studies but rather were using the term ‘grounded theory’ quite loosely (Welch, Piekkari, Plakoyiannaki, and Paavilainen-Mäntymäki, 2010). For instance, we found that some articles using other qualitative methods like descriptive case studies were claiming to be grounded theory studies, without using grounded theory. In the main study, we control for this occurrence by focusing on grounded theory studies solely.

Another interesting tendency noted was that more papers tend to report the positivistic criterion of rigor (especially generalizability, the counterpart of transferability) than the interpretivistic criterion. 25% of the articles mention generalizability, whilst 6.3% mention reliability. These articles were coded ‘1’ for their respective interpretivist criterion.

PART THREE

WRITING

Chapter VII*

Judging by its title: The link between article title attributes and citation count in management research

Lakshmi Balachandran Nair

Prof. Michael Gibbert

Journal

Nair, L.B., & Gibbert, M. Scientometrics (Accepted for publication)

Conferences

Nair, L.B. (2016, August). 75th Academy of Management Annual Meeting, Anaheim, USA (Awaiting decision).

Nair, L.B. (2016, June). European Academy of Management Conference, Paris, France (Awaiting decision).

Nair, L.B. (2016). Western Academy of Management Conference, Portland, USA.

Nair, L.B. (2015). Western Academy of Management Conference, Hawaii, USA.

7.1. Abstract

What makes a “good” title for an article, i.e. one which attracts citations in the academic community? Answers to this question are manifold, though inconclusive across disciplines. In an attempt to provide cohesion, we integrate significant title characteristics from previous studies into a comprehensive model and link it with citation count. Keeping the disciplinary context constant, we focus on management science and find that only non-alpha numeric characters have a small, but significant negative effect on citation count. Surprisingly, attributes which tended to show significant effects in other disciplines (though often in opposite directions), such as length, structure, context, and linguistic attributes exhibited no relationship with citation count.

Key words: *Article title, citation count, scholarly impact, title attributes, management*

7.2. Introduction

The title of an academic article is easily one of its most important features (Subotic and Mukherjee, 2014). So what, specifically, makes a title “good”? – And, perhaps even more interestingly, what makes a “good title”? Regarding the first question, for academics, an article’s title should be the first (Paiva, Lima and Paiva, 2012), and most concise statement of its content (Yitzhaki, 2002). A good title presents a publication effectively to readers and captures their attention, thereby facilitating knowledge-flow (Jakobovits and Nahl-Jakobovits, 1987; Soler, 2011). Good titles might also be influential in making an article visible to practitioners and other stakeholders, or even to academics in other fields/areas of research (Stremersch, Verniers and Verhoef, 2007). In academia, an average researcher scans through approximately 1142 titles a year (Mabe and Armin, 2002). Coining an article title in an effective and attractive way (Soler, 2007) can therefore be instrumental in triggering the interest of readers and reviewers; making the title stand out, raising interest for the article and increasing citations (Paiva et al., 2012).

Naturally, as a first step, title characteristics might be more likely to influence downloads rather than citations. For instance, if a title looks informative or attractive users might click, download, and peruse the article in terms of its usefulness for the citing author.³ However, this only underscores the fundamental importance of the title for an article’s appreciation and eventual citation. We therefore chose to measure what makes an article’s title “good”, i.e. the visibility, appreciation, and attention given to an article as impact (citation count). Article impact (or simply, ‘impact’) constitutes one of the ‘strongest currencies’ (Aguinis, Suarez-Gonzalez, Lannelongue and Joo, 2012) in academia and is coincidentally also an influencer of the authors’ own impact and reputation (e.g., Cole and Cole, 1972; Garfield, 2006; Judge, Cable, Colbert, and Rynes, 2007). Gomez-Mejia and Balkin (1992) went as far as estimating the marginal dollar value of a single citation to articles in top-tier management journals at \$192 in 1988, with a future value of \$1,522 and a cumulative annuity of \$13,350 in 2011 (cited in

³ We would like to thank one of the reviewers for alerting us to the link between downloads and citations.

Judge et al., 2007: 492). Impact of scholarly work also plays a part in determining the allocation of resources and rewards to individuals and departments (Aguinis et al, 2012). So it quite literally “pays” for authors to pay attention to title characteristics and their potential implications for impact.

Clearly, then, article titles matter. What, however, makes a “good title”, i.e. one which attracts impact in the academic community? The answer to this question has so far produced equivocal evidence, in large part since vastly different characteristics of article titles were investigated in myriad disciplines (each with their own customs and traditions), from biomedicine (Lewison and Paraje, 2004), biology, medicine, physics (Lewison and Hartley, 2005), to marketing (Stremersch et al., 2007). To illustrate, authors from various disciplines have undertaken studies on article titles and their characteristic features such as the number of words and presence of colons (Lewison and Hartley, 2005), the relationship between title characteristics, downloads and citations in psychology (Subotic and Mukherjee, 2014), the relation of title length to the article length (Yitzhaki, 2002), and the role of titles in informing and attracting audience in information science (Diener, 1984). Some authors have even explored the relationship between various (and often eclectic) selections of article title attributes and citation counts. Buter and van Raan, for instance, studied the occurrence of non-alpha numeric characters and their correlation with citation count. Jacques and Sebire (2010) investigated the influence of article titles on impact. Jamali and Nikzad (2011) explored the type of titles and their relation with the number of article downloads and their subsequent citation rates. Paiva et al. (2012) explored the relationship between the citation count of articles and the length and format of their titles, as did Stremersch and colleagues (Stremersch et al., 2007).

Unfortunately, after decades of research, many of these studies have produced results which do not add up, and often appear outright contradictory. Thus, while there is some consensus on the question what makes a title “good” (i.e. its implications for article impact), there is much less clarity on our initial question of what makes a “good title”. As we see it, the main issue is that prior studies, perhaps due to the different

traditions and customs prevalent in each discipline, have tended to narrowly focus on a very limited set of title attributes (for instance, studied titular colons, focused only on title word count etc.).

The objective of the present study is to contribute to theory development on effective article titles by offering an overall model of article title characteristics combining significant title characteristics from diverse disciplines, and to ascertain the implications for article impact. We focus on one discipline, management science, as an application context. This limitation is voluntary in that it allows us, primarily, to reduce the complexity of this undertaking by keeping the discipline constant while building an integrative model of article characteristics from diverse studies in diverse contexts. Since theoretical development in the area of article titles is still in an early stage, we consider it important to gather relevant data in all major fields and disciplines before attempting to make interdisciplinary or inter-field generalizations. It is also voluntary, secondarily, in that it allows us to make a contribution to management science, which is conspicuously absent from the long list of disciplines which have probed article characteristics and their implications in the past. As such, management-specific examination seems appropriate at this stage and we chose to focus on top management outlets to investigate the relationship between relevant title attributes and their implications for impact.

7.3. Article title attributes

Several authors, from various disciplines, have undertaken studies on article titles, their characteristic features, and the link with impact (e.g. Anthony, 2001; Diener, 1984; Lewison and Hartley, 2005; Subotic and Mukherjee, 2014; Yitzhaki 2002).

Unfortunately, these results remain equivocal, and no general model of a “good” article title has emerged so far. In addition, the observations made in the literature regarding titles have so far ignored the effects of discipline and field variation (Anthony, 2001) and are thus difficult to generalize across. For instance, Habibzadeh and Yadollahie (2010) found that articles with longer titles have a higher citation count in general medicine, science and multidisciplinary journals (medicine, clinical science and science). Whereas, in Psychology, the opposite effect was found: Articles with shorter titles had a higher

citation count (Subotic and Mukherjee, 2014). We focus specifically on attributes which were studied or pointed out to have a significant relationship with citation count. Depending on their characteristics, we group them together as ‘length attributes’, ‘character attributes’, ‘structure attributes’, ‘context attributes’ and ‘linguistic attributes’.

Length attributes

Prior studies have explored length-related title attributes and how they have changed over time (Whissell, 2010), unfortunately with very limited cross-pollination and progress over the years and across disciplines, emphasizing the need for more recent and more comprehensive interdisciplinary studies on titles and their role in enhancing scholarly communication. Consider two studies in particular, which contradict each other. Diener (1984) calculated the change in the informational dynamics of physics and social science journal article titles, and found only a very small increase in the number of words, key words and the informational dynamics of journal article titles. Inversely, Lewison and Hartley (2005) examined the spatial (across disciplines and countries) and temporal (across 5 years) differences in article title length in science and cancer research articles. The findings show that there is an increase in length of article titles over time. The later study (Lewison and Hartley, 2005) does not cite the earlier study (Diener, 1984). One of the reasons the two studies do not add up might be due to the fact that Diener's (1984) article was conducted prior to the internet revolution, and Lewison and Hartley's work (2005) is more recent. As such, the disagreement between their findings could be due to the advent of search engines and online databases which tend to favor shorter titles.⁴ Another possible reason might lie in the disciplinary differences in the journals/articles. This apparent controversy surrounding the length of titles is not limited to these two studies, however. In fact, it seems endemic across disciplines. For instance, Paiva et.al. (2012) explored the relationship between the length of titles published in Public Library of Science (PLoS) and Biomed Central (BMC) journals in 2008. They find that articles with shorter titles have a higher citation count than articles with longer titles, and argue that shorter titles are more attractive and simpler to understand (Paiva et

⁴ We thank one of the reviewers for pointing this out.

al., 2012). The study by Subotic and Mukherjee (2014) also shows similar trends in the field of psychology. The results, however, were quite the reverse when Jacques and Sebire (2010) explored article titles (January- March, 2006), in five major general medicine articles. Their results, in fact, show a very strong positive association between title length and citation rate. In fact, the authors found that articles with the highest citation rates tend to have as many as twice the number of words in their titles than articles with the lowest citation rates. Similarly, Habibzadeh and Yadollahie (2010) found that in general medicine and multidisciplinary science journals, longer titles seemed to lead to higher number of citations.

Given these non-additive prior results when it comes to title length, formulating an a-priori expectation is not straightforward. One might argue that a shorter title would be more attractive and concise, signaling a clear focus and therefore making it stand out among the over one thousand articles an average researcher scans in the course of a year (Mabe and Armin, 2002). Other things being equal, though, it seems more plausible to expect that a longer article could provide the readers more information about the contents of the paper, and thus increase both initial attention and exposure, as well as its readership and subsequently its citation rate. Thus, building on the conceptualization of the length attribute by building on previously- attributes such as ‘number of words’ (Lewison and Hartley, 2005), ‘share of substantive words’ (Diener, 1984; Nagano, 2015), and ‘number of characters’ (Paiva et.al, 2012), we would expect:

Hypothesis 1: Length attributes have a significant positive influence on the citation count of top management journal articles.

Character attributes

On the one hand, Paiva et al. (2012) noticed that including a question mark, colon or a hyphen is usually concurrent with a lower citation count. Likewise, Michelson (1994) also observed that there is a significant inverse relationship between colons in titles and the status of journals in industrial relations. This could be because non-alphanumeric characters like question marks are now included by authors just to market their paper (Ball, 2009), and this awareness makes these papers less appealing to the audience.

However, according to Dillon (1982), ‘Titular colonicity [colons in titles] is the primary correlate of scholarly quality’. Buter and Van Raan (2011) find that 29 types of non-alphanumeric characters are utilized in article titles, the most common of them being hyphen, colon, comma, and the two parentheses. Jacques and Sebire (2010) also found that in general medicine articles, the presence of a colon in the title positively correlates with the citation count. Buter and Van Raan (2011) discovered that the presence of common non-alphanumeric characters have a small, but significant positive influence on the citation count. They explain that this could be due to the fact that titles without such characters appear ‘odd’ (Buter and Van Raan, 2011: 617) to the reader. The social sciences tend to have a higher percentage (33- 50%) of titles with colons (Hartley, 2007), so this variable is particularly relevant for our application context (management studies). Given the slightly more pronounced evidence for the negative relationship between titular colonicity and impact, we focus specifically on the variable ‘non-alphanumeric characters’ into our study and hypothesize:

Hypothesis 2: Character attributes have a significant negative influence on the citation count of top management journal articles.

Structure attributes

Titles usually occur either in the format of a single sentence or question, or as a compound sentence (Hartley, 2007b). Soler (2007) examined 570 biology and social science titles and found that compound sentences are more common in social science research papers. The study also provided evidence that interrogative formats are more common in titles of review papers in the social sciences. She reasons that compound constructions make the titles more clear, precise and informative.

Lewis and Hartley (2005) and Hartley (2005) examined the proportion of compound sentence formats. They differentiated titles according to the length of the two parts, separated by a colon (non-alpha numeric character). The different types include ‘short: long’, if the first part of the compounded sentence is shorter than the subsequent part (s), ‘long: short’, if the former part is longer than the latter, and ‘even’, if both parts are fairly balanced.

Previous authors have examined the influence certain other structure related attributes might have on citation count. Jamali and Nikzad (2011), for instance, studied the influence article title type has on the number of citations and downloads. They specify three types of article titles, ‘declarative’ (including the study’s main conclusions), ‘descriptive/neutral’ (describing the subject), and ‘interrogative’ (in a question format). The findings showed that articles with interrogative titles are downloaded more, but cited less than articles with descriptive or declarative titles. We consider these attributes (‘*declarative*’, ‘*descriptive*’, ‘*interrogative*’) under the categorical variable ‘title format’ for our study.

Furthermore, Paiva et.al. (2012) examined the classification of titles as ‘methods-describing’ and ‘results-describing’, and found that articles with results-describing titles are cited more often than the others. We include these attributes (‘*methods-describing*’ and ‘*results-describing*’) also into the study. We group these under the variable ‘title classification’. Finally, we also include a couple of features, ‘share of the substantive words’, and ‘share of non-informative words’ in this category. Considering the variety of title structures and their associations with the citation count, we postulate:

Hypothesis 3: Structure attributes have a significant positive influence on the citation count of top management journal articles.

Context attributes

We then examine whether title attributes providing specific contextual information influence citation count. Jaques and Sebire (2010) and Paiva et al. (2012) found that reference to a specific country/geographical location in the title predicts poor citation. This could be due to the fact that researchers may discount information which they perceive to be limited to a specific country (Jacques and Sebire, 2010). Presence of contextual attributes might also suggest limited generalizability and visibility (Paiva et al., 2012). Correspondingly, it seemed plausible to expect that ‘country/continent name’ in the titles would have a negative influence on the citation count of articles in management:

Hypothesis 4: Context attributes have a significant negative influence on the citation count of top management journal articles.

Linguistic attributes

Previously, the study by Jacques and Sebire (2010) found evidence that in medical journals, the citation count of articles positively correlates with the presence of acronyms in titles. Following up on that, we examine whether the inclusion or exclusion of ‘acronyms’, might have any direct association with citation count in management discipline. Furthermore, we include the variable ‘linguistic tools’ (constituting proverbs, metaphors etc.) and examine whether they are positively associated with citation count. To include psychological factors of a linguistic kind, we also include the variable ‘amusement’ in the study. We followed the Oxford dictionary definition used by Sagi and Yechiam (2008) and Subotic and Mukherjee (2014) in their studies: *Amuse: Cause someone to laugh and smile*. We call all these variables ‘linguistic attributes’ collectively:

Hypothesis 5: Linguistic attributes have a significant positive influence on the citation count of top management journal articles.

7.4. Methods

Data collection

Sample

We concentrated on five top tier management journals, following previous work rating the impact of management journals (Gomez-Mejia & Balkin, 1992; Podsakoff et al., 2005; Siggelkow, 2001; Tahai & Meyer, 1999). The journals are *Academy of Management Journal*, *Administrative Science Quarterly*, *Journal of Management*, *Organization Science*, and *Strategic Management Journal*. We specifically focused on this array of journals since they are consistently the top ones across various ratings, and as such are good indicators of the current trends in the academic community.

We looked into all types of articles (quantitative and qualitative field studies, methodology papers, secondary data empirical studies, meta-analyses, experimental studies, and papers using mathematical modeling or simulations etc.). We excluded editorials, book reviews, comments, and letters to editors since they might behave differently from regular research papers. We randomly sampled articles from the time period 1997-2006, to allow for meaningful analyses of article citation counts, given that some time is needed after publication of an article for citations to materialize (Walters, 2011). The total number of articles in these journals during the time period was 2597 (source: Business Source Premier). Our sample consisted of 553 titles (approximately 110 articles/journal), following suggestions by Cohen (1992), and Ferguson and Ketchen (1999), for meeting power requirements for statistical analyses.

Measures and Data Extraction

Data extraction followed the format of previous studies on article impact (Gibbert and Ruigrok, 2010). We considered article impact as our dependant variable and article title attributes as our independent variables. We also controlled for the journal of publication, author attributes, and article attributes.

Article impact

In this study, we use citation count for measuring article impact. In management, citation count is considered the most popular, objective, and standard metric for measuring impact and appraising the influence of a scientist's work on another (Bergh et al., 2005). Citation count is also the traditional and most frequently used method (Adam, 2002; Leung, 2007). Information on article citations was collected using ISI web of science and Google scholar. We used both these sources since they are different from each other in various aspects and considering them both for the study would ensure the internal and external validity of the study. For instance, Google scholar is more widely distributed when compared to ISI web of science. In addition, it involves citations not only in other academic articles, but also in student papers, dissertations, other non-scholarly sources etc. On the other hand, considering Google scholar alone would not be ideal either, since sometimes Google scholar inadvertently count a citation more than once. For example, it

might count a citation by the preprint and paper version of the same article twice (Meho and Yang, 2007). It might also provide phantom or false citations due its 'frequent inability to recognize real matches between cited and citing items claiming a match where there is not even minimal chemistry' (Jacsó, 2006; Meho and Yang, 2007: 2111). The citation data for all the articles was collected on the same day to prevent distortion or possible errors. For the analysis, we considered both the absolute number of citations and citations per year, leading to four outcome variables (ISI citation count, ISI citation count per year, Google scholar citation count, and Google scholar citation count per year).

Article title attributes.

To obtain the data on article title attributes, we manually extracted information on the features listed in Table 7.1. Since information on many of our attributes were extractable only by human coding (for example, the code pertaining to 'linguistic tools'), we used this technique instead of seeking the help of a bibliometric software. Likewise, for calculating the 'share of substantive words', we had to first ascertain the number of 'substantive' or 'significant' words by considering all the words in the title after discounting the articles, prepositions, conjunctions, pronouns and auxiliary verbs (Yitzaki, 2002). After manually coding the titles for the substantive word count, we calculated their share with respect to the overall number of words in the title.

<i>Length attributes</i>
Number of words
Number of substantive words
Number of non-informative words
Number of characters
<i>Character attributes</i>
Presence of non-alpha numeric characters
<i>Structure attributes</i>
Title structure
· Short: Long
· Long: Short
· Balanced
· Non compounded articles
Title format
· Declarative
· Descriptive
· Interrogative
Title classification
· Methods-describing
· Results-describing
· Neither
Share of substantive words
Share of non-informative words
<i>Context attributes</i>
Country/continent name
Company name
Industry name
<i>Linguistic attributes</i>
Acronyms
Linguistic tools

Table 7.1. Article title attributes

Controls

Considering the journal level policies regarding the word or character count of article titles, we presumed that the specific journal where an article is published might influence the title attributes. Hence we controlled for the journal in which articles have been published, by applying four dummy variables representing the journals in our sample (with *Administrative Science Quarterly* as the reference group). Furthermore, in our analyses regarding the relationship between article title attributes and impact, we included the variables used to examine antecedents of title attributes in all equations. These potential antecedents of title attributes included core author and article attributes which were found to have a significant relationship with citation count (Bergh et al., 2006; Conlon et al., 2006; Mingers & Xu, 2010; Stremersch et al., 2007).

Author attributes

We borrow the measure ‘number of authors’ from the Bergh et al. (2006) study on SMJ article impact, which shows that the number of authors negatively correlates with citation count. Lewison and Hartley (2005) also found that except when the number of coauthors is high, single authors produce more titles with colons than multiple authors, in science and oncology. Apart from these variables, we include, ‘continent of first author’ to control for geographical and cultural differences and their influences on article nomenclature.

Article attributes

Bergh and colleagues (2005) explored the effects of article characteristics on citation count and found that article age has a positive relationship with citation count. Article length was also identified as a predictor of article citation count. Yitzaki (2002) also observed that there is a moderate positive correlation between the length of a paper (number of pages) and the number of significant words in its title, in the sciences. Hence we include ‘article age’ and ‘article length’ into our study as controls. In addition, we consider ‘number of references’ as a control variable since academics who get cited might tend to ‘return the favor’ by citing the referrer’s article too, thereby influencing the citation count of the concerned article (Gilbert, 1977; Judge et al., 2007; Van Wesel, Wyatt, and ten Haff, 2014).

Coding and data analysis

A standardized coding scheme was developed from prior literature. For the variables with yes or no answers, we allotted a ‘0’ for absence of the concerned variable and a ‘1’ for presence. For example, a title with non-alphanumeric characters was given a ‘1’ for the concerned code and a title without the characters, a ‘0’. For nominal variables with more than 2 possible values, we assigned different numbers to different categories. Categorical variables with more than two categories were converted into dummy variables in the analysis and assigned an omitted reference category. For the variable ‘title format’ (Jamali and Nikzad, 2011), we gave the value ‘0’ to *declarative titles*, ‘1’

to *descriptive titles* and '2' to *interrogative titles*. The omitted category during regression analysis was *descriptive titles*. For the categorical variable 'title structure' (Hartley, 2005), we assigned '0' to *non-compounded*, '1' to *short: long*, '2' to *long: short*, and '3' to *balanced* structures. We omitted *non-compounded* during the regression analysis. Likewise, for 'title classification', we gave '0' to *methodology/design describing*, '1' to *results/conclusion describing* and '2' to titles in *neither* format. Here, the omitted categories were *results/conclusion describing* and *neither*, the reason for omitting the latter being the fact that only very few articles use this format (1.1%). We followed the same criteria for coding similar author and article attributes ('continent of first author', 'journal of publication' etc.). For variables involving different values like 'number of words', 'number of characters', 'article length' etc., we included the exact count in the coding sheet. For the variable 'amusement', we did the rating using Mturk. Each title was assessed by three native English speakers from USA. The judges used a seven point likert scale for rating the attribute. For this particular coder, the average rWG inter rater agreement was 0.83 (James, Demaree, and Wolf, 1993).

As illustrated earlier, many of our codes do not allow for a mechanistic coding or extraction via a software (for example, consider the code: title classification). Therefore, we manually coded all the titles in the sample. The titles were coded by three coders. Two of them were independent, blind coders who were neither involved in the study, nor aware of the research questions and lines of inquiry. The third coder was the first author of the present study. A consensus coding approach was followed. The pre-consensus coding inter-rater reliability was quite high for all codes (>90%). Subsequently, we did stepwise regression analyses using ordinary least squares model on the title attributes and citation count; with the journals, author and article attributes as controls.

7.5. Results

The descriptive statistics showed some noteworthy trends. The frequency of reporting of the variables are mentioned in Table 7.2. Tables 7.3 and 7.4 show the descriptives and the correlations.

Item reported		Frequency (%)
Compounded titles		69.3
Title structure	Not reported	30.6
	Short: Long	44.8
	Long: Short	18.8
	Balanced	5.8
Title format	Declarative	1.4
	Descriptive	92.9
	Interrogative	5.6
Title classification	Research methodology/design describing	83.5
	Results/Conclusion describing	15.4
	Neither	1.1
Non-alpha numeric characters		88.1
Acronyms		6.9
Linguistic tools		7.2
Country/continent name		8.3
Company name		2.2
Industry name		7.6

Table 7.2. Article title attributes: Frequencies

	Min.	Max.	Mean	SD
Number of words	2	23	11.75	3.78
Number of substantive words	2	19	8.13	2.44
Number of non-informative words	0	12	3.63	1.86
Number of characters	20	189	90.82	28.05
Share of substantive words	0.50	1.00	0.70	0.09
Share of non-informative words	0.00	0.86	0.30	0.10

Table 7.3. Article title attributes: Descriptives

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Article age	-																			
Article length	0.05																			
References	-0.07	0.54**																		
Number of authors	0.01	-0.01	0.02																	
Continent of first author	-0.04	-0.08	-0.02	-0.03																
Number of characters	-0.01	0.08	0.06	0.08	-0.01															
Non-alphanumeric characters	-0.05	-0.03	0.06	0.07	-0.07	0.26**														
Short: long structure	-0.01	0.17**	0.11*	-0.02	-0.05	0.25**	0.33**													
Long: short structure	-0.08	-0.09*	0.01	0.07	0.06	0.18**	0.16**	-0.43**												
Balanced structure	0.02	-0.04	0.01	0.03	-0.01	0.00	0.09*	-0.22**	-0.12**											
Declarative format	0.09*	0.13**	0.04	0.03	-0.06	0.01	0.00	0.04	-0.06	0.03										
Interrogative format	-0.04	-0.05	-0.06	0.02	-0.01	0.06	0.09*	0.05	0.04	-0.03	-0.03									
Methods/design specifying classification	-0.02	-0.02	-0.01	-0.07	-0.05	-0.09*	-0.10*	-0.02	-0.15**	-0.14**	-0.07	0.00								
Share of substantive words	-0.07	-0.12*	-0.06	-0.01	0.00	-0.18**	0.08	-0.07	0.02	0.05	-0.04	-0.08	-0.01							
Country/continent name in title	0.05	-0.04	-0.04	0.09*	0.10*	0.15**	0.07	0.03	0.07	0.01	0.07	-0.04	-0.04	-0.01						
Company name in title	0.00	0.05	0.01	0.00	0.00	0.02	0.05	0.02	0.06	0.07	-0.02	0.02	-0.07	-0.05	-0.04					
Industry name in title	-0.09*	0.12**	0.01	0.00	-0.03	0.13**	0.04	0.13**	-0.03	-0.01	0.02	-0.07	0.02	-0.01	0.19**	0.10*				
Acronyms	0.02	0.02	0.00	0.01	-0.01	0.09*	0.06	0.04	0.02	0.02	0.03	0.00	-0.05	0.05	0.33**	0.06	0.19**			
Linguistic tools	0.02	0.14**	0.10*	0.10*	0.03	0.07	0.10*	0.28**	-0.12**	-0.04	0.14**	-0.01	-0.05	-0.13**	-0.01	0.01	-0.03	0.03		
Amusement	-0.17**	0.06	0.06	-0.03	-0.02	-0.01	0.13**	0.19**	0.01	-0.06	-0.01	0.16**	0.02	-0.10*	-0.09*	-0.01	0.07	-0.03	0.27**	-

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 7.4. Correlations

Title attributes and citation count

We did a step-wise regression analysis. We included relevant indicators for variables from each category. For instance, from the category ‘length attributes’, we included the variable ‘number of characters’, since including all the variables together would have caused multicollinearity. We ran the analysis by first including all the different categories of attributes separately and then including them altogether in the same model. For the final step (including all variables), we checked for results with and without interaction effects. Each analysis thus had eight models, and we had a total of four analyses. The analysis was rerun with different dependent variables (ISI citations, ISI citations per year, Google scholar citations, and Google scholar citations per year). In all combinations, we found that the presence of non-alphanumeric characters in an article title has a significant negative relationship with citation count (see Tables 7.5-7.8). This confirms our second hypothesis which states that character attributes in titles have a significant influence on the citation count of top management journal articles.

In the analyses containing Google scholar citation count and Google scholar citation count per year, ‘number of characters’ was found to have a significant, negative relationship with citation count when analyzed in a model without the other variables. In the final model, however, ‘number of characters’ was found to have less significant negative relationships with citation count. This could be due to the influence of all the other variables (‘non-alphanumeric characters’ in particular). However, this relationship was not significant across the four dependent variable categories we examined. Therefore, our first hypothesis stating that length attributes have a significant influence on top management journal article citation count, was rejected.

		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
		Beta	Error														
Controls																	
<i>Journals</i>	Dummies for journals	Included	-														
<i>Article attributes</i>	Article age	0.18**	2.10	0.18**	2.10	0.18**	2.09	0.19**	2.12	0.18**	2.12	0.18**	2.14	0.17**	2.15	0.16**	2.20
	Article length	0.19**	1.25	0.18**	1.25	0.17**	1.25	0.19**	1.26	0.20**	1.27	0.19**	1.26	0.18**	1.28	0.18**	1.28
	References	0.00	0.25	0.01	0.25	0.02	0.25	0.01	0.25	0.00	0.25	0.00	0.25	0.01	0.25	0.00	0.25
<i>Author attributes</i>	Number of authors	-0.07+	5.57	-0.07+	5.56	-0.06+	5.54	-0.07	5.61	-0.07+	5.59	-0.07+	5.61	-0.06	5.64	-0.07+	5.64
	First author: Europe	-0.06	17.83	-0.07+	17.86	-0.07+	17.82	-0.06	18.01	-0.06	17.89	-0.06	17.88	-0.07+	18.03	-0.07	17.99
	First author: Asia	-0.06	25.58	-0.06	25.54	-0.06	25.44	-0.06	25.81	-0.06	26.11	-0.06	25.67	-0.07+	26.26	-0.07+	26.20
	First author: Oceania	-0.05	44.14	-0.05	44.06	-0.05	43.87	-0.04	44.40	-0.05	44.21	-0.05	44.36	-0.05	44.29	-0.05	44.17
Variables																	
<i>Length attributes</i>	Number of characters			-0.07+	0.22									-0.06	0.25	-0.06	0.25
<i>Character attributes</i>	Non-alphanumeric characters					-0.11**	19.12							-0.15**	23.12	-0.15**	23.06
<i>Structure attributes</i>	Short: long structure							-0.02	15.17					0.11+	19.01	0.11+	18.96
	Long: short structure							-0.05	18.81					0.05	21.72	0.06	21.72
	Balanced structure							0.04	28.83					0.09*	30.26	0.09*	30.20
	Declarative format							-0.02	53.00					-0.02	53.05	-0.02	52.95
	Interrogative format							0.06	27.08					0.07+	27.26	0.08+	27.48
	Method/design specifying classification							0.03	17.32					0.04	17.26	0.04	17.22
	Share of substantive words							0.02	66.62					0.03	68.69	0.03	68.51
<i>Context attributes</i>	Country/continent name in title									0.01	23.45			0.03	24.91	0.04	24.88
	Company name in title									-0.02	42.95			-0.02	43.27	-0.02	43.19
	Industry name in title									-0.05	24.25			-0.04	24.75	-0.04	24.69
<i>Linguistic attributes</i>	Acronym											-0.04	24.62	-0.03	26.31	-0.04	26.29
	Linguistic tools											0.03	26.08	0.03	26.74	0.03	26.67
	Amusement											-0.03	7.33	-0.03	7.54	-0.02	7.56
Interaction	Article age*Interrogative format															-0.08*	8.95
R squared			0.13		0.13		0.14		0.14		0.13		0.13		0.16		0.17
F			7.23		6.89		7.34		5.00		5.81		5.78		3.86		3.87

**p \leq .01, *p \leq .05, +p \leq .10

Table 7.5. Step-wise regression analysis (ISI citation count)

		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
		Beta	Error														
Controls																	
<i>Journals</i>	Dummies for journals	Included	-														
<i>Article attributes</i>	Article age	-0.02	0.16	-0.02	0.16	-0.02	0.16	-0.02	0.16	-0.03	0.16	-0.02	0.17	-0.03	0.17	-0.05	0.17
	Article length	0.20**	0.10	0.20**	0.10	0.18**	0.10	0.20**	0.10	0.22**	0.10	0.20**	0.10	0.20**	0.10	0.20**	0.10
	References	0.02	0.02	0.03	0.02	0.04	0.02	0.03	0.02	0.01	0.02	0.02	0.02	0.03	0.02	0.02	0.02
<i>Author attributes</i>	Number of authors	-0.09*	0.43	-0.08*	0.43	-0.08*	0.43	-0.08+	0.43	-0.09*	0.43	-0.09+	0.43	-0.08+	0.44	-0.08*	0.44
	First author: Europe	-0.06	1.38	-0.07	1.38	-0.07+	1.38	-0.06	1.39	-0.06	1.38	-0.06	1.38	-0.07	1.39	-0.06	1.39
	First author: Asia	-0.06	1.97	-0.06	1.97	-0.07	1.96	-0.06+	2.00	-0.07	2.01	-0.06	1.98	-0.08+	2.03	-0.07+	2.03
	First author: Oceania	-0.05	3.41	-0.05	3.40	-0.05	3.39	-0.05	3.43	-0.05	3.41	-0.06	3.43	-0.05	3.42	-0.05	3.42
Variables																	
<i>Length attributes</i>	Number of characters			-0.07+	0.02									-0.07	0.02	-0.07	0.02
<i>Character attributes</i>	Non-alphanumeric characters					-0.11**	1.48							-0.15**	1.79	-0.15**	1.78
<i>Structure attributes</i>	Short: long structure							-0.01	1.17					0.12+	1.47	0.12+	1.47
	Long: short structure							-0.04	1.45					0.06	1.68	0.07	1.68
	Balanced structure							0.03	2.23					0.09+	2.34	0.09+	2.34
	Declarative format							-0.02	4.10					-0.02	4.10	-0.02	4.10
	Interrogative format							0.04	2.09					0.05	2.11	0.06	2.13
	Method/design specifying classification							0.03	1.34					0.03	1.33	0.03	1.33
	Share of substantive words							0.03	5.15					0.03	5.31	0.03	5.30
<i>Context attributes</i>	Country/continent name in title									0.01	1.81			0.03	1.92	0.04	1.92
	Company name in title									-0.03	3.31			-0.02	3.34	-0.03	3.34
	Industry name in title									-0.07+	1.87			-0.06	1.91	-0.06	1.91
<i>Linguistic attributes</i>	Acronym											-0.03	1.90	-0.03	2.03	-0.03	2.03
	Linguistic tools											0.02	2.01	0.02	2.07	0.02	2.06
	Amusement											-0.03	0.57	-0.02	0.58	-0.02	0.58
Interaction	Article age*Interrogative format															-0.07	0.69
R squared		0.10		0.11		0.12		0.11		0.11		0.12		0.14		0.14	
F		5.70		5.52		5.88		3.66		4.73		4.54		3.17		3.15	

**p \leq .01, *p \leq .05, +p \leq .10

Table 7.6. Step-wise regression analysis (ISI citation count per year)

		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8		
		Beta	Error															
Controls																		
<i>Journals</i>	Dummies for journals	Included	-															
<i>Article attributes</i>	Article age	0.20**	7.09	0.21**	7.06	0.20**	7.05	0.21**	7.15	0.20**	7.15	0.20**	7.20	0.20**	7.25	0.18**	7.42	
	Article length	0.22**	4.23	0.21**	4.21	0.20**	4.23	0.22**	4.25	0.23**	4.27	0.22**	4.24	0.21**	4.30	0.21**	4.29	
	References	-0.03	0.83	-0.02	0.83	-0.01	0.83	-0.02	0.84	-0.03	0.84	-0.03	0.84	-0.02	0.84	-0.02	0.84	
<i>Author attributes</i>	Number of authors	-0.07+	18.78	-0.07+	18.73	-0.07	18.69	-0.07+	18.90	-0.07+	18.87	-0.08+	18.92	-0.07+	19.00	-0.07+	18.98	
	First author: Europe	-0.05	60.18	-0.06	60.12	-0.06	60.12	-0.05	60.70	-0.05	60.36	-0.05	60.23	-0.06	60.70	-0.06	60.61	
	First author: Asia	-0.06	86.32	-0.06	85.97	-0.06	85.84	-0.06	87.01	-0.06	88.08	-0.06	86.48	-0.07+	88.42	-0.07	88.24	
	First author: Oceania	-0.05	148.94	-0.05	148.31	-0.05	148.03	-0.04	149.64	-0.05	149.14	-0.05	149.46	-0.05	149.11	-0.05	148.77	
Variables																		
<i>Length attributes</i>	Number of characters			-0.10**	0.75									-0.08+	0.85	-0.09+	0.85	
<i>Character attributes</i>	Non-alphanumeric characters					-0.11**	64.53							-0.13**	77.84	-0.14**	77.66	
<i>Structure attributes</i>	Short: long structure							-0.04	51.13					0.09	64.00	0.08	63.88	
	Long: short structure							-0.06	63.40					0.04	73.14	0.05	73.16	
	Balanced structure							0.03	97.17					0.09+	101.89	0.09*	101.73	
	Declarative format							-0.03	178.63					-0.04	178.61	-0.03	178.34	
	Interrogative format							0.06	91.26					0.07	91.78	0.08+	92.57	
	Method/design specifying classification							0.03	58.38					0.03	58.13	0.03	58.00	
	Share of substantive words							0.04	224.54					0.04	231.26	0.04	230.76	
<i>Context attributes</i>	Country/continent name in title									-0.01	79.10			0.03	83.88	0.03	83.79	
	Company name in title									-0.02	144.88			-0.01	145.68	-0.02	145.47	
	Industry name in title									-0.06	81.81			-0.04	83.35	-0.04	83.16	
<i>Linguistic attributes</i>	Acronym											-0.05	82.95	-0.04	88.57	-0.05	88.55	
	Linguistic tools											0.04	87.89	0.04	90.03	0.04	89.83	
	Amusement											-0.05	24.70	-0.05	25.38	-0.04	25.46	
Interaction	Article age*Interrogative format															-0.08+	30.16	
R squared		0.12		0.13		0.14		0.13		0.13		0.13		0.13		0.16		0.16
F		6.87		6.83		7.01		4.56		5.56		5.63		5.63		3.81		3.81

**p \leq .01, *p \leq .05, +p \leq .10

Table 7.7. Step-wise regression analysis (Google scholar citation count)

		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
		Beta	Error														
Controls																	
<i>Journals</i>	Dummies for journals	Included	-														
<i>Article attributes</i>	Article age	0.03	0.54	0.03	0.54	0.03	0.54	0.03	0.55	0.02	0.54	0.02	0.55	0.02	0.55	0.00	0.57
	Article length	0.23**	0.32	0.23**	0.32	0.21**	0.32	0.23**	0.32	0.25**	0.33	0.24**	0.32	0.23**	0.33	0.23**	0.33
	References	-0.01	0.06	0.00	0.06	0.00	0.06	0.00	0.06	-0.02	0.06	-0.01	0.06	0.00	0.06	-0.01	0.06
<i>Author attributes</i>	Number of authors	-0.09*	1.43	-0.08*	1.43	-0.08*	1.43	-0.09*	1.44	-0.09*	1.44	-0.10*	1.44	-0.08*	1.45	-0.09*	1.45
	First author: Europe	-0.06	4.59	-0.07	4.58	-0.07+	4.58	-0.06	4.63	-0.06	4.60	-0.05	4.60	-0.07	4.63	-0.06	4.63
	First author: Asia	-0.07	6.58	-0.06	6.55	-0.07+	6.54	-0.06+	6.64	-0.07	6.71	-0.07+	6.60	-0.08+	6.74	-0.08+	6.73
	First author: Oceania	-0.05	11.36	-0.05	11.30	-0.05	11.29	-0.05	11.42	-0.05	11.36	-0.06	11.40	-0.05	11.37	-0.05	11.35
Variables																	
<i>Length attributes</i>	Number of characters			-0.11**	0.06									-0.09*	0.07	-0.09*	0.06
<i>Character attributes</i>	Non-alphanumeric characters					-0.12**	4.92							-0.14**	5.93	-0.14**	5.93
<i>Structure attributes</i>	Short: long structure							-0.04	3.90					0.10	4.88	0.10	4.88
	Long: short structure							-0.06	4.84					0.05	5.58	0.06	5.58
	Balanced structure							0.03	7.42					0.08+	7.77	0.09+	7.76
	Declarative format							-0.03	13.64					-0.03	13.62	-0.03	13.61
	Interrogative format							0.04	6.97					0.05	7.00	0.06	7.07
	Method/design specifying classification							0.03	4.46					0.03	4.43	0.03	4.43
	Share of substantive words							0.04	17.14					0.04	17.63	0.04	17.61
<i>Context attributes</i>	Country/continent name in title									0.00	6.02			0.03	6.39	0.03	6.39
	Company name in title									-0.03	11.03			-0.02	11.11	-0.02	11.10
	Industry name in title									-0.07+	6.23			-0.06	6.35	-0.06+	6.35
<i>Linguistic attributes</i>	Acronym											-0.05	6.33	-0.04	6.75	-0.04	6.76
	Linguistic tools											0.03	6.71	0.03	6.86	0.03	6.86
	Amusement											-0.05	1.88	-0.04	1.94	-0.04	1.94
Interaction	Article age*Interrogative format															-0.06	2.30
	R squared		0.09		0.10		0.11		0.10		0.10		0.10		0.13		0.13
	F		5.00		5.18		5.29		3.34		4.21		4.12		3.03		3.00

**p ≤ .01, *p ≤ .05, +p ≤ .10

Table 7.8. Step-wise regression analysis (Google scholar citation count per year)

In all the four analyses, certain structure attributes were found to have feebly significant ($p < .1$) relationships with citation count which was not consistent across models or across analyses. For example, in the case of ISI citation count, both the 'short: long' and 'interrogative' formats had weakly significant, positive relationships with the dependent variable, in the final two models. Likewise, in the case of Google scholar, 'balanced format' had a significant positive relationship with citation count in the final model. However, these relationships were neither stable, nor sufficiently significant. For this reason, our third hypothesis which states that structure attributes have a significant influence on article citation count in management, was rejected.

Regarding context attributes, 'industry name in titles' was found to have weakly significant ($p < .1$) relationships with Google scholar and ISI citation count per year. However, neither this nor the other variables in this category had any strongly significant or steady influence on citation count as per our four analyses. Hence, our fourth hypothesis about the significant influence of context attributes on citation count in management journal articles was also rejected. Finally, we found that none of the linguistic attributes we examined had any significant relationships with citation count in any of the analyses. Hence our final hypothesis stating linguistic attributes have a significant positive effect on citation count was rejected, as well.

Additionally, we checked for interaction effects between variables that might systematically influence the title attributes-scholarly impact relationship. We tested for potential interactions with 'article age' since this variable was found to have influence on citation count in prior studies (Bergh et al., 2006). This analysis revealed one significant interaction effect for article age with the presence or absence of a title's 'interrogative format'. This interaction was noticed in the case of both ISI and Google scholar

citation count, even though there were miniscule differences in the magnitude of the revealed results.

To further explore the nature of the detected interaction effect we probed the simple slopes, revealing a significant positive relationship between article title attributes and article impact for high article age, while there was no significant relationship for articles of low article age. The plots of these simple slopes are shown in Figure 7.1.

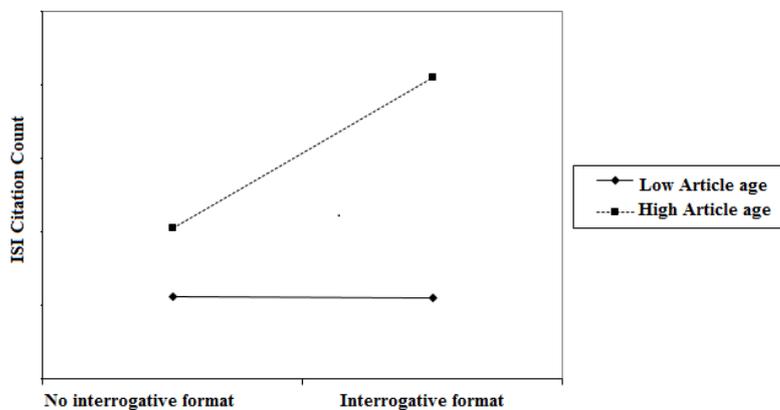


Figure 7.1. Plot of interaction effect

To ensure the rigor of our overall results, we performed robustness checks by exchanging variables from each category, with others from the same category (for example, in one of the robustness checks, we exchanged the variable ‘number of characters’, with ‘number of words’). The results of the robustness checks were consistent with our original results. The results observed refute hypotheses 1, 3, 4 and 5, which respectively stated that length, structure, context and linguistic attributes in titles have a significant influence on the citation count of top management journal articles.

7.6. Discussion

The question, what makes a “good title”, i.e. one which produces scholarly impact via citations in the academic community, has so far not provided

conclusive results. Our study is a first step towards building an overall model of article characteristics and linking them with impact. At this early stage of theory development on article titles, we do this in a specific application context (management science), in order to keep the discipline constant, reducing unnecessary variation across disciplines. Focusing on five of the top journals in the management discipline, our results show that certain title attributes do indeed predict the article citation count. Our step-wise regression analyses showed that the article title attribute ‘non-alphanumeric characters’ has a significant negative relationship with ISI and Google scholar citation counts and citation counts per year; thus confirming our hypothesis 2. This corroborates the findings by Jamali and Nikzad (2011) and Paiva et al (2012), in the field of biology and life sciences and Michelson (1994), in the field of industrial relations. On the other hand, it contradicts the findings by Jacques and Sebire (2010) in the field of general medicine and the interdisciplinary study by Buter and Van Raan (2011) both of which suggested a positive relationship between non-alphanumeric characters and citation count.

A plausible reason for this relationship between the title attribute ‘presence of non-alphanumeric characters’ and citation count could be that non-alphanumeric characters such as colons, question marks etc. usually denote complex, distinctive titles (Dillon, 1981). These titles used to be quite popular in the past, due to their containing more keywords, and therefore being more ‘findable and visible’ in databases (Moore, 2010). However, the arrival of search engines decreased their relevance. Search engines nowadays allow scanning keywords, abstracts, and often the full text of papers (Rostami et al., 2014), reducing the need for long and informative titles, and, coincidentally, also reduce the need for non-alphanumeric characters. In fact, they seem to propagate the use of titles which are simpler to understand, play the role of ‘interest-grabbers’, and are ultimately more attractive (Paiva et al., 2012). Another reason for the non-alphanumeric characters predicting low

citation count could be that some of these characters might be used as a kind of marketing strategy to win the attention of the readers (Ball, 2009), which, however, might suggest a lack of credibility and even frivolity (Fox and Burns, 2015), without accurately relaying an article's subject matter (Aleixandre-Benavent, Montalt-Resurrecció, and Valderrama-Zurián, 2014).

Hypotheses 1, 3, 4, and 5 were rejected. Length, structure, context and linguistic attributes were found to have no relationship with citation count. This contradicts many prior studies, where an increase in these attributes (for instance, length attributes) were found to have either a positive (Habibzadeh and Yadollahie, 2010; Jacques and Sebire, 2010) or negative effect (Jamali and Nikzad, 2011; Moore, 2010; Paiva et al., 2012) on the citation count. Thus, isolated effects from previous studies which focused narrowly on individual variables did not materialize in our application context, where significant attributes were examined together, and where individual effects cancel each other out. As an instance of this is, we found that certain attributes, in particular 'number of characters' tended to display significant relationships with citation count when entered into individual models. However, when entered into our final model, along with all the other attributes, 'number of characters' no longer exhibited any significant relationship with citation count. Thus, including more specific variables, in particular 'non-alpha numeric characters' helped us to better ascertain the actual relationships between relevant title attributes and citation count. Overall, then, a full model of article title such as ours represents a first in theory development delineating just what makes a "good title". Prior studies, which concentrated narrowly on specific title characteristics, without putting them in the context of other important attributes might lead to erroneous conclusions.

The study has a number of theoretical implications, both for scientometrics more generally, as well as for the application context, management science, specifically. Regarding the first field of study, our

model makes a first step towards exploring the influence of relevant title aspects, allowing us to observe how various title attributes contribute to citation count, in the presence of other factors. Including significant attributes from prior scientific literature into one model brings together different, and so far largely separate, lines of research focusing on article title attributes. Since a title often displays multiple attributes simultaneously, analyzing their collective implications represents a starting point for advancing theory development on title consequences. In this way, this study contributes to other scientometric studies on titles. Especially, our findings regarding the (negative) influence of non-alpha numeric characters on titles confirm the findings by certain prior researchers in the fields of medicine, physics (Ball, 2009), life sciences (Ball, 2009; Paiva et al., 2012.), and industrial relations (Michelson, 1994), whilst disconfirming the findings in the fields of chemistry, etymology (Buter and Van Raan, 2011) and general medicine (Jaques and Sebire, 2010). As can be observed, most of the application contexts explored so far are in life sciences. The choice of management as our setting allows us to add an additional context to the field of scientometric research on titles. Further studies could build upon our findings and examine more title aspects, in other application contexts, or even across disciplinary contexts.

Several disciplines have different title constructions and discourse conventions (Anthony, 2001), which can partly explain the observed inconclusiveness in the results from prior studies. For instance, in medicine, using full sentences represents the most widely-used article title structure (Soler, 2007). In management, as our results demonstrate, the most common title structure is compounded. Only 30.6% of the titles in management were not compounded, i.e. were full sentences. Hence what influences citation count of journals in one discipline (medicine) might not have a similar effect in the latter discipline (management), thus providing additional leads for further research.

Regarding management science specifically, scholarly impact is often defined and measured in terms of the number of citation counts an article receives (Adler and Harzing, 2009; Aguinis et al., 2012). Several articles have examined the relationship between various aspects of a scientific article and its subsequent citation count (e.g., Judge et al., 2007). However, not many studies have examined the influence of article titles and their attributes and how they relate to scholarly impact. Through this study, we examine the association of academic article title attributes with their scholarly impact in terms of citation counts. While four of the five proposed relationships did not materialize, our study did find evidence that management article titles without non-alpha numeric characters have a higher citation count than titles including them.

Interestingly, the specific application context (management science) and the idiosyncratic attributes we considered (industry name, company name) did not show any implications on the citation count of management articles. However, the frequency of these attributes in our sample was quite small (7.6% and 2.2% respectively). Further studies involving titles with additional context-relevant attributes could examine whether this effect holds even when these attributes are present in higher frequency. Due to field and discipline specificity and the scarcity of findings (Subotic and Mukherjee, 2014), management scholars often find it difficult to construct a ‘good’ title which is an informative and attractive short version of the article. There are several guidelines put forth by management journals which suggest titles of a particular length or format. However, those guidelines are not based on scientific data (Paiva et al., 2012). This study would thus be the first to offer some recommendations regarding the effective composition of management article titles. We provide an empirically grounded starting point for developing appropriate standards for formulating titles in management research.

7.7. Limitations and conclusion

One of the main limitations of our study is that we focused on article titles of one decade only (1997-2006). Also, our sample size is quite small considering the number of variables involved. Future studies could address this limitation by using a bigger sample. Furthermore, we did not perform any temporal comparisons of the trends, to see if there were significant changes in the titular practices over different time periods. Secondly, our sample consists of top management journals only. Including journals which are at the lower end of management scholarship would have facilitated a more detailed comparison. Additionally, we focus on article title attributes for explaining the citation count of articles. Citation count of articles can be due to several factors including the transparency, rigor, content, theoretical contributions etc. of the published articles. However, since prior studies have shown that titles do exert a significant, though small influence on an article's reception in the academic community, we limit this study to the characteristics of titles only. In addition, even though our study considered a wide variety of article title attributes, it is still not an exhaustive study of all possible title aspects out there. Future research could have a look at new, unexplored title attributes and their relationship with scholarly impact.

This study could be a starting point towards building a general model of article characteristics and its implications for impact. Ours was an attempt to describe recent trends with respect to management article titles, analyze the relationship between title aspects and citation count, and account for the relationship between various precursors of title aspects and the aspects themselves. Clearly, using a 'good title' is not the only strategy to increase citations. But an optimized title would help an article gain up-front attention and interest (Moore, 2010), and this study makes the case for (and take the first steps in) enhancing the visibility of an article.

Chapter VIII

Conclusion

Lakshmi Balachandran Nair

8.1. General summary of the findings

Many new developments have happened in qualitative research methods in the last twenty five years. From being prominent only in a few disciplines like anthropology and sociology, qualitative methods have now moved to become increasingly popular in various new disciplines, venues, perspectives, theories, and problem areas (Altheide and Johnson, 1994). This increasing popularity of qualitative methods in new arenas also brought with it a multitude of criticisms regarding its trustworthiness, relevance, and importance. Qualitative research has been questioned regarding its conceptual and analytical capabilities for demonstrating the rigor, plausibility, and defensibility of conclusions (Gioia, Corley, and Hamilton, 2013).

Several studies in several disciplines have pondered upon the rigor parameters of qualitative research. However, most of the studies which focus on qualitative research methods have solely focused on its generalizability or transferability (or lack thereof). Nonetheless, generalizability is not the only criterion for measuring the quality of a scientific inquiry. In particular, an inquiry's potential to build or enhance a rigorous (internally valid) theory and present it comprehensibly to the audience has not been explored before in management research. We identify a gap here. Therefore in this thesis, we focus on the process, problems, and solutions involved in investigating and communicating (Altheide and Johnson, 1994) a phenomenon in a rigorous, transparent and interesting manner.

Most of the studies in this thesis are concerned with qualitative research methods. However, it is impossible to talk about qualitative research methods completely separated from or as being completely different from quantitative methods. Therefore, we do not exclude quantitative research methods altogether from this dissertation. Rather, the individual studies included in this dissertation jointly demonstrate how qualitative research

methods (individually, or in conjunction with quantitative methods) contribute to new and rigorous idea/theory generation in the field of management, and how to present them in a way which ensures maximum readership and scientific impact. By doing so, we suggest new ways for authors and gatekeepers of management academia to ensure a credible and scientific dialogue in empirical management research, which in turn can inform management practice (Eden, 2002). Chapters III, V, and VII thus pertain to management research methods in general, albeit with a keener focus on qualitative inquiry.

Likewise, it is equally impossible to assume all the methods within the qualitative paradigm are uniform (Rolfe, 2006). Therefore, three of the studies in the thesis are dedicated to specific research methods in the qualitative research paradigm : Chapter II on case study research, Chapter IV on fsQCA (post-analysis) and Chapter VI on grounded theory research. We discuss the findings and conclusions of the individual chapters in the subsequent paragraphs.

Chapter I gives a short introduction to the thesis. We cover the three parameters of scientifically impactful research: idea, rigor, and writing. The thesis is divided into three parts. Part One of the thesis (Chapters II, III, and IV) elaborates on the parameters of ‘idea’ and ‘rigor’, within the framework of replication logic. Chapter II discusses the potential of replication logic in ensuring internal and external validity of a multiple case study design. We notice that replication logic is under-utilized as a technique for rigorous theory enhancement. We empirically differentiate three types of replication logic, with different degrees of theoretical and methodological sophistication. Then we build on prior studies in management case study research which focus on external validity (Yin, 2003; Eisenhardt 1989), and discusses replication logic in the context of a so far neglected rigor criteria: internal validity. We find evidence that replication logic designs, especially the most

sophisticated theoretical replication design enhances the internal validity and, as a favorable side-product, the external validity of the study.

In Chapter III, we concentrate on analyzing deviant cases or outliers identified during the course of a qualitative or quantitative study. We provide a definition of outliers and discuss why analyzing certain outliers contribute to developing theory that is more internally and externally valid. Subsequently, we infer three concrete strategies for analyzing deviant cases, based on replication logic. We term these strategies collectively as ‘Deviant Case Method’ (‘DCM’) and discuss their relative methodological sophistication with reference to published articles in top management journals. We provide exemplars which show that analyzing deviant cases identified during the course of a study (qualitatively or quantitatively) using the three DCM strategies provide a widely applicable approach for theory building.

In Chapter IV, we combine DCM with ‘most similar systems design’ and ‘most different system design’ (Faure, 1994; Levi-Faur, 2006) adapted from comparative politics, to look into the deviant cases identified during the course of fsQCA studies. We term this design the ‘Comparative Outlier Analysis’ (‘COA’). We observe that post-QCA deviant case analysis opens up several venues for new theory development. We describe how COA is a perfectly supplementary and at the same time, complementary design for QCA studies. We also discuss its potential in enhancing the generality, accuracy, and coverage of the concerned study. This is the last study in the thesis which applies replication logic or its derivatives.

Part Two of the thesis focuses on ‘rigor’, and ‘writing’. Specifically, in this part we focus on the ‘transparency’ and ‘auditability’ of writing (with a brief discussion of its implications for the replicability of the concerned study). Even though prior studies have examined the influence of ‘writing’ in scientific impact, they have not specifically examined the influence of

‘transparency’ and ‘auditability’ in enhancing and maintaining impact. Therefore, Chapter V reviews qualitative and quantitative field studies across seven major management journals, develop a transparency index, and link it to article impact. Our findings show that transparency increases impact across both methods. On comparison, quantitative studies were found to be more transparent than qualitative studies, although key transparency indicators were underutilized in both methods.

Chapter VI is a follow-up of the previous chapter. It adapts the transparency framework from Chapter V and builds an audit trail, specifically designed for grounded theory studies. After doing a preliminary analysis to ensure the applicability of the audit trail parameters, we examine published grounded theory studies in management across four decades (1970-2010). As a next step, we would empirically examine the adherence of grounded theory articles to audit trail, in top management journals across a more recent time period (2010-2015). Using the research questions and findings of this chapter as a template, we will formulate hypotheses regarding the audit trail tendencies in management grounded theory research, and their influence on scholarly impact.

Part Three of the thesis focuses on the influence of another ‘writing’ factor on enhancing citation count. We investigate various article title characteristics and how they relate to the article citation rate of academic articles in management. Authors in various disciplines (biology, life sciences, psychology etc.) have examined the importance of titles and some have even examined titles’ influence on scholarly impact. However, no such study has been conducted in management so far. In Chapter VII, we build on previous studies on titles from other disciplines, and do a comprehensive study of management article titles exclusively, in an attempt to understand what drives article impact in management. We first discuss the different types of article attributes (length, character, structure, context and linguistic attributes). Then we examine whether certain title attributes do indeed

predict the article citation count. Our findings show that the character attribute 'presence of non-alphanumeric characters' has a significant negative relationship with ISI and Google scholar citation counts and citation counts per year.

8.2. Implications of the findings and suggestions for future research

Management, as a discipline, is still comparatively young (Smith and Hitt, 2005) when compared to like sociology, psychology, and political science. It still has the scope to examine and learn from other, more mature disciplines. This is true especially with regards to methodological advancements, and in particular, qualitative research methods. Several management scholars, even today, feel that the standards followed by qualitative management research, are not high enough to demonstrate scientific advancement when compared to its counterparts (Campbell, 1975; Gioia et al., 2013).

In this dissertation, we adapt a multidisciplinary and eclectic stance to examine various aspects of impactful theory building in the light of empirical evidence. We adapt from other disciplines with state-of-the-art methodologies, as well as from other research methods and paradigms within the discipline, and contribute to the reconciliation of qualitative research with the conflicting demands of the scientific tradition (Gioia et al., 2013) in management research. Thus we contribute to theory building in management.

To explicate, Chapter II borrows the theoretical lens offered by positivistic paradigms to explore the relationship between the replication logic design undertaken and the internal and external validity parameters addressed in management multiple case study research. By primarily focusing on internal validity in the light of replication logic, Chapter II unravels the importance one of the most important rigor parameters for theory building. Without internal validity, it is not possible to build a theory with certainty about its causal mechanisms. Without certainty about casual mechanisms in a study, it is not possible to replicate it or generalize it to

other scenarios. Thus the main implication of this chapter is to establish internal validity as the most important rigor parameter in case study research (any research design, so as to speak). Also, by building on previous studies examining replication logic, we provide evidence for the importance of replication logic in ensuring internal validity in multiple case study research. This is a completely new direction from prior studies, which focused on replication logic only from an external validity point of view. A plausible follow-up study would examine the internal and external validity potential of single, embedded case studies.

Chapters III and IV discuss theory building through analysis of outliers or deviant cases. As Hambrick (2005) notes, we observe phenomenon first and notice puzzles. By thinking of ways to solve these puzzling observations or outliers, theory development is triggered. In Chapter III, we focus on the theory building potential of deviant cases identified during the course of qualitative and quantitative field research in management. We observe that fields as diverse as biology (Hagstrum 2013), comparative politics (Emigh 1997, Gerring 2007, Lieberman 2005), health care (Mays and Pope 2000), law (Gordon 1947), and criminology (Sullivan 2011) have identified the potential of deviant cases in theory enhancement. Such a constructive analysis of deviant cases is lacking in management discipline. To fill this gap, we offer Deviant Case Method for theory building in management, with suitable evidence from published articles regarding their rigor and theory building potential.

Likewise, in Chapter IV, we combine the most and least likely designs from comparative politics and combine them with our replication logic based deviant case analysis techniques to come up with Comparative Outlier Analysis. Our exemplar cases in this chapter are from politics, marketing and education. Through these two chapters we explicate the importance of deviant case analysis in theory building. It has a wide range of implications for management research. Deviant cases often point to a new

variable, or causal mechanism which led to the deviancy of the said case. By examining such cases we not only learn about this new variable or mechanism, but also about the typical cases in a better way.

A follow-up study of Chapters III and IV would be to examine outlier analysis tendencies beyond published papers. Articles represent the final product of a research process, and as such capture only a fraction of the procedures, considerations, and practices which lead to the practice of handling outliers as reported in the write-up which gets published. A rigorous engagement with outliers therefore requires going straight to the desks of researchers, and this is the approach and objective of the third sub-project. Thus, whereas Chapters III and IV rely exclusively on the outlier-handling strategies as reported in the published articles, the follow up study we envision would go beyond the published article and intends to probe the upstream processes which eventually led to the methodological procedures reported in the final paper.

Chapter V and VI discusses transparency of field research and auditability of grounded theory studies in management. Transparency of research procedures are quite important, not only in the early pre-publication stages, but also during and after the publication. Psychology and medicine, among other fields, have well-developed transparency parameters for academic articles (e.g., APA, 2008; Hancock & Mueller, 2010; Kilkenny, Browne, Cuthill, Emerson, & Altman, 2010; Schulz, Altman, & Moher, 2010). In Chapter V, we attempt to do the same in management (qualitative and quantitative) research.

In Chapter VI, we build upon the operationalization of transparency and study the auditability of grounded theory articles. We borrow the concepts of first, second, and third party audits from the field of accounting. We bring forth auditability parameters from prior literature and see the extent to which they are reported in grounded theory studies. Together, Chapters V

and VI offer a framework of reporting parameters for management researchers to follow and use as a foundation for future studies.

A logical follow-up of these two chapters would be to examine the transparency of specific methods. Since different methods under the same paradigm may exhibit different characteristics (for example, case study research and ethnography), this follow-up study would help further refine our transparency index and make it even more specialized for the specific method in hand. Furthermore, understanding the role of authors, reviewers, and editors in influencing the transparency of a manuscript would also be an interesting next-step to take.

Chapter VII, the final empirical chapter in thesis, explores the influence of title characteristics in citation count. For this study, we draw from prior literature in multiple fields as diverse as computer science, medicine, biology, psychology, agriculture etc. (Anthony, 2001; Lewison and Hartley, 2005; Subotic and Mukherjee, 2014; Yitzhaki 2002; Diener, 1984). We incorporate attributes from previous studies in these fields, to examine how titles influence management scholarly impact. By examining published articles in management journals, we provide evidence for our assertion that certain title attributes do indeed facilitate readership and subsequent impact. Examining how titles behave in one specific subject area (e.g.: ‘sense making’ literature) over an extended time period, would be something interesting to pursue. Besides, the main source of data we use for all our content analyses are published articles in management journals. Future studies could methodically examine other faucets of idea, rigor, and writing parameters by exploring it using other data sources (extensive interviews, surveys, focus group discussions etc).

To conclude, as a fast developing discipline, management is in dire need to refine and develop sophisticated research methods. Many prominent researchers in the past have pointed out the importance of one or the other

research approach (Popper, 1959; Glaser and Strauss, 1967). On the contrary, we believe theory building is a process which involves different approaches at different stages of development and/or integration of several methods almost simultaneously (Smith and Hitt, 2005). This dissertation thus takes a multidisciplinary and heterogeneous stance to address some of the major issues which qualitative research methods face. To quote R.L. Emerson, through this study we aspire to move away from ‘where the path may lead’ and ‘leave a trail’ (Emerson, 1888) for fellow researchers to follow, to develop on, or to diverge from.

REFERENCES

- Adam, D. (2002). Citation analysis: The counting house. *Nature*, 415(6873), 726-729.
- Adler, N. J., & Harzing, A. W. (2009). When knowledge wins: Transcending the sense and nonsense of academic rankings. *Academy of Management Learning & Education*, 8(1), 72-95.
- Duran, R. P., Eisenhart, M. A., Erickson, F. D., Grant, C. A., Green, J. L., Hedges, L. V., ... & Schneider, B. L. (2006). Standards for Reporting on Empirical Social Science Research in AERA Publications-American Educational Research Association.. *Educational Researcher*, 35(6), 33-40.
- Aguinis, H., Shapiro, D. L., Antonacopoulou, E. P., & Cummings, T. G. (2014). Scholarly impact: A pluralist conceptualization. *Academy of Management Learning & Education*, 13(4), 623-639.
- Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, 16(2), 270-301.
- Aguinis, H., & Joo, H. (2014). *10 Debunking Myths and Urban Legends about how to Identify Influential Outliers. More Statistical and Methodological Myths and Urban Legends*. New York: Routledge.
- Aguinis, H., Suárez-González, I., Lannelongue, G., & Joo, H. (2012). Scholarly impact revisited. *The Academy of Management Perspectives*, 26(2), 105-132.
- Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, 13(3), 515-539.
- Aldrich, H. E., & Fiol, C. M. (1994). Fools rush in? The institutional context of industry creation. *Academy of management Review*, 19(4), 645-670.
- Aleixandre-Benavent, R., Montalt-Resurrecció, V., & Valderrama-Zurián, J. C. (2014). A descriptive study of inaccuracy in article titles on bibliometrics published in biomedical journals. *Scientometrics*, 101(1), 781-791.
- Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., ... & Lang, T. (2001). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals of Internal Medicine*, 134(8), 663-694.
- Altheide, D. L., & Johnson, J. M. (1994). Criteria for assessing interpretive validity in qualitative research. In Denzin, Norman K. (Ed); Lincoln, Yvonna

- S. (Ed), (1994). In *Handbook of Qualitative Research*, (pp. 485-499). Thousand Oaks, CA: Sage Publications.
- Ancona, D. G. (1990). Outward bound: strategic for team survival in an organization. *Academy of Management Journal*, 33(2), 334-365.
- Anthony, L. (2001). Characteristic features of research article titles in computer science. *Professional Communication, IEEE Transactions on*, 44(3), 187-194.
- APA. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63(9), 839-851.
- Atinc, G., Simmering, M. J., & Kroll, M. J. (2011). Control variable use and reporting in macro and micro management research. *Organizational Research Methods*, 15(1): 57-74.
- Atwater, L. E., Mumford, M. D., Schriesheim, C. A., & Yammarino, F. J. (2014). Retraction of leadership articles: Causes and prevention. *The Leadership Quarterly*, 25(6), 1174-1180.
- Aytug, Z. G., Rothstein, H. R., Zhou, W., & Kern, M. C. (2012). Revealed or concealed? Transparency of procedures, decisions, and judgment calls in meta-analyses. *Organizational Research Methods*, 15(1), 103-133.
- Ball, R. (2009). Scholarly communication in transition: The use of question marks in the titles of scientific articles in medicine, life sciences and physics 1966–2005. *Scientometrics*, 79(3), 667-679.
- Banks, G. C., O’Boyle, E. H., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., ... & Adkins, C. L. (2016). Questions about questionable research practices in the field of management A guest commentary. *Journal of Management*, 42(1), 5-20.
- Beach, D., & Pedersen, R. B. (2013). *Process-tracing methods: foundations and guidelines*. Michigan: University of Michigan Press.
- Behling, O. (1980). The case for the natural science model for research in organizational behavior and organization theory. *Academy of Management Review*, 5(4), 483-490.
- Berg, B. L. & Lune, H. (2004). *Qualitative Research Methods for the Social Sciences*. Boston: Pearson.
- Bergh, D. D., Perry, J., & Hanke, R. (2006). Some predictors of SMJ article impact. *Strategic Management Journal*, 27(1), 81-100.
- Berg-Schlosser, D., De Meur, G., Rihoux, B., & Ragin, C.C. (2009). Qualitative Comparative Analysis (QCA) as an approach. *Configurational Comparative Methods*, 1–18.

- Berthon, P., Pitt, L., Ewing, M., & Carr, C. L. (2002). Potential research space in MIS : A framework for envisioning and evaluating research replication, extension, and generation. *Information Systems Research*, 13, 416–427.
- Bettis, R., Gambardella, A., Helfat, C., & Mitchell, W. (2014). Quantitative empirical analysis in strategic management. *Strategic management journal*, 35(7), 949–953.
- Beyer, J. M., Chanove, R. G., & Fox, W. B. (1995). The review process and the fates of manuscripts submitted to AMJ. *Academy of Management Journal*, 38(5), 1219–1260.
- Blatter, J., & Haverland, M. (2012). *Designing Case Studies. Explanatory Approaches to Small-n Research*. Basingstoke: Palgrave Macmillan.
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In S. W. J. Kozlowski & K. J. Klein (Eds.), *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*: 349–382. San Francisco: Jossey-Bass.
- Boone, C., Van Witteloostuijn, A. (2005). Team locus-of-control composition, leadership structure, information acquisition, and financial performance: A business simulation study. *Academy of Management Journal* 48(5), 889–909.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40.
- Bradach, J. (1997). Using the plural form in the management of restaurant chains. *Administrative Science Quarterly*, 42, 276–303.
- Bråten, I., Strømsø, H. I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction*, 21(2), 180–192.
- Browning, L. D., Beyer, J. M., & Shetler, J. C. (1995). Building cooperation in a competitive industry: SEMATECH and the semiconductor industry. *Academy of Management Journal*, 38(1), 113–151.
- Brusoni, S., & Prencipe, A. (2006). Making design rules: A multi domain perspective. *Organization Science*, 17, 179–189.
- Burgelman, R. A. (2002). Strategy as vector and the inertia of coevolutionary lock-in. *Administrative Science Quarterly*, 47, 325–357.
- Buter, R. K., & van Raan, A. F. (2011). Non-alphanumeric characters in titles of scientific publications: an analysis of their occurrence and correlation with citation impact. *Journal of Informetrics*, 5(4), 608–617.

- Calder, B. J., Phillips, L. W., & Tybout, A. M. (1982). The concept of external validity. *Journal of Consumer Research*, 9(3), 240–244.
- Campbell, D. T. (1975). "Degrees of freedom" and the case study. *Comparative Political Studies*, 8(2), 178.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago, Illinois: Rand McNally.
- Carney, M., Gedajlovic, E. (2002). The co-evolution of institutional environments and organizational strategies: The rise of family business groups in the ASEAN region. *Organization Studies*, 23(1), 1-29.
- Carney, M., Gedajlovic, E. R., Heugens, P. P., Van Essen, M., & Van Oosterhout, J. H. (2011). Business group affiliation, performance, context, and strategy: A meta-analysis. *Academy of Management Journal*, 54(3), 437-460.
- Charmaz, K. (2001). Grounded theory: Methodology and theory construction. *International Encyclopedia of the Social and Behavioral Sciences*, 6396-6399.
- Charmaz, K. (2003). Grounded theory. *Qualitative Psychology: A Practical Guide to Research Methods*, 81-110.
- Charmaz, K., & Mitchell, R. G. (2001). Grounded theory in ethnography. *Handbook of Ethnography*, 160-174.
- Church, A. H. (2001). Is there a method to our madness? The impact of data collection methodology on organizational survey results. *Personnel Psychology*, 54(4), 937-969.
- Clemen, R. T., & Ulu, C. (2008). Interior additivity and subjective probability assessment of continuous variables. *Management Science*, 54(4), 835-851.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, New Jersey: Routledge.
- Cole, J. R., & Cole, S. (1972). The Ortega Hypothesis. *Science*, 178(4059), 368-375.
- Conlon, D. E., Morgeson, F. P., McNamara, G., Wiseman, R. M., & Skilton, P. F. (2006). Examining the impact and role of special issue and regular journal articles in the field of Management. *Academy of Management Journal*, 49(5), 857-872.

Cook, T. D. & Campbell, D. T. (1979). *Quasi-Experimental Design: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally.

Cooper, B., & Glaesser, J. (2012, March). Qualitative work and the testing and development of theory: Lessons from a study combining cross-case and within-case analysis via Ragin's QCA. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 13, No. 2).

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1), 3-21.

Cortina, J.M. (2002). Big things have small beginnings: An assortment of "minor" methodological misunderstandings. *Journal of Management*, 28(3), 339-362.

Cote, S., & Miners, C. T. (2006). Emotional intelligence, cognitive intelligence, and job performance. *Administrative Science Quarterly*, 51(1), 1-28.

Cramton, C.D. (2001). The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science*, 12(3), 346-371.

Cronin, B. (1984). *The Citation Process*. London: Taylor Graham.

Curtis, S., Gesler, W., Smith, G., & Washburn, S. (2000). Approaches to sampling and case selection in qualitative research: Examples in the geography of health. *Social Science & Medicine*, 50(7), 1001-1014.

Daft, R. L. & Lewin, A. Y. (1990). Can organization studies begin to break out of the normal science straitjacket? An editorial essay. *Organization Science*, 1(1), 1-9.

Daft, R. L. & Lewin, A. Y. (2008). Rigor and relevance in organization studies: Idea migration and academic journal evolution. *Organization Science*, 19(1), 177-183.

Denk, N., Kaufmann, L., & Carter, C. R. (2012). Increasing the rigor of grounded theory research-a review of the SCM literature. *International Journal of Physical Distribution & Logistics Management*, 42(8/9), 742-763.

Denzin, N. K. (1970). *The Research Act: A Theoretical Introduction to Sociological Methods*. London: Transaction publishers.

Denzin, N. K., & Lincoln, Y. S. (1994). Introduction : Entering the field of qualitative research. In N.K. Denzin, N. K. & Lincoln, Y. S. (Ed). *The Sage Handbook of Qualitative Research*. London: Sage.

Denzin, N. K., & Lincoln, Y. S. (2005). *The Sage Handbook Of Qualitative Research*. London: Sage.

- Diener, R. A. (1984). Informational dynamics of journal article titles. *Journal of the American Society for Information Science*, 35(4), 222-227.
- Dillon, J. T. (1982). In pursuit of the colon: A century of scholarly progress: 1880-1980. *The Journal of Higher Education*, 93-99.
- Dion, D. (1998). Evidence and inference in the comparative case study. *Comparative Politics*, 30(2), 127-146.
- Doz, Y. (2011). Qualitative research for international business. *Journal of International Business Studies*, 42(5), 582-590.
- Dubois, A. & Gibbert, M. (2010). From complexity to transparency: Managing the interplay between theory, method and empirical phenomena in IMM case studies. *Industrial Marketing Management*, 39(1), 129-136.
- Dubois, A., & Araujo, L. (2007). Case research in purchasing and supply management: opportunities and challenges. *Journal of Purchasing and Supply Management*, 13(3), 170-181.
- Eden, D. (2002). From the editors. *Academy of Management Journal*, 45(5), 841-846.
- Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of management review*, 32(4), 1246-1264.
- Edwards, J. R. (2008). To prosper, organizational psychology should... Overcome methodological barriers to progress. *Journal of Organizational Behavior*, 29(4), 469-491.
- Eisenhardt, K. M. & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, 50(1), 25-32.
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532-550.
- Eisenhardt, K. M. (2014). Primer: Qualitative research in strategic management. *Strategic Management Journal*. Retrieved from <http://strategicmanagement.net/pdfs/qualitative-research-in-strategic-management>
- Eisenhardt, K. M., & Tabrizi, B. N. (1995). Accelerating adaptive processes: Product innovation in the global computer industry. *Administrative Science Quarterly*, 84-110.
- Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, 50(1), 25-32.

- Elsbach, K.D. (1994). Managing organizational legitimacy in the California cattle industry: The construction and effectiveness of verbal accounts. *Administrative science quarterly*, 57-88.
- Emerson, R. W. (1888). *Select Writings of Ralph Waldo Emerson* (Vol. 33). London: W. Scott.
- Emigh, R.J. (1997). The power of negative thinking: The use of negative case methodology in the development of sociological theory. *Theory and Society*, 26(5), 649-684.
- Faure, A. M. (1994). Some methodological problems in comparative politics. *Journal of Theoretical Politics*, 6(3), 307-322.
- Ferguson, T. D. & Ketchen, J. D. J. (1999). Organizational configurations and performance: The role of statistical power in extant research. *Strategic Management Journal*, 20(4), 385-395.
- Ferlie, E., Fitzgerald, L., Wood, M., & Hawkins, C. (2005). The nonspread of innovations: the mediating role of professionals. *Academy of management Journal*, 48(1), 117-134.
- Fischer, J., Kaiser, A., & Rohlfing, I. (2006). The push and pull of ministerial resignations in Germany, 1969-2005. *West European Politics*, 29(4), 709-735.
- Fiss, P.C. (2007). A set-theoretic approach to organizational configurations. *Academy of Management Review*, 32, 1180-1198.
- Fiss, P.C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal*, 54, 393-420.
- Flick, U. (2014). *An Introduction to Qualitative Research*. London: Sage.
- Flickinger, M., Tuschke, A., Gruber-Muecke, T., & Fiedler, M. (2013). In search of rigor, relevance, and legitimacy: What drives the impact of publications? *Journal of Business Economics*, 1-30.
- Ford, J.D. (1981). Departmental context and formal structure as constraints on leader behavior. *Academy of Management Journal*, 24(2), 274-288.
- Forte, M., Hoffman, J.J., Lamont, B.T., Brockmann, E.N. (2000). Organizational form and environment: An analysis of between-form and within-form responses to environmental change. *Strategic Management Journal*, 21(7), 753-773.
- Fossey, E., Harvey, C., McDermott, F., & Davidson, L. (2002). Understanding and evaluating qualitative research. *Australian and New Zealand Journal of Psychiatry*, 36(6), 717-732.

- Fox, C. W., & Burns, C. S. (2015). The relationship between manuscript title structure and success: editorial decisions and citation performance for an ecological journal. *Ecology and Evolution*, 5(10), 1970-1980.
- Fox-Wolfgramm, S.J., Boal, K.B., Hunt, J.G. (1998). Organizational adaptation to institutional change: A comparative study of first-order change in prospector and defender banks. *Administrative Science Quarterly*, 43(1), 87-126.
- Freedman, D.A. (2008). On types of scientific inquiry: The role of qualitative reasoning *The Oxford Handbook of Political Methodology*, 300-318.
- Fuchs, H. M., Jenny, M., & Fiedler, S. (2012). Psychologists are open to change, yet wary of rules. *Perspectives on Psychological Science*, 7(6), 639-642.
- Gardner, W.L., Avolio, B.J. (1998). The charismatic relationship: A dramaturgical perspective. *Academy of Management Review*, 23(1), 32-58.
- Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA*, 295(1), 90-93.
- Geddes, B. (1990). How the cases you choose affect the answers you get : Selection bias in comparative politics. *Political Analysis*, 2(1), 131-152.
- George, A. (1979). Case studies and theory development: The method of structured, focused comparison. In P. G. Lauren (Ed.), *Diplomacy. New Approaches in History, Theory, and Policy*. New York: The Free Press.
- George, A. L., & Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*. London: MIT Press.
- Gephart, R. P. (2004). Qualitative research and the Academy of Management Journal. *Academy of Management Journal*, 47(4), 454-462.
- Gerring, J. (2007). *Case Study Research : Principles and Practices*. Cambridge : Cambridge University Press.
- Gerring, J. (2007b). Is there a (viable) crucial-case method? *Comparative Political Studies*, 40(3), 231-253.
- Gibbert, M. (2006). Munchausen, black swans, and the RBV Response to Levitas and Ndofor. *Journal of Management Inquiry*, 15(2), 145-151.
- Gibbert, M., & Nair, L.B. (2013, June). *Boosting internal and external validity in case study research: A review (1996-2006)*. Paper presented at European Academy of Management Conference, Istanbul, Turkey.
- Gibbert, M., Nair, L.B., & Weiss, M. (2014). "Oops, I've got an outlier in my data—what now?" *Using the Deviant Case Method for theory building*.

Paper presented at the 74th Academy of Management Annual Meeting, Philadelphia, USA.

Gibbert, M. & Ruigrok, W. (2010). The “what” and “how” of case study rigor: Three strategies based on published work. *Organizational Research Methods*, 13(4), 710-737.

Gibbert, M., Ruigrok, W., & Wicki, B. (2008). What passes as a rigorous case study? *Strategic Management Journal*, 29(13), 1465-1474.

Gilbert, G. N. (1977). Referencing as persuasion. *Social Studies of Science*, 113-122.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research notes on the Gioia methodology. *Organizational Research Methods*, 16(1), 15-31.

Giordano, P.C. (1989). Confronting control theory's negative cases. In *Theoretical Integration in the Study Of Deviance and Crime: Problems and Prospects*. (pp. 261-278). Albany: State University of New York Press.

Gittell, J.H. (2001). Supervisory span, relational coordination and flight departure performance: A reassessment of postbureaucracy theory. *Organization Science*, 12(4), 468-483.

Glaesser, J., & Cooper, B. (2011). Selectivity and flexibility in the German secondary school system: A configurational analysis of recent data from the German socio-economic panel. *European Sociological Review*, 27(5), 570-585.

Glaser, B. G. & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.

Glaser, B. G. (1978). *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*. Mill Valley: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967, 2006). *The Discovery of Grounded Theory : Strategies for Qualitative Research*. Chicago : Aldine Pub. Co.

Glass, B. (1965). *Science and Ethical Values*. Chapel Hill, NC: University of North Carolina Press.

Gomez-Mejia , L. R., & Balkin, D. B. (1992). Determinants of faculty pay : An agency theory perspective. *Academy of Management Journal* , 35, 921–955.

Gordon, M.M. (1947). Sociological law and the deviant case. *Sociometry*, 10(3), 250–258.

- Gottfredson, S. D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist*, 33, 920–934.
- Grant, R. M. (2003). Strategic planning in a turbulent environment : evidence from the oil majors. *Strategic Management Journal*, 24(6), 491-517.
- Greckhamer, T., Misangyi, V.F., Elms, H., & Lacey, R. (2008). Using Qualitative Comparative Analysis in strategic management research: An examination of combinations of industry, corporate, and business-unit effects. *Organizational Research Methods*, 11(4), 695-726.
- Habibzadeh, F., & Yadollahie, M. (2010). Are shorter article titles more attractive for citations? Cross sectional study of 22 scientific journals. *Croatian Medical Journal*, 51(2), 165-170.
- Hackman, J.R., Oldham, G.R. (1980). *Work Redesign*. New York: Addison-Wesley.
- Hagstrum, J.T. (2013). Atmospheric propagation modeling indicates homing pigeons use loft-specific infrasonic ‘map’ cues. *The Journal of Experimental Biology*, 216(4), 687–699.
- Hambrick, D. C. (2005). Upper echelons theory: Origins, twists and turns, and lessons learned. In *Great minds in management: The Process of Theory Development*. (pp. 109-127). Oxford: Oxford University Press.
- Hancock, G. R. & Mueller, R. O. (Eds.). (2010). *The Reviewer's Guide to Quantitative Methods in the Social Sciences*. New York: Taylor & Francis.
- Hartley, J. (2005). To attract or to inform: What are titles for? *Journal of Technical Writing and Communication*, 35(2), 203-213.
- Hartley, J. (2007). Colonic titles. *Journal of the European Medical Writers Association*, 16(4), 147-149.
- Hartley, J. (2007 b). Planning that title: Practices and preferences for titles with colons in academic articles. *Library & Information Science Research*, 29(4), 553-568.
- Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., & Wilson, S. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, 76(1), 169-185.
- Haynes, K.T., Hillman, A. (2010). The effect of board capital and CEO power on strategic change. *Strategic Management Journal*, 31(11), 1145-1163.

- Hendrick, C. (1990). Replications, strict replications, and conceptual replications: are they important? *Journal of Social Behavior and Personality*, 5(4), 41.
- Hiles, D. (2008). Transparency. In L. M. Given (Ed.), *The Sage Encyclopedia of Qualitative Research*. Thousand Oaks: Sage.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21(5), 967-988.
- Hitt, M.A., Gimeno, J., Hoskisson, R.E. (1998). Current and future research methods in strategic management. *Organizational Research Methods*, 1(1), 6-44.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
- ISI Web of Knowledge. (2013). *Social Sciences Citation Index: Journal Citation Reports*. Philadelphia: Institute for Scientific Information.
- Jacques, T. S., & Sebire, N. J. (2010). The impact of article titles on citation hits: an analysis of general and specialist medical journals. *JRSM Short Reports*, 1(1), 2.
- Jacsó, P. (2006). Dubious hit counts and cuckoo's eggs. *Online Information Review*, 30(2), 188-193.
- Jakobovits, L. A., & Nahl-Jakobovits, D. (1987). Learning the library: Taxonomy of skills and errors. *College & Research Libraries*, 48(3), 203-214.
- Jamali, H. R., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, 88(2), 653-661.
- James, L. R., Demaree, R. G., & Wolf, G. (1993). R WG: An assessment of within-group interrater agreement. *Journal of Applied Psychology*, 78(2), 306-309.
- Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science*, 322(5905), 1259-1262.
- Jones, R., & Noble, G. (2007). Grounded theory and management research: a lack of integrity? *Qualitative Research in Organizations and Management: An International Journal*, 2(2), 84-103.
- Judge, G. G., Hill, R. C., Griffiths, W., Lutkepohl, H., & Lee, T. C. (1988). *Introduction to the theory and practice of econometrics*. New York: Wiley.

- Judge, T. A., Cable, D. M., Colbert, A. E., & Rynes, S. L. (2007). What causes a management article to be cited—article, author, or journal? *Academy of Management Journal*, 50(3), 491-506.
- Katz, J. (1988). *Seductions of Crime*. New York: Basic Books.
- Kaufmann, L., & Denk, N. (2011). How to demonstrate rigor when presenting grounded theory research in the supply chain management literature. *Journal of Supply Chain Management*, 47(4), 64-72.
- Kendall, P.L., Wolf, K.M. (1949). The analysis of deviant cases in communications research. In P.F. Lazarsfeld, F.W. Stanton (Eds.), *Communications research*. New York: Harper.
- Kepes, S., McDaniel, M., Brannick, M., & Banks, G. (2013). Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to mars (the meta-analytic reporting standards). *Journal of Business and Psychology*, 28(2), 123-143.
- Khanna, T., Palepu, K. (1997). Why focused strategies may be wrong for emerging markets. *Harvard Business Review*, 75(4), 41-48.
- Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., & Altman, D. G. (2010). Improving bioscience research reporting: The arrive guidelines for reporting animal research. *PLoS Biology*, 8(6), e1000412.
- King, G., Keohane, R., & Verba, S. (1994). *Designing Social Inquiry : Scientific Inference in Qualitative Research*. Princeton, NJ : Princeton Univ. Press.
- Koch T. (1994). Establishing rigor in qualitative research: The decision trail. *Journal of Advanced Nursing*, 19, 976–986.
- Koenker, R., Hallock, K. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 43-56.
- Kuhn, T.S. (1977). The relations between the history and the philosophy of science. In T.S. Kuhn (Eds.) *The essential tension*. (pp. 3-20). University of Chicago Press: Chicago.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, 9(2), 202-220.
- Lance, C. E. & Vandenberg, R. J. (2009). *Statistical and Methodological Myths and Urban Legends*. New York: Routledge.
- Lawrence, J. & Tar, U. (2013). The use of grounded theory technique as a practical tool for qualitative data collection and analysis. *The Electronic Journal of Business Research Methods*, 11(1), 29-40.

- LeBreton, J. M. & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11(4), 815-852.
- Lee, A., & Baskerville, R. L. (2003). Generalizing generalizability in information systems. *Information Systems Research*, 14(3), 221–243.
- Lee, T. L., Mitchell, T. R., & Sablinski, C. J. (1999). Qualitative research in organizational and vocational psychology: 1979-1999. *Journal of Vocational Behavior*, 55(2), 161-187.
- Lepage, L., Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gøtzsche, P. C., & Lang, T. (2001). The revised consort statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine*, 134(8), 663-694.
- Leung, K. (2007). The glory and tyranny of citation impact: An East Asian perspective. *Academy of Management Journal*, 50(3), 510-513.
- Levi-Faur, D. (2006). A Question of Size?. In *Innovative comparative methods for policy analysis*. (pp. 43-66). US: Springer.
- Lewin, A.Y. (1992). On learning from outliers. F.Y. Phillips, J.J. Rousseau, eds. *Systems And Management Science By Extremal Methods*. Boston: Kluwer Academic Publishers.
- Lewison, G., & Hartley, J. (2005). What's in a title? Numbers of words and the presence of colons. *Scientometrics*, 63(2), 341-356.
- Lewison, G., & Paraje, G. (2004). The classification of biomedical journals by research level. *Scientometrics*, 60(2), 145-157.
- Lieberman, E.S. (2005). Nested analysis as a mixed–method strategy for comparative research. *American Political Science Review*, 99(3), 435-452.
- Lieberson, S. (1992). Einstein, Renoir, and Greeley: Some thoughts about evidence in sociology. *American Sociological Review*, 57(1), 1-15.
- Lijphart, A. (1971). Comparative politics and the comparative method. *The American Political Science Review*, 65(3), 682-693.
- Lijphart, A. (1975). The comparable–cases strategy in comparative research. *Comparative Political Studies*, 8, 158-177.
- Likert, R. (1961). *New patterns of management*. New York: McGraw-Hill.
- Lincoln, Y. S., & Guba, E. G. (1985). Establishing trustworthiness. *Naturalistic Inquiry*, 289-331.
- Lindesmith, A. R. (1968). *Addiction and opiates*. Chicago, IL: Aldine–Atherton.

- Locke, E.A. (2007). The case for inductive theory building. *Journal of Management*, 33(6), 867-890.
- Locke, K. (2001). *Grounded Theory in Management Research*. London: Sage.
- Locke, K. & Golden-Biddle, K. (1997). Constructing opportunities for contribution: Structuring intertextual coherence and “problematizing” in organizational studies. *Academy of Management Journal*, 40(5), 1023-1062.
- Locke, K., Golden-Biddle, K., Feldman, M.S. (2008). Making doubt generative: Rethinking the role of doubt in the research process. *Organization Science*, 19(6), 907–918.
- Lounsbury, M., Glynn, M.A. (2001). Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources. *Strategic Management Journal*, 22(6-7), 545-564.
- Mabe, M.A., Amin, M. (2002). Dr. Jekyll and Mr. Hyde: Author-reader asymmetries in scholarly publishing. *ASLIB Proceedings*, 54(3), 149-157.
- Madden, C. S., Easley, R. W., & Dunn, M. G. (1995). How journal editors view replication research. *Journal of Advertising*, 24(4), 77-88.
- Maggetti, M., Radaelli, C. M., & Gilardi, F. (2012). *Designing research in the social sciences*. London: Sage.
- Mahoney, J., & Goertz, G. (2004). The possibility principle: Choosing negative cases in comparative research. *American Political Science Review*, 98(4), 653–669.
- Mahoney, J., & Goertz, G. (2006). A tale of two cultures : Contrasting quantitative and qualitative research. *Political Analysis*, 14(3), 227-249.
- March, J. G., Sproull, L. S., & Tamuz, M. (1991). Learning from samples of one or fewer. *Organization Science*, 2(1, Special Issue : Organizational Learning : Papers in Honor of [and by] James G. March), 1–13.
- Marcus, A.A., Goodman, R.S. (1991). Victims and shareholders: The dilemmas of presenting corporate policy during a crisis. *Academy of Management Journal*, 34(2), 281-305.
- Mays, N., Pope, C. (2000). Qualitative research in health care: Assessing quality in qualitative research. *British Medical Journal*, 320(7226), 50-52.
- McGrath, J. E., & Brinberg, D. (1983). External validity and the research process: A comment on the Calder/Lynch dialogue. *Journal of Consumer Research*, 10(1), 115-124.
- McGrath, J. E., Martin, J., & Kulka, R. A. (1982). *Judgment Calls in Research*. Beverly Hills: Sage Publications.

- McKinley, W. (2007). Special issues as vertical integration: A rejoinder to priem and mowday. *Journal of Management Inquiry*, 16(3), 240-245.
- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105-2125.
- Merriam, S.B. (1998). *Qualitative Research and Case Study Applications in Education*. Revised and Expanded from "Case Study Research in Education." San Francisco: Jossey-Bass Publishers.
- Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. *American Sociology Review*, 22, 635-659.
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159, 56-63.
- Meyer, A.D. (1982). Adapting to environmental jolts. *Administrative Science Quarterly*, 27(4), 515-537.
- Michelson, G. (1994). Use of colons in titles and journal status in industrial relations journals. *Psychological Reports*, 74(2), 657-658.
- Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B. A., Petersen, M., Sedlmayr, R., Simmons, J. P., Simonsohn, U., & Van der Laan, M. (2014). Promoting transparency in social science research. *Science*, 343(6166), 30-31.
- Miles, M. B. & Huberman, A. M. (1984). *Qualitative Data Analysis: A Sourcebook of New Methods*. London: Sage publications.
- Miles, R.E., Snow, C.C. (1978). *Organizational strategy, structure, and process*. New York: McGraw-Hill.
- Mill, J. (1875). *A system of logic, ratiocinative and inductive : Being a connected view of the principles of evidence and the methods of scientific investigation*. London : Longmans, Green, Reader and Dyer.
- Miller, C. C., Washburn, N. T., & Glick, W. H. (2013). The myth of firm performance. *Organization Science*, 24(3), 948-964.
- Mingers, J. & Xu, F. (2010). The drivers of citations in Management science journals. *European Journal of Operational Research*, 205(2), 422-430.
- Molina-Azorin, J. F. (2012). Mixed methods research in strategic management impact and applications. *Organizational Research Methods*, 15(1), 33-56.
- Moore, A. (2010). What's in a title? A two-step approach to optimisation for man and machine. *Bioessays*, 32(3), 183-184.

- Moravcsik, A. (2014). Transparency: The revolution in qualitative research. *Political Science & Politics*, 47(01), 48-53.
- Morrow, S. L. (2005). Quality and trustworthiness in qualitative research in counseling psychology. *Journal of Counseling Psychology*, 52(2), 250.
- Nagano, R. L. (2015). Research Article Titles and Disciplinary Conventions: A Corpus Study of Eight Disciplines. *Journal of Academic Writing*, 5(1), 133-144.
- Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17(4), 372-411.
- Newman, J. M., & Cooper, E. (1993). Determinants of academic recognition: The case of the Journal of Applied Psychology. *Journal of Applied Psychology*, 78, 518-526.
- Nicolaou, A. I. & McKnight, D. H. (2006). Perceived information quality in data exchanges: Effects on risk, trust, and intention to use. *Information Systems Research*, 17(4), 332-351.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E. L., Simonsohn, U., Soderberg, C., Spellman, B. A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E. J., Wilson, R., & Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425.
- O'Boyle, E., Banks, G. C., & Gonzalez-Mulé, E. (2014). The chrysalis effect how ugly initial results metamorphosize into beautiful articles. *Journal of Management*, 0149206314527133.
- O'Boyle, E., Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology*, 65(1), 79-119.
- Olk, P. & Griffith, T. L. (2004). Creating and disseminating knowledge among organizational scholars: The role of special issues. *Organization Science*, 15(1), 120-129+131.
- Orlikowski, W.J. (1993). CASE tools as organizational change: Investigating incremental and radical changes in systems development. *MIS quarterly*, 30-340.
- Orton, J. D. (1997). From inductive to iterative grounded theory: Zipping the gap between process theory and process data. *Scandinavian Journal of Management*, 13(4): 419-438.

- Paiva, C. E., Lima, J. P. D. S. N., & Paiva, B. S. R. (2012). Articles with short titles describing the results are cited more often. *Clinics*, 67(5), 509-513.
- Parkhe, A. (1993). "Messy" research, methodological predispositions, and theory development in international joint ventures. *Academy of Management Review*, 18(2), 227-268.
- Patton, M.Q. (1990). *Qualitative Evaluation and Research Methods*. London: SAGE Publications.
- Pearce, L.D. (2002). Integrating survey and ethnographic methods for systematic anomalous case analysis. *Sociological Methodology*, 32(1), 103-132.
- Pettigrew, A. M. (1990). Longitudinal field research on change: Theory and practice. *Organization Science*, 1(3), 267-292.
- Piekkari, R., Welch, C., & Paavilainen, E. (2008). The case study as disciplinary convention: Evidence from international business journals. *Organizational Research Methods*, 10.1177/1094428108319905. New version: (2009). The case study as disciplinary convention: Evidence from international business journals. *Organizational Research Methods*. 12(3), 567-589.
- Pirson, M. & Malhotra, D. (2011). Foundations of organizational trust: What matters to different stakeholders? *Organization Science*, 22(4), 1087-1104.
- Pisano, G.P. (1994). Knowledge, integration, and the locus of learning: An empirical analysis of process development. *Strategic Management Journal*, 15(S1), 85-100.
- Podsakoff, P. M., MacKenzie, S. B., Bachrach, D. G., & Podsakoff, N. P. (2005). The influence of management journals in the 1980's and 1990's. *Strategic Management Journal*, 26(5), 473-488.
- Popper, K. (1934, 1959). *The logic of scientific discovery*. London: Hutchinson.
- Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research*, 258-284.
- Pratt, M. G. (2009). From the editors: For the lack of a boilerplate: Tips on writing up (and reviewing) qualitative research. *Academy of Management Journal*, 52(5), 856-862.
- Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

- Ragin, C.C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkley: University of California Press.
- Ragin, C.C. (2000). *Fuzzy-set social science*. Chicago: University of Chicago Press.
- Ragin, C.C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago: University of Chicago Press.
- Ragin, C.C., & Schneider, G. A. (2012). Comparative Political Analysis: Six Case-Oriented Strategies. *The Wiley-Blackwell Companion to Political Sociology*, 78–91.
- Rihoux, B., & Ragin, C.C. (Eds.). (2009). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Thousand Oaks, CA: Sage Publications.
- Rodgers, B. (2008). Audit trail. *The Sage encyclopedia of qualitative research methods*, 1, 43-44.
- Rodgers, B. L., & Cowles, K. V. (1993). The qualitative research audit trail: A complex collection of documentation. *Research in Nursing & Health*, 16(3), 219-226.
- Rodwell, M. K., & Byers, K. V. (1997). Auditing constructivist inquiry: Perspectives of two stakeholders. *Qualitative Inquiry*, 3(1), 116-134.
- Rolfe, G. (2006). Validity, trustworthiness and rigor: Quality and the idea of qualitative research. *Journal of Advanced Nursing*, 53(3), 304-310.
- Rostami, F., Mohammadpoorasl, A., & Hajizadeh, M. (2014). The effect of characteristics of title on citation rates of articles. *Scientometrics*, 98(3), 2007-2010.
- Rowley, J. (2002). Using case studies in research. *Management Research News*, 25(1), 16–27.
- Russell, J. P. (Ed.). (2012). *The ASQ Auditing Handbook*. Wisconsin: ASQ Quality Press.
- Ryan, G. W., & Bernard, H. R. (2000). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Ed.), *Handbook of Qualitative Research*. (pp. 959-978). Thousand Oaks, CA: Sage.
- Sagi, I., & Yechiam, E. (2008). Amusing titles in scientific journals and article citation. *Journal of Information Science*, 34(5), 680-687.
- Sandelowski, M. (1998). Writing a good read: Strategies for re-presenting qualitative data. *Research in Nursing & Health*, 21(4), 375-382.

- Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, 43(6), 1248-1264.
- Schneider, C.Q., & Rohlfing, I. (2013). Combining QCA and process tracing in set-theoretic multi-method research. *Sociological Methods & Research*, 42(4), 559-597.
- Schoenecker, T.S., Cooper, A.C. (1998). The role of firm resources and organizational attributes in determining entry timing: A cross-industry study. *Strategic Management Journal*, 19(12), 1127-1143.
- Schulz, K. F., Altman, D. G., & Moher, D. (2010). Consort 2010 statement: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, 340(7748), 698-702.
- Schweizer L. (2005). Organizational integration of acquired biotechnology companies into pharmaceutical companies: The need for a hybrid approach. *Academy of Management Journal*, 48(6), 1051-1074.
- Seawright, J., Gerring, J. (2008). Case selection techniques in case study research a menu of qualitative and quantitative options. *Political Research Quarterly*, 61(2), 294-308.
- Shapira, Z. (2011). "I've got a theory paper—do you?": Conceptual, empirical, and theoretical contributions to knowledge in the organizational sciences. *Organization Science*, 22(5), 1312-1321.
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2), 63-75.
- Shortell, S.M., Morrison, E.M., Friedman, B. (1990). *Strategic choices for america's hospitals: Managing change in turbulent times*. San Francisco, CA: Jossey-Bass.
- Siggelkow, N. (2001). *Who reads my paper anyways? A survey of the journal readership and reputation*. Working paper, The Wharton School, University of Pennsylvania.
- Singh, J., Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality? *Management Science*, 56(1), 41-56.
- Slater, D., & Ziblatt, D. (2013). The enduring indispensability of the controlled comparison. *Comparative Political Studies*, 46(10), 1301-1327.
- Smith, K. G. (1999). Editor's comments. *Academy of Management Review*, 24(4), 611-613.
- Smith, K. G., & Hitt, M. A. (Eds.). (2005). *Great Minds in Management: The Process of Theory Development*. Oxford: OUP Oxford.

- Soler, V. (2007). Writing titles in science: An exploratory study. *English for Specific Purposes*, 26(1), 90-102.
- Soler, V. (2011). Comparative and contrastive observations on scientific titles written in English and Spanish. *English for Specific Purposes*, 30(2), 124-137.
- John, C. S., & Harrison, J. S. (1999). Manufacturing-based relatedness, synergy, and coordination. *Strategic Management Journal*, 20(2), 129-145.
- Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks: Sage.
- Sterba, S. K., Christ, S. L., Prinstein, M. J., & Nock, M. K. (2011). Beyond treating complex sampling designs as simple random samples: Data analysis and reporting. In A. T. Panter & S. K. Sterba (Eds.), *Handbook of Ethics in Quantitative Methodology*. New York: Routledge.
- Stevens, J.P. (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin*, 95(2), 334.
- Stone, A. & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, 24(3), 236-243.
- Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research: Procedures and Techniques for Developing Grounded Theory*. Thousand Oaks, CA: Sage.
- Stremersch, S., Verniers, I., & Verhoef, P. C. (2007). The quest for citations: Drivers of article impact. *Journal of Marketing*, 71(3), 171-193.
- Subotic, S., & Mukherjee, B. (2014). Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of Information Science*, 40(1), 115-124.
- Suddaby, R. (2006). From the editors: What grounded theory is not. *Academy of Management Journal*, 49(4), 633-642.
- Sullivan, C.J. (2011). The utility of the deviant case in the development of criminological theory. *Criminology*, 49(3), 905-920.
- Sutton, R.I., Callahan, A.L. (1987). The stigma of bankruptcy: Spoiled organizational image and its management. *Academy of Management Journal*, 30(3), 405-436.
- Tahai, A. & Meyer, M. J. (1999). A revealed preference study of management journals' direct influences. *Strategic Management Journal*, 20(3), 279-296.
- Tams, S. & Grover, V. (2010). The effect of an article's structure on its impact. *Information Systems*, 27, 10.

- Tannenbaum, A. (1968). *Control in organizations*. New York: McGraw-Hill.
- Thorngate, W. (1976). "In General" vs. "It Depends": Some Comments of the Gergen-Schlenker Debate. *Personality and Social Psychology Bulletin*, 2(4), 404-410.
- Toncar, M. F., & Munch, J. M. (2010). Meaningful Replication : When is a replication no longer a replication? A rejoinder to Stella and Adam (2008). *Journal of Marketing Theory and Practice*, 18(1), 71-80.
- Tripsas, M., & Gavetti, G. (2000). Capabilities, cognition, and inertia : Evidence from Digital Imaging. *Strategic Management Journal*, 21, 1147-1161.
- Tsang, E. K., & Williams, J. N. (2012). Generalization and induction : Misconceptions, clarifications, and a classification of induction. *MIS Quarterly*, 36(3), 729-748.
- Tsang, E. W. (2014). Generalizing from research findings: The merits of case studies. *International Journal of Management Reviews*, 16(4), 369-383.
- Tsang, E. W., & Ellsaesser, F. (2011). How contrastive explanation facilitates theory building. *Academy of Management Review*, 36(2), 404-419.
- Tsang, E.W.K. & Kwan, K. M. (1999). Replication and Theory Development in Organizational Science: A Critical Realist Perspective. *Academy of Management Review*, 24(4),759-780
- Tsui, A. S. (1999). From the editor. *Academy of Management Journal*, 42(4), 349-350.
- Udy, S.H. (1959). *Organization of Work*. New Haven: HRAF Press.
- Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'theory' back into grounded theory: guidelines for grounded theory studies in information systems. *Information Systems Journal*, 20(4), 357-381.
- Van Maanen, J. (1998). Different strokes: Qualitative re-search in the Administrative Science Quarterly from 1956 to 1996. In *Qualitative Studies of Organizations*, (pp. 1X—XXXII). Thousand Oaks, CA: Sage.
- Van Maanen, J., Sørensen, J. B., & Mitchell, T. R. (2007). The interplay between theory and method. *Academy of Management Review*, 32(4), 1145-1154.
- Van Wesel, M., Wyatt, S., & ten Haaf, J. (2014). What a difference a colon makes: how superficial factors influence subsequent citation. *Scientometrics*, 98(3), 1601-1615.
- Walters, G. D. (2011). The citation life cycle of articles published in 13 American psychological association journals: A 25-year longitudinal

- analysis. *Journal of the American Society for Information Science and Technology*, 62(8), 1629-1636.
- Walton, R.E., Hackman, J.R. (1986). Groups under contrasting management strategies. In P.S. Goodman (Eds.) *Designing Effective Work Groups*. San Francisco: Jossey-Bass.
- Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods*, 18(1), 53-70.
- Weiss, M., Nair, L.B., Gibbert, M., & Koeplin, H. (2015). What passes as a transparent field study in management? *Proceedings of the 75th Academy of Management Annual Meeting*, Vancouver, Canada, 7-11 August, 2015.
- Welch, C., Piekkari, R., Plakoyiannaki, E., Paavilainen-Mäntymäki, E. (2010). Theorising from case studies: Towards a pluralist future for international business research. *Journal of International Business Studies*, 42, 740-762
- Western, B. (2001). Bayesian thinking about macrosociology, *American Journal of Sociology* 107(2), 353–378.
- Westphal, J.D., Zajac, E.J. (1998). The symbolic management of stockholders: Corporate governance reforms and shareholder reactions. *Administrative Science Quarterly*, 43(1), 127-153.
- Whissell, C. (2012). The trend toward more attractive and informative titles: American Psychologist 1946–2010 1. *Psychological Reports*, 110(2), 427-444.
- Wiggins, R.R., Ruefli, T.W. (2002). Sustained competitive advantage: Temporal dynamics and the incidence and persistence of superior economic performance. *Organization Science*, 13(1), 82-105.
- Wiggins, R.R., Ruefli, T.W. (2005). Schumpeter's ghost: Is hypercompetition making the best of times shorter? *Strategic Management Journal*, 26(10), 887-911.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604.
- Woodside, A. G. (2010). Bridging the chasm between survey and case study research: Research methods for achieving generalization, accuracy, and complexity. *Industrial Marketing Management*, 39(1), 64-75.
- Woodside, A. G., & Wilson, E. J. (2000). Constructing thick descriptions of marketers' and buyers' decision processes in business-to-business relationships. *Journal of Business & Industrial Marketing*, 15(5), 354-369.

- Woodside, A.G., & Baxter, R. (2013). Achieving accuracy, generalization-to-contexts, and complexity in theories of business-to-business decision processes. *Industrial Marketing Management*, 42(3), 382-393.
- Worren, N., Moore, K., Cardona, P. (2002). Modularity, strategic flexibility, and firm performance: A study of the home appliance industry. *Strategic Management Journal*, 23(12), 1123-1140.
- Yin, R. K. (1994, 2003). *Case study research : Design and methods*. Thousand Oaks: Sage.
- Yin, R. K. (1981). The case study crisis: Some answers. *Administrative Science Quarterly*, 58-65.
- Yitzhaki, M. (2002). Relation of the title length of a journal article to the length of the article. *Scientometrics*, 54(3), 435-447.
- Zajac, E.J., Shortell, S.M. (1989). Changing generic strategies: Likelihood, direction, and performance implications. *Strategic Management Journal*, 10(5), 413-430.
- Zimmerman, M.A., Zeitz, G.J. (2002). Beyond survival: Achieving new venture growth by building legitimacy. *Academy of Management Review*, 27(3), 414-431.
- Zott, C., Huy, Q.N. (2007). How entrepreneurs use symbolic management to acquire resources. *Administrative Science Quarterly*, 52(1), 70-105.

