

A thesis submitted in accordance with the requirements for the degree of
Doctor in Education (EdD)

Addressing Pedagogical Solitude.

A Realist Evaluation of Organisation Development at a German Higher Education Institution

Michael Vogel
28 June 2014

Department of Lifelong and Comparative Education
Faculty of Policy and Society
Institute of Education, University of London

Author's Declaration

I undertake that all the material presented for examination is my own work and has not been written for me, in whole or in part, by any other person. I also undertake that any quotation or paraphrase from the published or unpublished work of another person has been duly acknowledged in the work which I present for examination.

Word count of chapters 1-10 (including diagrams): 44,997 words

Michael Vogel

28 June 2014

Acknowledgements

I would like to express my sincere appreciation and heartfelt gratitude to my supervisors, Dr Gwyneth Hughes and Professor Ron Barnett, for their ideas, advice, feedback and challenging questions. It has been a delight and a privilege to work, do research and grow under their thoughtful and experienced guidance.

I am also deeply indebted to:

- all my research ‘subjects’ at the *Hochschule Bremerhaven*;
- my fellow doctoral students, especially Judith, Laura, Sarah, Gordon and Shepherd, for many wonderfully stimulating and enlightening conversations;
- Professor Joseph Stockemer, Rector of the *Hochschule Bremerhaven*, for granting me permission to participate in all doctoral research weeks at the Institute of Education during term time in Bremerhaven;
- the *Stifterverband für die Deutsche Wissenschaft* and the State of Bremen for funding the GUUGLE programme, on which this thesis is based;
- Barbara Voges for proofreading the manuscript.

My very special thanks go to my wife Cristina for putting up with my second doctorate. Her understanding, support, love and admirable patience at all times made this thesis possible. Luckily, she enjoyed the London research weeks as much as I did.

Abstract

To establish a culture of collegial exchange about teaching and learning among its academic staff, a German higher education institution is running a seemingly quite successful organisation development (OD) programme, comprising professional learning communities, conferences and other interventions. But how fit for purpose is the programme? A formative realist evaluation is conducted to establish whether and why the programme works for whom and in what circumstances.

On the basis of Coleman's (1987; 1990) social macro-micro-macro scheme, a programme theory is developed and generalised as a framework for theorising, planning, visualising and evaluating OD. Pawson & Tilley's (1997) Realistic Evaluation is chosen as research methodology, modified to match the programme theory's structure and applied to a large data pool covering the OD programme's first four years. Using an explanatory sequential mixed methods research design involving path analysis, content analysis and realist interviews, the programme theory is tested and gradually refined.

The detailed realist evaluation reveals a number of problems at the level of the social mechanisms on which the OD programme's effectiveness and sustainability depend. Unintended self-selection mechanisms limit the programme's prospective fitness for purpose. Also the programme's own history and organisational ramifications interfere with its regular functioning. Various possibilities for improvements are considered and thoughts on the programme's transferability to other contexts are offered. The thesis concludes with critical reflections on Realistic Evaluation.

Table of Contents

Author's Declaration	2
Acknowledgements.....	3
Abstract.....	4
List of Figures.....	10
List of Tables	12
List of Abbreviations	13
Personal Statement.....	14
My motives.....	14
My research preceding the thesis	15
Fitting the pieces together	18
Contribution to my professional development and knowledge.....	20
International and intercultural dimensions.....	21
Chapter 1 Introduction.....	22
1.1 The problem of pedagogical solitude	23
1.2 Purpose and research question of this thesis	25
1.3 Evaluation as research	28
1.4 Insider research.....	30
1.5 Overview of this thesis	34
Chapter 2 Organisation Development and Its Evaluation in a HE Context.....	36
2.1 Organisational problem-solving through adaptive coping	36
2.2 Key features of OD.....	38
2.2.1 Organisational learning and OD.....	39
2.2.2 Dialogic OD	41
2.3 The need for OD at HE institutions.....	43
2.3.1 Three characteristics of HE institutions	43
2.3.2 Problems of adaptive coping at HE institutions.....	45

2.4	Evaluation as OD intervention	47
2.4.1	Examples of OD through evaluation	48
2.4.2	Using Realistic Evaluation of OD in a HE context	49
2.5	Summary	50
Chapter 3	Realist Evaluation – Paradigm, Approach and Design	52
3.1	The philosophy of critical realism	52
3.1.1	The real, the actual and the empirical	53
3.1.2	Causation through generative mechanisms	55
3.1.3	Agency, structure and emergence	58
3.1.4	Retroduction and retrodiction	60
3.2	The realist evaluation of social programmes	62
3.2.1	Realist evaluation and the role of theory	63
3.2.2	Pawson & Tilley's Realistic Evaluation	64
3.2.3	Realist evaluation cycle	67
3.2.4	Theory-driven methodical pluralism	69
3.3	Design of the GUUGLE programme evaluation	70
3.4	Summary	73
Chapter 4	The GUUGLE Programme and Its Context	75
4.1	A context marked by pedagogical solitude	75
4.1.1	Institutional context: a German Fachhochschule	76
4.1.2	Social context I: network structure and collegial exchange	78
4.1.3	Social context II: discourses about teaching and learning	80
4.2	The GUUGLE programme	82
4.2.1	Professional learning communities	83
4.2.2	Pilot PLC 2009-2010	85
4.2.3	GUUGLE programme design	86
4.2.4	GUUGLE in practice	87
4.2.5	GUUGLE-X	90
4.3	GUUGLE as adaptive coping facilitator	91
4.4	Summary	93

Chapter 5 GUUGLE Programme Theory.....	95
5.1 Choosing a realist framework.....	95
5.1.1 Coleman's social macro-micro-macro scheme	96
5.1.2 Bathtub versus rugby ball	97
5.2 The GUUGLE bathtub	100
5.3 From bathtub to whirlpool.....	103
5.4 Reconciling the GUUGLE whirlpool with CMO configurations.....	107
5.5 Rival programme theories	109
5.6 Summary	110
Chapter 6 First Evaluation Cycle: Quantitative Analysis.....	111
6.1 Deriving the hypotheses	111
6.2 Validating the normative outcomes (a)-(c)	113
6.2.1 Validating outcome (b)	114
6.2.2 Validating outcome (a).....	116
6.2.3 Validating outcome (c).....	117
6.3 Testing the hypothesised causal model	118
6.3.1 Path model.....	118
6.3.2 Extending the path model	121
6.3.3 Beneficiaries of GUUGLE and the path model	123
6.4 Specifying the GUUGLE programme	127
6.5 Refining the GUUGLE programme theory	127
Chapter 7 Second Evaluation Cycle: Content Analysis.....	130
7.1 Deriving the hypotheses	130
7.2 DMOC 2: professors' expectations of CETL.....	132
7.3 DMOC 1: communicated reasons for CETL.....	134
7.4 DMOC 4: the nature of CETL in PLCs.....	138
7.5 Critical voices about the GUUGLE programme	143
7.5.1 Preaching to the converted	144
7.5.2 Deficient interfaces	145
7.5.3 Insufficient substitute for leadership and structures.....	146
7.5.4 Disparagement.....	148
7.5.5 Little perceived cultural impact	149

7.6	Re-specifying the GUUGLE programme	151
7.7	Refining the GUUGLE programme theory	152
Chapter 8	Third Evaluation Cycle: Realist Interviews	154
8.1	Deriving the hypotheses	154
8.2	Realist interviews	156
8.2.1	The method of realist interviewing	156
8.2.2	Realist interviews to test the GUUGLE programme theory	158
8.3	Testing the hypotheses	160
8.3.1	DMOC 1: Learning to regard CETL as an important idea.....	160
8.3.2	DMOC 1a: Perceiving GUUGLE as a threat	162
8.3.3	DMOC 2: Learning to regard CETL as an important practice.....	163
8.3.4	DMOC 3: Engaging in CETL	164
8.3.5	DMOCs 4 and 4a: Experiencing CETL as helpful or ineffective	165
8.3.6	DMOC 5 and 6: Turning CETL into culture.....	167
8.4	Re-specifying the GUUGLE programme	168
8.5	Refining the GUUGLE programme theory	171
Chapter 9	Fitness for Purpose – A Discussion.....	174
9.1	Operationalising the concept of fitness for purpose for GUUGLE.....	174
9.2	GUUGLE’s theoretical fitness for purpose	175
9.3	GUUGLE’s practical fitness for purpose	176
9.3.1	A theory-implementation gap	177
9.3.2	The problem of offering motivating reasons for CETL	178
9.4	GUUGLE’s prospective fitness for purpose.....	180
9.5	Enhancing GUUGLE’s prospective fitness for purpose	182
9.5.1	Providing extrinsic incentives	182
9.5.2	Networking and leveraging relationships.....	184
9.5.3	Profiling and managing PLCs	186
9.5.4	Experimenting with new intervention formats.....	187
9.6	Transferring GUUGLE to other contexts	188
9.7	Summary	191

Chapter 10	Concluding Remarks.....	193
10.1	Evaluation findings and implications recapitulated	193
10.2	Contributions of this thesis.....	195
10.3	Critical reflections on Realistic Evaluation.....	197
Appendix.....		202
A1	Images	202
A2	Tables	204
A3	Interview notes	206
References		210

List of Figures

Figure A. Links between my EdD essays, studies and the GUUGLE programme.....	19
Figure B. My EdD essays and studies in the method(ology)-perspective matrix	20
Figure 1.1. Structures, positions and agency	31
Figure 2.1. Adaptive coping cycle	37
Figure 2.2. Cycles of experiential learning and OD compared.....	39
Figure 3.1. Successionist and generative causation	56
Figure 3.2. Society-person connection in the TMSA.....	59
Figure 3.3. Successful programmed social change	65
Figure 3.4. Realist evaluation cycle	68
Figure 3.5. Framework for research design	71
Figure 3.6. Explanatory sequential evaluation design	72
Figure 4.1. Strong-tie and weak-tie networks of 40 professors	79
Figure 4.2. Discourses sustaining the normality of pedagogical solitude.....	81
Figure 4.3. Intended impacts of GUUGLE on the adaptive coping cycle	92
Figure 5.1. Coleman's bathtub	97
Figure 5.2. Coleman bathtub depicting the situation before GUUGLE.....	101
Figure 5.3. Coleman bathtub depicting the intended effect of GUUGLE	102
Figure 5.4. Whirlpool diagram of the GUUGLE programme theory	105
Figure 6.1. Linearised causal pathway described by the DMOCs 2-4.....	119
Figure 6.2. Estimated path model with standardised coefficients.....	121
Figure 6.3. Estimated extended path model with standardised coefficients	122
Figure 6.4. PLC participation frequency and proximity to the GUUGLE team	126

Figure 6.5. Re-looping the path model	128
Figure 6.6. Whirlpool diagram of GUUGLE after the first evaluation cycle	129
Figure 7.1. Reasons stated for PLC (non-) participation	133
Figure 7.2. Normalised topic-purpose impact factors.....	137
Figure 7.3. Whirlpool diagram of GUUGLE after the second evaluation cycle.....	153
Figure 8.1. Basic structure of the realist interview	157
Figure 8.2. Hierarchical clustering dendrogram of the five realist interviews.....	170
Figure 8.3. Whirlpool diagram of GUUGLE after the third evaluation cycle	173
Figure 9.1. Assessing GUUGLE's prospective fitness for purpose.....	175
Figure 9.2. The GUUGLE funnel of self-selection.....	181
Figure 9.3. The high/low context continuu	191
Figure 10.1. Influence of a theory-implementation gap on evaluation	200
Figures A1 a-d. Impressions from GUUGLE PLC sessions and workshops.....	202
Figures A2 a-d. Scenes from GUUGLE Forums	202
Figures A3 a-c. GUUGLE website, newsletter, posters and postcards.....	203

List of Tables

Table 5.1. Sequential DMOCs representing the Coleman bathtub in Figure 5.1.....	108
Table 6.1. Normative GUUGLE programme theory translated into DMOCs	112
Table 6.2. Coverage of DMOCs in different evaluation cycles	113
Table 6.3. PLC participation by professors 2009-2013	114
Table 6.4. Professors' past and intended future participation in GUUGLE	115
Table 6.5. Professors' views regarding the importance of the idea of CETL.....	116
Table 6.6. Professors' views regarding the importance of PLCs.....	117
Table 7.1. Refined GUUGLE programme theory represented by DMOCs.....	131
Table 7.2. Cluster analysis of CETL taking place in PLCs	141
Table 7.3. Professors' views regarding the progress made towards regular CETL.....	151
Table 8.1. The re-refined GUUGLE programme theory in terms of DMOCs.....	155
Table 8.2. Subjects of realist interviews	158
Table 8.3. Summary of the realist interviews	169
Table A1. Path model fit summary	204
Table A2. Correlations between PastGP and possible explanatory variables.....	205

List of Abbreviations

CETL	collegial exchange about teaching and learning
CMOC	context-mechanism-outcome configuration
DMOC	delta context-mechanism-outcome configuration
FoP	Foundations of Professionalism in Education (taught EdD module)
FutuGP	future GUUGLE participation (variable name)
GUUGLE	programme name and acronym for <i>gut und gerne lernen & lehren</i> (German for ‘successful and enjoyable teaching and learning’)
HE	higher education
ICETL	importance of CETL as idea (variable name)
IFS	Institution-Focused Study
IPLC	importance of PLCs (variable name)
MoE1, 2	Methods of Enquiry 1 and 2 (taught EdD modules)
OD	organisation development
PastGP	past GUUGLE participation (variable name)
PLC	professional learning community
SEM	structural equation model
TMSA	Bhaskar’s Transformational Model of Social Activity

Personal Statement

This statement provides a summary and synthesis of my learning experience over the EdD (International) programme as a whole, makes links between its elements, demonstrates how the programme has contributed to my professional development and knowledge and points out the international and intercultural dimensions of the thesis (IOE 2013, 87).

My motives

It all started with Ron Barnett's (2000) 'Realizing the University in an Age of Super-complexity'. The title had made me curious, and I read the book over Christmas and New Year 2006/2007. I had been teaching at a German *Fachhochschule* (or university of applied sciences) for three years and loved it, but I had no background in education studies. Needless to say, much of the book went over my head. Yet it gave me an idea of questions one can ask about higher education (HE), and it aroused in me the desire to gain a deeper understanding of the institutional, social and philosophical context of my work as an academic educator. Six weeks later, I applied for the EdD programme.

A second motive for applying was my wish to extend my research toolbox. A PhD in economics which I completed 17 years ago had provided me with decent skills in mathematical modelling and statistical analysis. But instead of pursuing an academic career I went into strategy consulting and corporate management. Seven years later I returned to academia, still lacking experience in qualitative research. The research training provided to EdD students, I hoped, would allow me to finally become the well-rounded scholar I wanted to be.

A third motive guiding my research emerged during the ‘Methods of Enquiry 1’ module of our programme. Academically speaking, I had grown up in the typical post-positivist monoculture which reputation-hungry economics departments tend to cultivate. I cannot say that I sneered at other theoretical perspectives; I simply did not know anything about them. Non-positivist perspectives had not been part of my economics education. All the more eye-opening was one of the set texts for the module: the first chapter of Crotty’s (2003) ‘Foundations of Social Research’, which introduces the concepts of method, methodology, theoretical perspective and epistemology. I devoured the whole book and resolved to use my EdD research to ‘try out’ different theoretical perspectives.

My research preceding the thesis

International EdD students at the Institute of Education are required to write four essays of 5,000 words each about aspects related to four taught modules, followed by a study of 20,000 words, focusing on the institution or organisation we work for, and the 45,000 word thesis. In the following, I briefly summarise each of my four essays and the institution-focused study before linking them with each other and the thesis.

Module: Foundations of Professionalism in Education. Essay title: The Professionalism of Professors at German Universities of Applied Sciences

In the debate about the (de-)professionalising effects of current HE reforms in Germany, the *Fachhochschulen* have mostly remained silent. Unlike the university professors, professors at *Fachhochschulen* did not seem to regard managerial and market-oriented reforms as a threat to their professionalism. The reasons may not be hard to find, but I asked myself how these professors, who are tenured academics with years of

professional work experience outside academia, construct their professional ideal. Based on a survey of 259 professors from 25 *Fachhochschulen*, I developed a model of normative professionalism, analysed what prevents the professors from achieving it and found that they perceive their substantial teaching obligation as the principal threat to their professionalism. The essay was published (Vogel 2009b).

Module: Methods of Enquiry I. Essay title: Implementing a Professional Learning Community at a German University of Applied Sciences

We were asked to write an outline of a research project which could later become the institution-focused study or the thesis. As a starting point, I chose one of the findings of the survey for the first essay, namely that many professors at *Fachhochschulen* seem to work in ‘pedagogical solitude’, treating their teaching as a private affair and refusing to engage in collegial exchange about teaching and learning. I developed an action research proposal to address the problem by implementing a professional learning community (PLC) for professors at my home institution and to explore its effects in cooperation with the participants. Moreover, to determine the extent of pedagogical solitude I proposed to perform a social network analysis prior to the PLC intervention.

Module: Specialist Course in International Education. Essay title: A Comparative Computational Approach to Understanding Emergent Social Phenomena

This essay deals with the use of comparative methods to study the emergence of social phenomena, such as trust or collective action, from complex interactions of individuals.

Referring to the appearance of new properties in self-organising systems, the concept of emergence combines causality and meaning, requiring researchers to adopt positivist and interpretive perspectives simultaneously to link micro and macro levels. In addition,

comparative research faces the ‘many variables, few cases’ problem, making it difficult to describe the complex dynamics which underlie emergence. I discussed the epistemological status of various emergent social phenomena and assessed the potential of agent-based comparative computational simulation to solve the ‘many variables, few cases’ problem, bridge micro-macro gaps and mediate between positivism and interpretivism. The essay was published (Vogel 2009c).

Module: Methods of Enquiry 2. Essay title: Exploring the Conditions for Academic Teachers’ Informal Collegial Learning About Teaching. A Social Network Approach.

Pedagogical solitude (see my second essay) can be an obstacle to academic teachers’ informal professional learning. In this essay I examined the extent to which this may have been the case at my home institution. Based on a network interpretation of communities of practice, I applied social network analysis to the communication and influence relations among 40 professors, which they had revealed to me in a survey. I found that while some professors were solitary workers, a strong minority was engaging in small communities of teaching practice, connected only by a few individuals who acted as hubs. Informal learning seemed to be obstructed more by a shortage of highly connected professors than by an abundance of isolated ones. Network coherence depended critically on just two professors. The essay was published (Vogel 2009a).

Institution-focused study. Title: Pedagogical Solitude at a German Higher Education Institution. A Critical Discourse Analysis.

The perceived normality and uncritical acceptance of pedagogical solitude by academic teachers in Germany suggested to me that it may have an ideological component. In this study I aimed to explore and expose the ideology behind pedagogical solitude. Using

critical discourse analysis on observations and interviews with professors at my home institution, I identified two discourse categories which appeared to sustain the normality of pedagogical solitude. Working conditions discourses represented the working conditions as obstacles to conversations with colleagues and the professors as victims of their circumstances. Professional identity discourses constructed collegial exchange about teaching and learning as unnecessary by emphasising the importance of content and trivialising the process of teaching. Together, these discourses preserved the professors' powerful position as experts and justified the privatisation of their teaching.

Fitting the pieces together

Pedagogical solitude has been the dominant theme of my work for the EdD and it continues to be so in this thesis. The research outline I wrote in 2008 for the Methods of Enquiry 1 module defined the direction of almost everything I did afterwards, including the programme GUUGLE, a vastly scaled-up version of the action research project I proposed in 2008. A colleague and I started GUUGLE in 2009 as a small, informal initiative with no resources. Over time we raised more and more money for it, added elements, hired staff and were joined by further colleagues. Today, the seven-year organisation development programme operates on a budget of 2.6 million euros and encompasses practically all teaching-related areas of our *Fachhochschule*.

For me the downside of this exciting evolution was that GUUGLE had become so big and demanding to manage that it delayed my thesis by two years. The only way I could work on the thesis whilst running the GUUGLE programme was by creating synergies between them. I made GUUGLE the object of the thesis and used the thesis to better understand and enhance GUUGLE. Figure A depicts the links between my essays,

studies and GUUGLE. Solid arrows indicate direct influence. The dotted line represents a shared concern with social micro-macro gaps and the reconciliation of positivist and interpretive perspectives. As can be seen, three of my written pieces fed directly into the development of GUUGLE, and the Methods of Enquiry 1 essay triggered much of what has kept me extremely busy during the past five years.

Foundations of Professionalism (FoP): Survey data indicate widespread pedagogical solitude among professors at *Fachhochschulen*.



Methods of Enquiry 1 (MoE1): Essay proposes a social network analysis to study, and a PLC for professors to address, pedagogical solitude at my home institution.



Methods of Enquiry 2 (MoE2): Social network analysis explores the extent and characteristics of pedagogical solitude at my home institution.



Institution-Focused Study (IFS): Critical discourse analysis identifies discourses sustaining the normality of pedagogical solitude at my home institution.



GUUGLE: *Organisation development programme at my home institution addressing pedagogical solitude through PLCs and other interventions*.



Thesis: Realist evaluation establishes why the GUUGLE programme works for whom and in what circumstances.



International Education: Studying emergent social phenomena involves closing social micro-macro gaps and combining positivist and interpretive perspectives.

Figure A. Links between my EdD essays, studies and the GUUGLE programme.

Given my initial intention to use the EdD programme to ‘try out’ different theoretical perspectives and get to know new research methods and methodologies, a second way of fitting my written pieces together suggests itself. Rather than emphasising influence

or chronological order, Figure B shows to what extent I did what I set out to do. It depicts a matrix with theoretical perspectives on the horizontal and method(ology)s on the vertical axis. The locations of my essays and studies in the matrix represent specific method(ology)-perspective combinations characterising these works. The thesis, for instance, adopts a critical realist perspective and uses a methodology involving mixed methods. By and large, I would say I realised my initial intention.

	Discourse analysis		IFS
	Action research		MoE1
Method(ology)	Realist interviewing		Thesis
	Content analysis		Thesis
	Cluster analysis		Thesis
	Path analysis		Thesis
	Social network analysis	MoE2	
	Correlation analysis	FoP	Thesis
	(Post-) positivism	Network perspective	Critical realism
			Critical inquiry
		Theoretical perspective	

Figure B. My EdD essays and studies in the method(ology)-perspective matrix.

Contribution to my professional development and knowledge

Practically all of my research and writing for the EdD was about the institution I work for, about my own professional species and about our professional practices. My awareness of what I do and who I am professionally has grown enormously as a result.

Also my research skills and my understanding of the research process have improved greatly over the past few years. Moreover, interacting on a regular basis with leading academics at the Institute of Education has been a most inspiring, motivating and memorable experience. However, perhaps the most important factor for my professional development was my literary encounter with critical writers like Freire, Habermas, Baumann and Kemmis, and also with Lyotards's criticism of critical theory. It is remarkable that I grew up 30 miles from Frankfurt but had to go to England to learn about the Frankfurt School. My teaching of business will never be the same.

International and intercultural dimensions

Since the focus of the thesis is on GUUGLE, a programme implemented at a single German HE institution, its international and intercultural dimensions are naturally limited. Firstly, the thesis was written in English by a German for British readers (presumably). It uses data from German-speaking research subjects and borrows theories and models from the English-speaking literature. Secondly, the referenced literature itself is international. For example, practically all publications on pedagogical solitude have appeared in the United States, whereas the literature on critical realism is predominantly from the UK. Thirdly, since education systems can differ considerably between countries, and the readers of this thesis may not be familiar with the German system, the thesis provides ample background information on German HE, the *Fachhochschulen* and their professors. And finally, the thesis addresses possibilities and problems of transferring the GUUGLE programme to other contexts, including other countries.

Chapter 1 Introduction

In March 2012, two members of my team and I attended the Dortmund Spring School for Academic Staff Developers in Germany to offer a workshop on ‘Organisation Development in Higher Education Through Professional Learning Communities’. We had planned to use the workshop to present and discuss a change programme called GUUGLE, which we had been running for almost three years at our home institution. The programme’s main concern is the stimulation of collegial exchange about teaching and learning (CETL) among teaching staff by means of professional learning communities (PLCs), annual conferences and various other interventions.

When our workshop participants learned that one-third of all full-time teaching staff at our home institution had already participated in at least one PLC, they looked at us in disbelief. One seasoned staff developer told us that at his university only junior academics sometimes discussed their teaching with one another, whereas among the professors this would be considered infra dig. An assistant professor of biochemistry had come to the Dortmund Spring School because at his university he felt isolated with his interest in teaching. Around him the only thing being discussed was research. An associate professor of law from another university said that she had been warned by a senior colleague not to talk with peers about their teaching, as teaching involves intellectual property and she might later be accused of plagiarism. All fifteen workshop participants agreed that CETL was marginalised or even tabooed at their respective universities.

1.1 The problem of pedagogical solitude

This experience confirmed findings of a survey I had conducted a few years earlier (Vogel 2009b) among 259 members of my own professional species, i.e. professors at German universities of applied sciences or *Fachhochschulen*. The purpose of the survey had been to elicit elements of the professors' concept of their own professionalism, but it also brought to light several other aspects. For example, it turned out that by far the most important source of knowledge about teaching and learning was the respondents' own experience. The emphasis is on 'own' because, as the survey showed, teaching-related advice-seeking among colleagues tended to be very uncommon. A majority considered collegial relations neither as important learning opportunities nor as a significant source of motivation, and most preferred working alone.

These findings as well as the stories about marginalised and tabooed CETL are indicators of a phenomenon which has been named *pedagogical solitude* (Shulman 1993). In this thesis I define pedagogical solitude as a cultural condition characterised by the tendency of teachers to refrain from any CETL which goes beyond the required planning and coordination of teaching-related activities and an occasional chat with direct colleagues. Not only individuals, also small, tight-knit groups of teachers can experience pedagogical solitude when they limit their CETL to their own members.

Pedagogical solitude is regarded as problematic for at least three reasons. The first reason concerns the individual academic teachers who treat their teaching as a private affair and shield it from colleagues in the same institution. What happens inside the lecture theatres and classrooms remains between the students and their teachers. Hutchings (2001/2002) observes that

‘teaching, which one might expect to be the most social of work, done in community with others, is in fact much less so than research. Indeed, teaching is lonely work for many faculty, work with a very underdeveloped set of habits and infrastructure for sharing what we learn with colleagues.’

Palmer criticises that, ‘instead of calling it the isolationism it is and trying to overcome it, we claim it as a virtue called “academic freedom” [...] We pay a high price for this privatization’ (Palmer 2007, 147), which consists in unnecessary frustrating experiences and in slow professional development. The resources needed to become better academic teachers faster are available from colleagues, but these resources remain inaccessible in pedagogical solitude.

The second reason why pedagogical solitude is considered a problem is associated with the fact that, unlike research, teaching produces no artefacts which would routinely be subjected to peer review and made available to others. One might say that ‘teaching is a bit like dry ice; it disappears at room temperature’ (Shulman 1993, 7). Moreover, Palmer (2007) points out that while surgeons or trial lawyers work in the presence of colleagues and fellow experts, teachers do not, so that ‘teachers can lose sponges or amputate the wrong limb with no witnesses except the victims’ (146). Without CETL, teaching risks going entirely unproblematised, unchallenged and uncritiqued.

The third reason is the detrimental effect of pedagogical solitude on organisational learning. In the face of rapid and profound transformations in the higher education (HE) landscape, the capacity of universities for organisational learning, understood as the creation, transmission and maintenance of shared mental models, in order to adapt to environmental change in a competent, coherent and timely manner, can be decisive for their legitimacy, sustainability and success. ‘Communication is the lifeblood of

organisational learning' (Marsick, Bitterman & Veen 2000, 14), so lack of communication must result in an organisational learning disability. Pedagogical solitude acts as a communication barrier among academic teachers, prevents important insights and innovations from spreading, obstructs collective problem-solving and can thus inhibit organisational adaptation processes.

Curiously, German-speaking literature on pedagogical solitude in HE is practically non-existent. The reason is surely not the absence of the problem. Rather, pedagogical solitude at German universities may not have been identified as a problem because of its normality. In fact, for centuries it seems to have been a cultural norm. Historically, German universities were organised so as to allow their professors to act as 'small monopolies in thousands of parts' (Clark 1983, 14). The *Lehrfreiheit* (freedom of teaching) was 'not simply the right of professors to speak without fear or favour [...] but denoted...] the absence of a prescribed syllabus, the freedom from tutorial duties, the opportunity to lecture on any subject according to the teacher's interest' (Hofstadter & Metzger 1955, 386). No cooperation, coordination or exchange of any sort was required from the professors with regard to their teaching. *Einsamkeit und Freiheit* (solitude and freedom), key principles of the Humboldtian idea of a university dating back to 1810, still resonate in German professors' professional self-image today.

1.2 Purpose and research question of this thesis

In 2009, a colleague and I initiated the aforementioned GUUGLE programme to address what we saw as widespread pedagogical solitude at our home institution. Our aim was to overcome this cultural condition by means of organisation development (OD). OD is a field of inquiry, a profession and a particular approach to planned organisational

change. OD tends to have an organisation-wide scope and aims to enhance an organisation's adaptive capacity through interventions in human and social processes.

According to a classic, often-cited definition,

'OD is a long-range effort to improve an organization's problem-solving and renewal processes with respect to personal, interpersonal, structural, cultural and technological aspects. This is achieved particularly through a more effective and collaborative management of organization culture [...] with the assistance of a change agent, or catalyst, and the use of the theory and technology of applied behavioural science, including action research' (French & Bell 1973, 15).

OD differs from other forms of planned change, such as organisational restructuring, project management or product innovation, by 'building the organization's ability to assess its [own] current functioning and to make necessary changes to achieve its goals' (Cummings & Worley 2014, 1).

For five years my colleague and I have been developing, expanding, promoting and managing GUUGLE, supported by a team of academic staff developers and student assistants. The OD programme is still in progress and funding is secured until 2016, with a likely prolongation until 2020.

The purpose of this thesis is to *conduct a formative evaluation of GUUGLE*. The findings of the evaluation will be of practical value to the programme management by helping to make GUUGLE more effective and its impact more sustainable. Moreover, in the course of the evaluation, a theory of the GUUGLE programme will be developed and empirically tested. Given the apparent pervasiveness of pedagogical solitude in German HE institutions, the GUUGLE programme theory may be of interest and use to a wider audience than today's stakeholders.

The main contribution of this thesis in academic terms is not the GUUGLE evaluation, but some of its methodological by-products. One of them is a framework for theorising, planning, visualising and evaluating OD and similar interventions. It will be referred to as ‘whirlpool’ scheme. It integrates social macro and micro levels of an organisation, makes explicit the social mechanisms operating between them and forces the user of the scheme to consider the sustainability of any intended organisational change. The whirlpool scheme constitutes the conceptual backbone of the GUUGLE programme theory and evaluation.

Evaluating GUUGLE is no simple matter of ticking boxes. First of all, the programme’s aim of addressing pedagogical solitude is hard to quantify, and no measurable criteria of success were defined when the programme took shape. Secondly, the programme is still ongoing. Even if measurable criteria existed, they may not yet be applicable. And thirdly, when my colleague and I first conceived GUUGLE, we did not have a particular formal theory, blueprint or model in mind. Working under time pressure towards an application deadline, we shaped GUUGLE based on what we felt was necessary for and acceptable within our institutional context. Since then, operational programme management, our teaching obligations and other projects have prevented us from returning to the fundamental questions of theory and design. But since ‘empirical work in program evaluation can only be as good as the theory which underpins it’ (Pawson & Tilley 1997, 83), the evaluation of GUUGLE calls for the prior development of a sound programme theory.

The evaluative research question of this study reads: *How fit is GUUGLE for the purpose of overcoming pedagogical solitude, and how might its fitness be enhanced?*

Fitness for purpose is a widely used quality criterion in education (Harvey & Green 1993; Wittek & Kvernbekk 2011; Woodhouse 2003). It is met when something does what it is intended to do. Other common concepts of quality, such as efficiency, excellence or perfection, are not implied by fitness for purpose. The research question can thus not be paraphrased as: Will GUUGLE always eliminate pedagogical solitude? or: Is GUUGLE the best way to overcome pedagogical solitude? Fitness for purpose is no binary concept. If quality is defined as fitness for purpose, quality improvement means ‘becoming more fit for purpose; that is, changing plans or actions so as to come nearer to achieving the specified purposes’ (Woodhouse 2003, 134).

1.3 Evaluation as research

Scriven (1967) introduced the distinction between the formative evaluation of ongoing programmes and the summative evaluation of completed ones.

‘*Formative evaluation* is evaluation designed, done, and intended to support the process of improvement, and normally commissioned or done by, and delivered to, someone who can make improvements. *Summative evaluation* is the rest of evaluation: in terms of intentions, it is evaluation done for, or by, any observers or decision makers (by contrast with developers) who need evaluative conclusions’ (Scriven 1991a, 20, original emphasis).

The summative evaluation of social intervention programmes is often restricted to the collection, analysis and interpretation of evidence on whether or not a certain desired effect has been achieved, when this is all the clients of an evaluation (e.g. funding body, policymaker or manager in charge) need to meet their accountability obligations. This approach to evaluation treats a programme as a black box. ‘When evaluators talk about the black box “problem,” they are usually referring to the practice of viewing social

programs primarily in terms of effects, with little attention paid to how those effects are produced' (Astbury & Leeuw 2010, 364).

For formative evaluation, however, a black-box approach makes little sense. Trying to improve a programme without understanding its inner workings means groping about in the dark. Formative evaluation needs to open the black box, identify the causal social mechanisms on which the evaluand is based and establish under which conditions they are triggered or fail to operate. To this end, various approaches have been proposed, including Theories of Change (Connell et al. 1995; Weiss 1995) and Realistic Evaluation (Pawson & Tilley 1997). Central to both is the development and empirical testing of theories on how, why and for whom a given programme is expected to work.

Programme theory evaluation avoids the black box problem, makes use of standard social research methods, may generate new theory through the refinement of a given programme theory and can therefore rightfully claim to be evaluation *research*. Unlike black-box evaluation, programme theory evaluation offers the potential to make 'a distinct contribution to the knowledge of the field of study' (IOE 2013, 89), which is one of the requirements of a doctoral thesis. Patton (2012) agrees when he points out that evaluation findings can contribute to general knowledge about interventions if they '*triangulate* with research results, social science theory, expert opinion, practitioner wisdom, and participant feedback' (131, original emphasis).

A key difference between the two aforementioned approaches of programme theory evaluation lies in their underlying research paradigms.¹ With its extensive stakeholder involvement and the participatory negotiation of working assumptions, e.g. about cause-effect relationships, the Theories of Change approach is rooted in the constructionist tradition. Realistic Evaluation, on the other hand, follows the realist tradition in the philosophy of science. It takes its inspiration especially from Bhaskar's (1998; 2008) critical realism, which asserts that the ontologically real may not be perceivable and the perceived may not be ontologically real. A key task of the realist evaluator is therefore to identify and theorise unobservable but real social mechanisms underlying a social programme, which only become manifest through the events they cause.

The evaluation of the GUUGLE programme will follow Pawson & Tilley's approach but modify it to make Realistic Evaluation fit the structure of the GUUGLE programme theory better.

1.4 Insider research

This thesis is the product of insider, or practitioner, research in which I occupy the triple position of GUUGLE programme initiator/director, programme evaluator/researcher and colleague of the researched academics. Figure 1.1 depicts my positions and links each of them to a group of stakeholders representing certain norms, values and role expectations I did not wish to or could not afford to disappoint.

¹ The term 'paradigm' here refers to metaphysical constructs made up of philosophical assumptions about axiology, ontology, epistemology and methodology (Guba & Lincoln 2005; Ponterotto 2005).

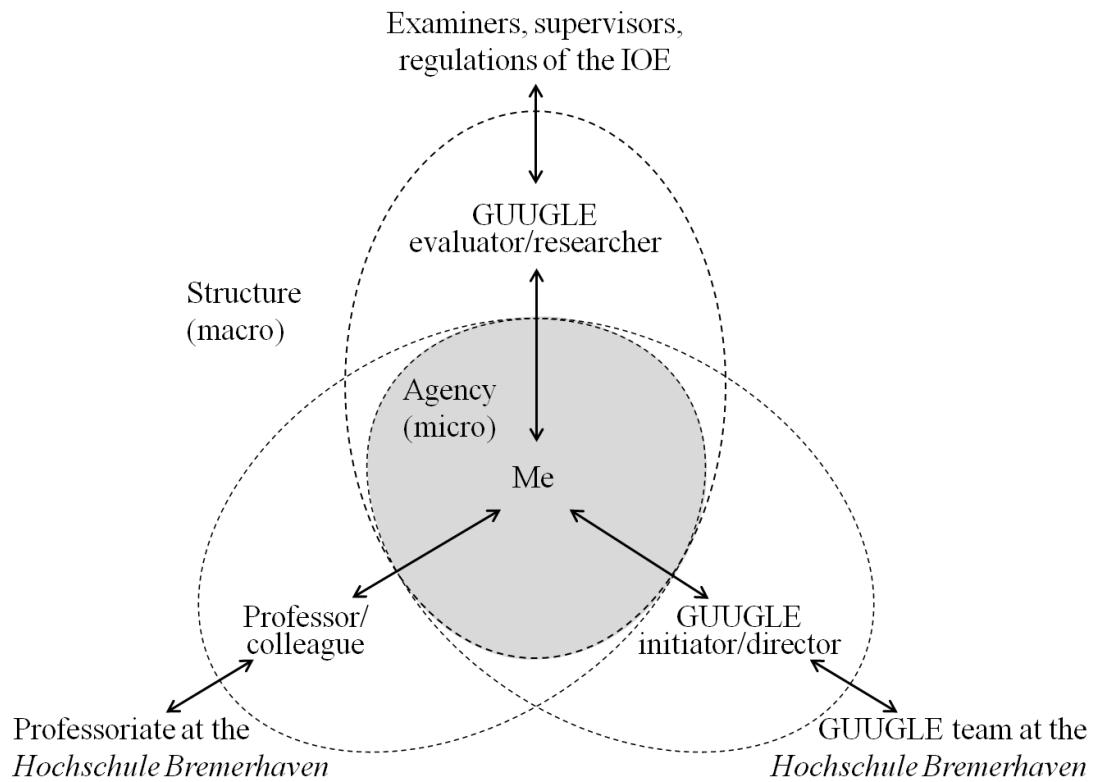


Figure 1.1. Structures, positions and agency

Each position is surrounded by a dashed ellipse which symbolises the socially acceptable role distance available to me in the respective position. The three ellipses overlap in the grey shaded area. Figuratively speaking, I could act freely within the boundaries of the shaded area but needed to beware of possible social sanctions when crossing them. For instance, I could never *only* be a programme evaluator when conducting interviews, as denying the collegial relationship or behaving as if GUUGLE had nothing to do with me would have made me inauthentic and prompted irritated reactions in my interviewees.

Inherent in my triple position is, obviously, a potential for ethically relevant role conflicts. For example, as programme director I may want the evaluation to make the

programme and my contributions to it look good. As a colleague I receive personal information which would be most useful for the evaluation. And as evaluator, I might be tempted to suppress critical assessments of the work of others to avoid possible retaliations on their part. However, the purpose of the evaluation and the chosen approach minimise such conflict potential. This study has no purpose other than to meet the requirements of a doctoral thesis and to help improve the GUUGLE programme. No formal programme evaluation is demanded either by the funding organisation or the senior management of my home institution, and no thesis examiner would be impressed by an uncritical ‘all is fine’ type evaluation, so there is little incentive for me to sugar-coat GUUGLE. Moreover, Realistic Evaluation is about testing and refining programme theory, not about praising, criticising or blaming.

Insider research has been the object of much controversy over the past decades. If the purpose of academic research is the development of theory in accordance with certain standards of rigour, insiders, with their personal stake and substantive emotional investment in the research setting, often find it difficult to legitimate their inquiry as academic research (Alvesson 2003; Anderson & Herr 1999). Insider research may also not be able to meet the same criteria of validity as those applied in more traditional forms of academic research. In constructionist social research, for example, the ‘major criterion of external validity is still the idea of presenting the researcher’s account back to the researched. To be valid, an account must have convergence with the experience of the researched’ (Edwards & Furlong 1985, 33). If the insider researcher’s politics, hidden agendas, loyalties or power relations, which ‘operate in a setting even when insiders think they are being collaborative’ (Herr & Anderson 2005, 36), lead her or him

to representations which do not meet the correspondence criterion, research validity and credibility are threatened.

Robson (2002, 382), on the other hand, points out the advantages of insider research, the most important of which are the easy access to an organisation and its people, as well as the pre-understanding of the research setting and its wider context. Without insiders, much social research could not be undertaken. Insider research can avoid the Hawthorn Effect, i.e. the ‘effect a researcher has on the events she wishes to study’ (Carspecken 2004, 52). Outside researchers attract attention in organisations where insiders simply blend in. Moreover, the results of practitioner research tend to be perceived as more useful and relevant by other practitioners than the results of traditional academic research, which appear to be too theoretical and ‘out of touch’ for them (Anderson & Herr 1999, 12).

In order to benefit from the advantages of my insider position, whilst limiting the influence of power relations on the data, I will conduct the evaluation with a minimum of invasiveness by working largely with existing data, to which also any contracted outside evaluator would be granted access (see section 3.3 for details). My pre-understanding of the institutional setting and the GUUGLE programme will be an asset but also a potential source of bias. To avoid this threat to the validity of my research, I will choose a research design which includes a set of (realist) interviews in which key stakeholders will be asked to comment on my theories and findings. With this feedback, I can test and ensure the convergence of my account with the experience of the researched, as demanded by Edwards & Furlong.

A point related to my different and overlapping roles is the choice of voice in this thesis. To linguistically separate the practitioner who is an interested party and ‘does’ GUUGLE from the more detached, neutral researcher who analyses it, I will largely limit the use of first person to those paragraphs in which I report on my role in the GUUGLE programme. For the most part, the thesis is therefore written from a third-person point of view.

1.5 Overview of this thesis

The remainder of this thesis is divided into nine chapters. Chapter 2 is dedicated to the purpose, concept and methods of OD. Special attention is paid to OD as a facilitator of organisational learning. The chapter then focuses on the role that OD can play at HE institutions, given their unique features. It presents evaluation as an OD intervention in its own right and argues that programme theory evaluation and especially Realistic Evaluation are well suited for evaluating in HE contexts.

Chapter 3 is dedicated to Realistic Evaluation and its realist research paradigm. Bhaskar’s critical realism is outlined, emphasising its assumption of a stratified reality, its reliance on the concepts of generative causation and mechanism and its preferred mode of reasoning, retrodiction. Building on these assumptions and concepts, the chapter introduces Pawson & Tilley’s Realistic Evaluation. Their approach serves as the research strategy of the thesis and informs the research design, which is presented in the final part of this chapter.

The GUUGLE programme’s context, origins, aims and structure are the subject of chapter 4. It sets out by characterising the institutional setting of GUUGLE and pointing

out relevant features of the German *Fachhochschule*. The chapter then presents selected findings of my research on pedagogical solitude that preceded and triggered the launch of GUUGLE. It describes the programme's evolution from a small experimental pilot initiative to a 2.6-million-euro, multi-year OD effort.

Chapter 5 prepares the realist evaluation of GUUGLE by developing a normative programme theory on the basis of a modified version ('whirlpool' scheme) of Coleman's macro-micro-macro model of social change. Also the Realistic Evaluation approach is modified to accommodate the whirlpool and match the structure of the GUUGLE programme theory.

In the three chapters that follow, different hypotheses are derived from the GUUGLE programme theory, tested empirically and used to refine the theory. Testing involves correlation and path analysis (chapter 6), different forms of content analysis (chapter 7) as well as realist interviews with professors (chapter 8).

Chapter 9 returns to the research question of this thesis: whether GUUGLE is fit for the purpose of putting an end to pedagogical solitude. It argues that, in theory, GUUGLE has the potential to be successful. In practice, however, a number of problems persist at the level of the social mechanisms on which the OD programme's effectiveness and sustainability depend. Possibilities for improvements are considered and thoughts on the programme's transferability to other contexts are offered.

Chapter 10, finally, recapitulates the main findings and implications of the GUUGLE programme evaluation, summarises the contributions of this thesis and reflects on merits and shortcomings of Realistic Evaluation.

Chapter 2 Organisation Development and Its Evaluation in a HE Context

The introductory chapter referred to GUUGLE as an OD programme and provided a definition of OD. The present chapter characterises OD in more detail and aims to show in what way OD can help organisations enhance their survivability, why OD is of value to HE institutions, what role evaluation can play in this context, and why Pawson & Tilley's Realistic Evaluation may be well suited to evaluate OD in HE settings.

2.1 Organisational problem-solving through adaptive coping

Organisations are often conceived as open systems embedded in an environment on which they depend, e.g. for their resources, their functions or their legitimacy. This dependence means that changes in the environment may provide opportunities for organisations to prosper and thrive, or they may put their existence at risk. As a consequence of their dependence, all organisations and their sub-systems continuously face two fundamental types of problem (Schein 1965): the need to externally adapt to a changing environment; and the need to maintain internally a degree of integration required for the successful external adaptation.

To illustrate how organisations deal with these two problems and to provide a tool for assessing organisational problem-solving effectiveness, Schein proposes a process model of organisational change, called the adaptive coping cycle. It is depicted in Figure 2.1 and alludes to the psychological concept of coping which denotes any behaviour that solves a stress-producing problem. Adaptive coping differs from

maladaptive coping (e.g. denial, escape or avoidance) by solving the underlying problem (Carver, Scheier & Weintraub 1989).

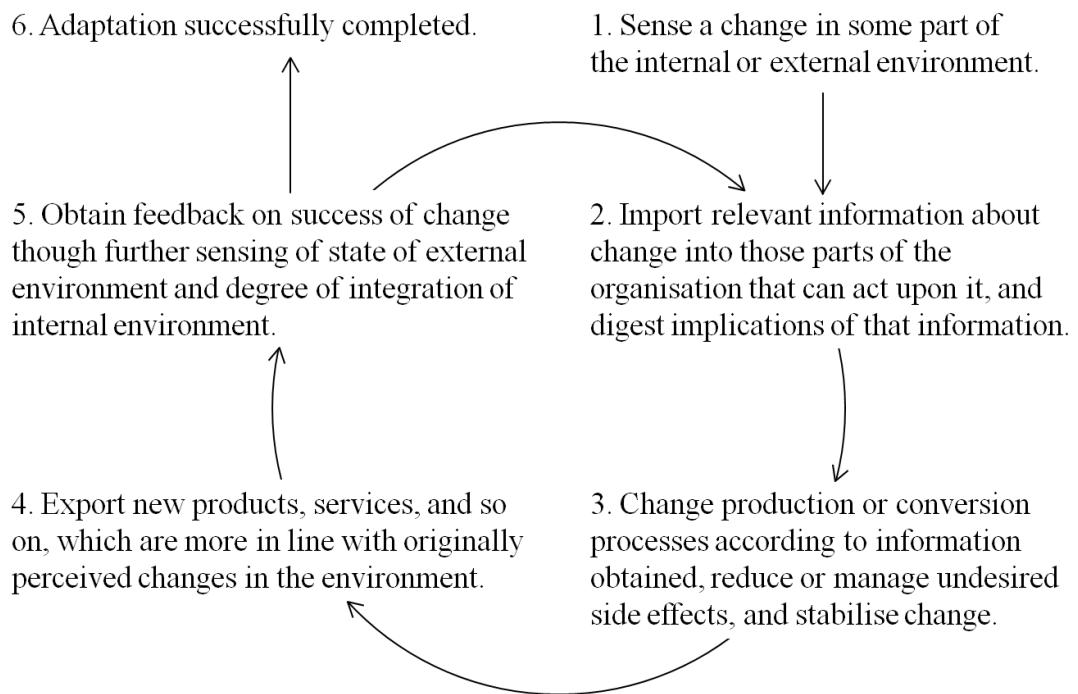


Figure 2.1. Adaptive coping cycle (adapted from Schein 1965, 233-234).

One reason for the model's continued popularity might be the compelling argument on which it is based. Adaptation is triggered by signals from the system's environment, which need to be sensed (phase 1). The signals must be processed by the system, and if the sensing sub-system itself is not able to do this, an internal signal transmission mechanism is indispensable (phase 2). Signal processing capacity without change capacity would make no difference to the system's survivability (phase 3). The system's changes ideally result in the solution of the initial adaptation problem (phase 4). But if this is not the case, a feedback mechanism is needed which tells the system whether further changes are required (phase 5). Strictly speaking, phases 1 and 5 are the same,

but their artificial separation highlights the difference between the initial trigger and the iterative nature of the adaptation process.

When Schein first presented the adaptive coping cycle 50 years ago, he noted that 'If we identify the various stages or processes of this cycle, we shall also be able to identify the points where organizations typically may fail to cope adequately' (quoted in Beckhard 2006, p. 6). To help identify those points and illustrate what OD can do about them is the role of the adaptive coping cycle in this thesis.

2.2 Key features of OD

OD is concerned with system-wide deliberate change in organisations to secure their survivability, uses behavioural science knowledge, targets human and social processes – especially the belief systems of individuals, work groups or culture – and aims to strengthen the capacity of organisations to adapt to change and renew themselves (Cummings & Worley 2001; French & Bell 1999). Characteristic of OD is its systemic orientation, taking into account an organisation's various stakeholders, constituents, organisational sub-entities, internal functions and processes, and aiming to foster greater system awareness among organisation members (Wirtenberg, Abrams & Ott 2004, 469).

The OD process is often depicted as a cycle or spiral consisting of four phases: diagnosis, planning, intervention and evaluation (Bullock & Batten 1985; Kolb & Frohman 1970). In Figure 2.2, this cycle is shown next to Kolb's (1984) experiential learning cycle to emphasise the parallels between the two: both cycles involve forms of learning by doing.

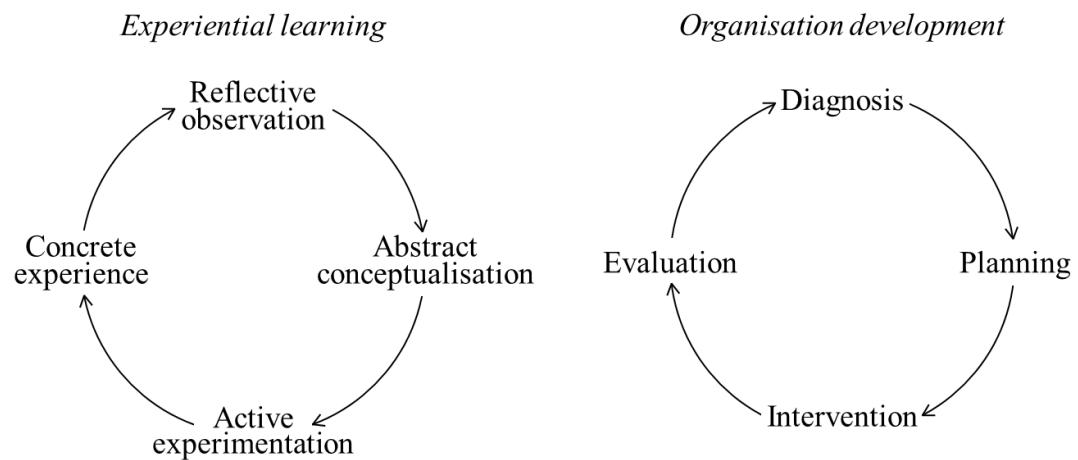


Figure 2.2. Cycles of experiential learning and OD compared.

The evaluation phase corresponds to the concrete experience in experiential learning in that it provides feedback to the learner (i.e. the organisation developer or the organisation itself) about the consequences, impact or success of an intervention. Without feedback, learning by doing is impossible. The reason learning through evaluation is so important for OD is that organisational responses to specific interventions cannot be predicted with any degree of certainty. The same intervention may produce very different outcomes in different contexts, a point which is also central to the realist evaluation approach of this study.

2.2.1 *Organisational learning and OD*

Learning is not only built into OD approaches and methods. An important mission of OD is the improvement of the conditions for '*learning in organisations* and *learning by organisations*' (Popper & Lipshitz 2000, 39, original emphasis), because to 'remain viable in an environment characterised by uncertainty and change, organisations and individuals alike depend upon an ability to learn' (Edmondson & Moingeon 1998, 9).

Learning in organisations is the individual learning of organisation members. For decades, it has been regarded as critical for organisational survivability, since it is people within organisations, not organisations themselves, that plan, decide, organise, lead, control, know, act etc. At the most general level, individual learning is defined as ‘the acquisition of knowledge or skills’ (Oxford Dictionaries 2013).

Learning by organisations, on the other hand, is an anthropomorphic metaphor, which ‘glosses over the fact that if and how organisations can learn is far from self-evident’ (Popper & Lipshitz 1998, 163). Perhaps as a consequence of this, definitions and explanations of organisational learning abound and have so far failed to converge towards a single core concept (Easterby-Smith, Burgoyne & Araujo 1999; Garvin 1993; Hawkins 1994; Huber 1991; Karataş-Özkan & Murphy 2010; Levitt & March 1988; D. Miller 1996; Popper & Lipshitz 1998; Popper & Lipshitz 2000; Rashman, Withers & Hartley 2009; Shrivastava 1983). Instead, two main positions regarding the ability of organisations to learn can be found in the literature. The first position holds that organisations cannot learn and that learning by and in organisations is the same thing. This does not imply, however, that learning in organisations is the same as learning without them: ‘an important component of organisational learning is internal learning – transmission of information from one organisational member or group of members to another. Individual learning in organisations is very much a social, not a solitary, phenomenon’ (Simon 1991, 125).

The alternative position claims that organisational learning cannot be reduced to individual learning because it is a process mediated by, but not bound to individuals. ‘Organizations do not have brains, but they have cognitive systems and memories. [...]’

Members come and go, and leadership changes, but organisations' memories preserve certain behaviors, mental maps, norms and values over time' (Hedberg 1981, 6). Over time, this alternative position has adopted a range of philosophical paradigms, from postpositivism to critical theory, social constructionism and postmodernism (Chiva-Gómez 2003; Cilliers 1998; Karataş-Özkan & Murphy 2010).

In the 1970s, for instance, organisational learning was regarded as the growing capacity of organisations to detect and correct errors through single-loop and double-loop learning (Argyris & Schön 1978), a view clearly inspired by postpositivist thinking. Fifteen years later, following the linguistic turn in the social sciences, social constructionism had become the dominant paradigm. Since then, organisational learning has increasingly been associated with the successive improvement of shared mental models (e.g. Kim 1993) and collective sensemaking, i.e. with a collaborative process of creating shared awareness, understanding and identity (Weick 1995). Organisational learning is seen to have an essential linguistic dimension, because 'Sense is generated by words [...] Words approximate the territory; they never map it perfectly. That is why sensemaking never stops' (ibid, 106-107).

2.2.2 *Dialogic OD*

The challenges posed by organisational learning in terms of sensemaking and creating, transmitting and sustaining shared mental models grow with the size of organisations and their internal functional, hierarchical, geographical or other differentiation. Schein (1993) suggests that each organisational subunit and hierarchical layer which is created in the course of differentiation will develop its own specific local culture, implying the co-existence of different languages and mental models or assumptions about reality

within the same organisation. The essential coordination problem, especially in large, complex organisations, for Schein (1993), is integration across subcultures, that is,

‘the ability to develop an overarching common language and mental model. Any form of *organizational* learning, therefore, will require the evolution of shared mental models that cut across the subcultures of the organisation. The evolution of new shared mental models is inhibited by current cultural rules about interaction and communication’ (41, original emphasis).

Recognising the power of communication and its cultural rules as vehicles for organisational change processes, organisation developers have come up with an arsenal of so-called dialogic OD approaches (Bushe 2013; Holman 2013; Marshak & Bushe 2013), such as appreciative inquiry (Cooperrider, Whitney & Stavros 2008), Open Space (Owen 2008), World Café (Brown & Issacs 2005), Art of Hosting (Büro für Zukunftsfragen 2013), Bohm dialogue (Bohm 1996; Senge 1990), search conferences and dialogic action research (Kristiansen & Bloch-Poulsen 2004; Maurer & Githens 2010; Shotter 2010). The shared idea of these approaches is to change ‘the conversations that shape everyday thinking and behavior by involving more and different voices, altering how and which people engage with each other’ (Marshak & Bushe 2013, 1).

The shift of emphasis from diagnostic to dialogic approaches in OD reflects an understanding of organisations not as machines that may break down and need to be repaired but as (mosaics of) cultures. In consequence, interventions tend to be targeted at value systems, beliefs, rituals, perceptions, shared visions and mental models. Most OD interventions today, including those of the GUUGLE programme, are dialogic, choreographed events creating conditions for stakeholders to ‘share their views of social reality and seek common agreements in real time’ (Bushe & Marshak 2009, 356).

2.3 The need for OD at HE institutions

The bulk of the OD literature refers to private sector companies and institutions of public administration. There is also a body of literature on OD in schools, whereas HE tends not to be in the focus of OD researchers and practitioners. Yet universities and similar institutions can benefit greatly from OD, owing to their nature as loosely coupled systems, professional bureaucracies and organised anarchies. The extent to which these three characteristics prevail will vary between countries and institutions. Since the GUUGLE programme is set in Germany, the following section relates them to the German HE context where appropriate.

2.3.1 *Three characteristics of HE institutions*

A loosely coupled system comprises elements which are responsive to each other, whilst retaining a degree of separateness and identity (Weick 1976, 3), i.e. they affect each other ‘suddenly (rather than continuously), occasionally (rather than constantly), negligibly (rather than significantly), indirectly (rather than directly), and eventually (rather than immediately)’ (Weick 1982, 380). Organisations can be loosely coupled externally with their environment and internally among subunits, levels of hierarchy and individuals as well as between intentions and actions or between activities (Orton & Weick 1990, 208).

One reason why in HE institutions subunits and individuals are loosely coupled is their internal disciplinary fragmentation, which has been circumscribed as ‘academic tribes and territories’ (Becher & Trowler 2001). Another reason is their particular structural configuration as professional bureaucracies: their operating core consists of highly trained specialists (professionals, meaning the academics) who:

‘work relatively freely not only of the administrative hierarchy but also of their own colleagues. [...] Hence] the structure is decentralized both in the vertical and horizontal dimensions. [...] Not only do the professionals control their own work, but they also tend to maintain collective control of the administrative apparatus of the organization.’ (Mintzberg 1980, 333-334)

Mintzberg suggests that due to their inertia, professional bureaucracies typically prevail in stable environments. However, the political, social and economic environment of HE institutions has changed profoundly over the past decades, raising doubts about the sustainability of the professional bureaucracy model in this sector. Reed (2002) declares that ‘Professional bureaucracy cannot be saved’ and that ““New managerialism” is its nemesis’ (169).

At UK universities, managerialism seems to have established a solid foothold and superseded collegial forms of control (Burnes, Wend & By 2013; Reed 2002; Trowler 2010). In Germany, however, managerialism has made comparatively few inroads into the traditional self-governance of HE institutions so far, possibly due to less significant public funding cuts than in the UK and the senior academics’ legal status as tenured civil servants. As such, they are guaranteed particularly high autonomy whilst hardly being held accountable individually.

In the past, German universities have been said to share ‘a unique social structure; almost an inverted hierarchy, with initiative, and even power, moving from below upwards’ (Ashby 1967, 420), forming an organisational environment allowing professors to act as ‘small monopolies in thousands of parts’ (Clark 1983, 140). At older German universities this pattern can still be observed today. At the *Fachhochschulen*, the second-tier HE institutions in Germany, the professors may not be as powerful, but the situation is not fundamentally different. No professor has authority over another or

can be forced to cooperate with colleagues. Senior management has practically no way of implementing decisions or enforcing measures against the will of professors, even individual ones. All major decisions must be made or confirmed by committees dominated by academics whose decision-making behaviour follows its own logic and, at least occasionally, may adequately be represented by the garbage-can model of organisational choice (M. D. Cohen, March & Olsen 1972). The model explains seemingly random outcomes of decision-making processes in ‘organised anarchies’ such as HE institutions, where preferences are problematic, technology is unclear and participation in decision-making is fluid.

2.3.2 Problems of adaptive coping at HE institutions

The combination of loose organisational coupling, professional bureaucracy and garbage can processes of decision-making affects the capacity of HE institutions to cope adequately with changes in their environment. Loosely coupled systems can more accurately sense environmental change (phase 1 of the adaptive coping cycle) than tightly coupled systems, as sensing ‘improves when elements become more numerous and the constraints among them weaken’ (Orton & Weick 1990, 210). However, sensing is selective. Academics register best those changes which take place in their domain of expertise, senior management pays special attention to changes in the political environment and the administration is most sensitive to regulatory changes. Blind spots are most likely to exist in areas which are neither domains of professorial expertise nor political nor government-regulated.

In professional bureaucracies, most decisions are made by relatively autonomous professionals in a decentralised manner. With academics acting simultaneously as

sensors of environmental change, as information processors and, if their own teaching or research is concerned, also as implementers, many small-scale loosely coupled (uncoordinated) adaptive activities may be undertaken locally with minimal delay and minimal impact on the organisation as a whole.

However, the systemic strength of distributed competence and power comes at a cost. Decentralised, uncoordinated adaptive efforts based on individual perceptions of necessity may be inadequate or wholly counterproductive in the face of larger-scale environmental change, especially when the distributed individual responses lead to tensions, contradictions and conflicts within the organisation, or when change would require fundamentally new patterns of thinking and acting. In such cases, organisational learning (i.e. collective sensemaking) and problem-solving is required to arrive at consistent responses to environmental change. This is where the loosely coupled professional bureaucracy shows its major weakness and greatest need of OD.

The import and internal transmission of information about sensed environmental change (phase 2 of the adaptive coping cycle) in loosely coupled systems is sudden rather than continuous, occasional rather than constant, negligible rather than significant, indirect rather than direct, eventual rather than immediate (see section 2.3.1) and therefore notoriously unreliable. The processing of information thus transmitted in ways characterised by the garbage can model almost certainly leads to the choice of poor organisational responses to sensed environmental change (phase 3).

Given the academics' high level of autonomy in what and how they teach and research, and given further the lack of effective control and sanctioning mechanisms on the part of the institutions, at least in the German HE context, the implementation of decisions

made as part of coping efforts (phase 4) is de facto largely voluntary and potentially a case of loose coupling between intentions and actions. As change takes long to work its way through the system, feedback on the success of an internal change agenda, if actually implemented, is received selectively and with a delay by the aforementioned sensory subsystem (phase 5).

In those points where HE institutions, in Schein's words, 'typically may fail to cope adequately' (quoted in Beckhard 2006, 6) due to prevailing cultural rules about interaction and communication, OD seems ideally suited to support the institutions in changing those rules in order to enhance internal communication, information-sharing, cooperation, collective sensemaking and hence organisational learning.

2.4 Evaluation as OD intervention

Evaluation is a phase in the OD cycle (see Figure 2.2) but it is also an OD intervention. The evaluation process with its stakeholder participation, data collection activities and, often, the diffusion of evaluation findings does not leave organisations unchanged. On the one hand, the term evaluation itself is not neutral. It has a negative connotation for many people and is associated with an exam situation, with fears of failing, with stress and frustration. Buchanan, Claydon & Doyle (1999) warn that 'the evaluation of certain types of change, particularly cultural change programmes, is a highly political process and the management of culture may actually be designed in part to forestall independent evaluation and criticism of senior management' (33). If such perceptions persist, the evaluation of an OD intervention turns into an (unintended) intervention in its own right, with counterproductive consequences.

2.4.1 Examples of OD through evaluation

On the other hand, if properly prepared, conducted and communicated, evaluations may have the power to influence organisational performance and culture in positive ways.

Guba & Lincoln's (1989) Fourth Generation Evaluation, for example, is a normative evaluation approach whose key feature is a negotiation process to generate consensus among different stakeholder groups with respect to as many constructions about the issues at hand and their related claims and concerns as possible. As a side effect, this process can result in substantial shifts in perspective and increased critical self-awareness of all parties.

A second example of a normative approach employing evaluation deliberately as an intervention is Fetterman's (1994) Empowerment Evaluation. It involves 'the use of evaluation concepts and techniques to foster self-determination. The focus is on helping people help themselves' (1). Participants of OD programmes are coached by the evaluator to develop their own capacity to carry out evaluations. Fetterman's idea is that the chances of getting evaluations institutionalised and used after the evaluator's departure will be higher if the members of an organisation learn to take ownership of the evaluation process.

Preskill & Torres (1999), finally, assert that their Evaluative Inquiry for Learning in Organisations, which requires systematically building evaluations into organisation members' everyday practices, boosts individual and organisational learning abilities. In their approach, evaluation serves as a mechanism for continual improvement, emphasising 'learning as an outcome of the process, in addition to the more summative, product-oriented outcomes normally expected of an evaluation study' (55).

The reason for stressing evaluation as an OD intervention is that evaluation design is likely to affect the acceptance, credibility, reputation and possibly even the effectiveness of the underlying OD programme. Evaluation design conveys a message to stakeholders. McLuhan's (1964) enigmatic paradox that 'the medium is the message' (7) cautions against the unintended, overlooked consequences of how, as opposed to what, things are done. He points out that 'the personal and social consequences of any medium – that is, of any extension of ourselves – result from the new scale that is introduced into our affairs by each extension of ourselves' (ibid). OD and its evaluation are social innovations and as such 'extensions of ourselves', since they enable individuals to do things they could not do otherwise. Given the scale of change they can trigger and the personal and social consequences potentially associated with it, the message which an OD evaluation sends out through the way it is conducted ought to be crafted carefully.

2.4.2 Using Realistic Evaluation of OD in a HE context

The choice of critical realism as evaluation (or research) paradigm and of Pawson & Tilley's (1997) Realistic Evaluation approach for OD evaluation in this thesis is partly the result of the above considerations. HE institutions comprise academics with backgrounds in different academic disciplines. Each discipline tends to have its favourite research paradigm(s) and to reject certain other ontological, epistemological, axiological and methodological convictions. In particular, positivist and relativist positions are still widely regarded as incompatible, mutually exclusive and to be at 'paradigm war' with each other. Evaluating OD programmes which span different faculties and departments, as in the case of GUUGLE, means that the choice of a

particular evaluation paradigm may be perceived as taking sides. In consequence, evaluation results might be dismissed as invalid by the proponents of other paradigms.

Critical realism, however, combines ontological realism with epistemological relativism, thus bridging the paradigmatic gap. The medium of Realistic Evaluation may therefore be suitable for conveying the message that an OD programme and its evaluation do not take sides with any discipline but aim to integrate different perspectives, encourage exchange between the disciplines and achieve integration across the academic subcultures (see section 2.2.2).

What is more, Realistic Evaluation does not ask *whether* a social intervention or programme works, which would be purely judgemental: it asks ‘*why* a program works for *whom* and in *what* circumstances’ (Pawson & Tilley 1997, xvi, original emphasis). Realistic Evaluation is a theory-driven and theory-generating evaluation methodology focusing on causation and ‘the mechanics of explanation’ (*ibid*, 55), rather than on judgement. For this reason, it should find acceptance especially among academics, stimulate further collegial exchange and set an example in a HE context.

2.5 Summary

HE institutions may not be in the focus of OD, but their nature as loosely coupled systems, professional bureaucracies and organised anarchies makes it difficult for them to cope with profound changes in their environment in a timely, competent and coherent manner. Certain cultural rules about interaction and communication tend to impede organisational learning and adaptive coping. OD, with its dialogic interventions which

target value systems, beliefs and perceptions, promises help by supporting institutions in changing those cultural rules.

OD programmes are often highly visible within organisations. The change processes accompanied or triggered by OD are associated with uncertainties for many stakeholders who therefore follow very attentively not only what the organisation developers do but also how they do it. Therefore, the way OD evaluations are designed and conducted sends a message to organisation members, potentially influencing the acceptance and impact of OD programmes, setting examples and defining new cultural rules. In this sense, programme evaluation can be said to be an OD intervention.

In HE contexts, the choice of an evaluation paradigm and approach is of particular importance for evaluation as an OD intervention. Here, Realistic Evaluation offers the distinct advantages of being based on a ‘unifying’ paradigm that combines a realist ontology with a relativist epistemology and of approaching evaluation as theory-led and theory-generating research.

Chapter 3 Realist Evaluation – Paradigm, Approach and Design

This chapter prepares the ground for the GUUGLE programme evaluation by outlining its three main ingredients: the paradigm of critical realism, the approach of Realistic Evaluation and the research design. The presentation of critical realism goes beyond the minimum necessary to understand Realistic Evaluation, since also the development of the GUUGLE programme theory and the interpretation and discussion of the evaluation findings will make use of critical realist concepts.

3.1 The philosophy of critical realism

The realist tradition in the philosophy of science asserts that the world, physical and social, exists independently from human experience or knowledge of it. Collier (1994) points out that ‘The word “real”, in many contexts, draws its content from its contrast with “apparent”’ (6). Since appearances can be deceiving, the realism of Roy Bhaskar is less concerned with empirical phenomena, a marked difference to positivism, than with the unobservable structures and mechanisms underlying and generating those phenomena. Bhaskar’s critical realism is based on a transcendental ontology, which is why originally he had termed it transcendental realism.

Critical realism takes the world as objectively given but sees the theories about it as subject to historical, cultural and other influences, that is, as socially constructed. Hence Bhaskar links ontological realism with epistemological relativism. Through his sceptical position towards epistemological realism he differs markedly from most scientific

realists. However, he stresses that epistemological relativism does not imply judgemental relativism, the view that all beliefs are equally valid (Bhaskar 1989, 24). On the contrary, the reference of critical realist theories to an objectively existing world is a prerequisite of their fundamental fallibility, refutability and rational judgeability.

3.1.1 The real, the actual and the empirical

Critical realism posits a stratified reality consisting not only of the world and human experience of it, but of three distinct domains: the real, the actual and the empirical. The empirical is the domain of experiences. An experience requires the perception of an event or phenomenon by a person. By contrast, events can occur whether or not they are perceived. Experienced events thus constitute a subset of all events which form the domain of the actual.

Events are actualisations of the unobservable mechanisms which cause them. For Bhaskar (2008), mechanisms are ‘neither phenomena (empiricism) nor human constructs imposed upon the phenomena (idealism), but real structures which endure and operate independently of our knowledge, our experience and the conditions which allow us access to them’ (15). Bhaskar emphasises that:

‘The world consists of mechanisms not events. Such mechanisms combine to generate the flux of phenomena that constitute the actual states and happenings of the world. They may be said to be real, though it is rarely that they are actually manifest and rarer still that they are empirically identified by men. [...] They are not unknowable, although knowledge of them depends upon a rare blending of intellectual, practice-technical and perceptual skills. [...] This is the arduous task of science: the production of the knowledge of those enduring and continually active mechanisms of nature that produce the phenomena of our world.’ (ibid, 37)

In the critical realist model of a stratified reality, the domain of the real comprises the mechanisms which can be thought of as representing all potential events. Some of these actually occur and some of the occurring events are experienced. Hence the domain of the real includes the domain of the actual, which, in turn, includes the domain of the empirical. This relationship can be expressed by the formula $D_r \geq D_a \geq D_e$ (Bhaskar 2008, 47 and 221).

In the natural sciences, experiments are designed in such a way that all three domains are brought in phase, i.e. $D_r = D_a = D_e$. This setting requires a closed system in which all causal mechanisms but one are deactivated, so that whatever event occurs and is observed, it must be attributable to this single mechanism. Only in closed systems is it possible to identify the constant conjunctions between a particular cause and effect, on which much inductive natural scientific reasoning rests.² In the social sciences, however, the objects of knowledge can only be experienced in open systems, making it impossible to apply the same methods and pursue the same kind of reasoning as in the natural sciences. Open systems are characterised by $D_r > D_a > D_e$, which says that there are always more mechanisms than events and more events than observations. The real can never be fully reconstructed on the basis of the actual and even less so on the basis of the empirical.

‘The ineradicably open calibre of social systems, accounts for the absence of (ontologically) crucial or decisive test situations, the breakdown of any significant explanatory/predictive symmetry and the need to rely on exclusively explanatory,

² Bhaskar (2008) argues that positivism, by assuming that the true nature of the world becomes available to humans through observation, commits the error of treating the three domains as one, leaving no room for underlying, unobservable mechanisms (221).

non-predictive criteria for the rational assessment and development of theories in the social sciences.' (ibid, 133)

This quotation highlights the significance of explanation, as opposed to prediction, in critical realism and the social sciences. An explanation articulates the causal mechanism of a phenomenon.

3.1.2 Causation through generative mechanisms

Critical realism holds that everything that happens has a real cause in the form of a mechanism. It takes mechanisms to be constitutive of events and therefore characterises them as generative. 'A mechanism in this sense is not necessarily mechanical in the sense of Newtonian mechanics. It could be an animal instinct, an economic tendency, a syntactic structure, a Freudian "defence-mechanism"' (Collier 1994, 43). Generative mechanisms 'exist as the causal powers of things [..., as] potentialities which may or may not be exercised' (Bhaskar 2008, 40). Put differently, in virtue of their inner structure, objects have the power to cause certain events, given the right conditions. A structure here is 'a set of internally related elements whose causal powers, when combined, are emergent from those of their constituents' (Sayer 2000, 14).

With this concept of generative causation, Bhaskar rejects the positivist or Humean 'successionist' view that there are regular successions of events so that, at its most basic, a change in X always leads to a change in Y (see Figure 3.1 (a)). In this view, X and Y are referred to as explanatory and explained variables or as cause and effect. According to the generative concept of causation, however, predetermination of events is only possible under the conditions of closed systems, whereas 'in open systems, generative mechanisms are not isolated; when triggered, they operate, but in

conjunction with other generative mechanisms, producing a complexly codetermined outcome' (Collier 1994, 62). This is shown in Figure 3.1 (b). For the generative view of causation, the causal explanation of an observed regularity is not a matter of one element exerting influence on another. It is the association as a whole which is to be explained. This is done by identifying the mechanism that brings it about and the contingent contextual conditions triggering the mechanism, because mechanism and context are 'equal partners in generative explanation' (Pawson 2008, 17).

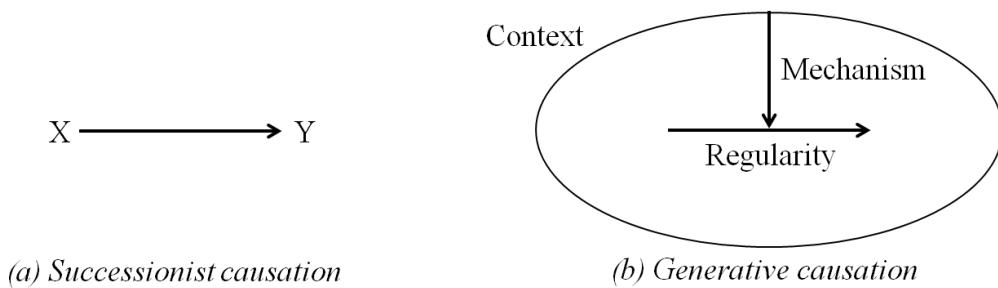


Figure 3.1. Successionist and generative causation

(based on Pawson & Tilley 1997, 72 and Sayer 2000, 14).

Objects exercising their power are always only able to give rise to a limited range of possible events or outcomes. Within this range, some events or outcomes will be more typical, usual or characteristic than others. In Bhaskar's (2008) words, 'some complex structured objects reveal, in virtue of their pre-formed structure, what I am going to metaphorically characterise as an "ontological preference" for some but not other of the natural possibilities open to them' (222). This ontological preference or disposition of objects is called their tendency.

All statements of laws in the natural and social sciences, for Bhaskar, are tendency statements, not statements about constant conjunctions. Only in closed systems do tendencies result in their manifestations and produce strict empirical regularities. In open systems, however, tendencies may not be actualised or observable and at best produce partial or ‘demi-regularities’ (Lawson 1997). Bhaskar’s concept of tendency is essential for the explanation of social phenomena which, on the one hand, display certain demi-regularities without which societies could not function; but, on the other hand, are varied enough to allow for innovation, change and surprise. As Collier (1994) notes, ‘Explanation in open systems is in terms of tendencies’ (63).

A final aspect to be mentioned here is the role of people in causation. The critical realist view is that since people are able to change real things, their beliefs, desires and reasons are real causes. Bhaskar (1998) suggests ‘that intentional human behaviour is caused, and that it is always caused by reasons, and that it is only because it is caused by reasons that it is properly characterised as intentional’ (89). He argues that ‘either a reason will make a difference to behaviour or it will not. In the former case it counts as a cause. In the latter case it is logically redundant, and deliberation, ratiocination (and indeed thought generally) become practically otiose’ (92).

If reasons belong to the causal order, they interact with other causes, which may be other people’s reasons. This introduces causation into human interaction. From a critical realist perspective, the exchange of ideas, opinions and reasons may induce its participants, through a causal mechanism, to change their mind and trigger learning processes.

3.1.3 Agency, structure and emergence

If reasons are causes of human action and human action shapes the social processes that constitute organisations, then individuals' belief systems, which are a source of reasons, can be said to have causal power within organisations. In line with this argument, OD interventions target the belief systems of individuals (see section 2.2). This approach reflects a conception of social explanation which regards organisations and other social objects 'as the results of (or as constituted by) intentional or meaningful human behaviour' (Bhaskar 1998, 34).

According to this view, social structures are determined by human agency. However, causal powers also work in the opposite direction. New members get socialised into the norms and values of the organisation they have just joined and internalise a set of locally shared beliefs, discourses and habits which will henceforth tend to direct, orientate or constrain their thinking and acting. The view that social entities have 'a life of their own, external to and coercing the individual' (Bhaskar 1998, 34), and that therefore social structures largely determine human behaviour corresponds to the functionalist conception of social explanation.

Bhaskar rejects both of these conceptions and also their dialectical combination in favour of his own transformational model of social activity (TMSA), which he summarises as follows:

'people do not create society. For it always pre-exists them and is a necessary condition for their activity. Rather, society must be regarded as an ensemble of structures, practices and conventions which individuals reproduce or transform, but which would not exist unless they did so. Society does not exist independently of human activity (the error of reification). But it is not the product of it (the error of voluntarism). [...] Society, then, provides the necessary conditions for intentional

human action, and intentional human action is a necessary condition for it. Society is only present in human action, but human action always expresses and utilises some or other social form.' (Bhaskar 1998, 36-37)

In Archer's (1998) words, 'Society is that which nobody wants, in the form they encounter, for it is an unintended consequence' (165). The TMSA posits the existence of a symbiotic relationship between society and people or between social structures and human agency, in which each enables and depends on the other without being reducible to it (see Figure 3.2). The mutual dependence and irreducibility of structure and agency allow Bhaskar to argue that social structures and human agency are '*existentially interdependent but essentially distinct*' (Bhaskar 1986, 123, original emphasis).

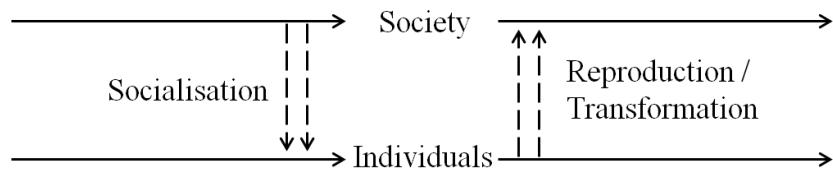


Figure 3.2. Society-person connection in the TMSA (Bhaskar 1998, 36).

The existential interdependence of structure and agency means that society and its structures are the condition as well as the outcome of human agency; and human agency involves the (mostly unconscious) reproduction and the (mostly conscious) transformation of society and its structures. 'One could refer to the former as the *duality of structure*, and the latter as the *duality of praxis*' (Bhaskar 1998, 38, original emphasis).

The essential distinctiveness of structure and agency lies in the different kinds of generative mechanisms governing them. Bhaskar describes an object's governing

mechanisms as being arranged in hierarchically ordered layers or strata (i.e. physical, chemical, biological, psychological and socio-cultural). Each stratum is rooted in and emergent from the more basic ones. All mechanisms are specific to the stratum in which they operate, so that knowledge of mechanisms of lower strata does not permit predicting mechanisms of the next higher stratum. Human behaviour, for instance, is dependent on and constrained by physical, chemical and biological mechanisms, yet it cannot be explained or predicted on their basis, because human behaviour is also governed by psychological mechanisms which are irreducible to the lower-strata mechanisms. Analogously, social structures are governed by socio-cultural generative mechanisms which are rooted in and emergent from, but essentially distinct from the mechanisms operating in the more basic stratum of human agency.

3.1.4 Retroduction and retrodiction

Bhaskar's posit of a stratified reality including the domain of the real, the role he assigns to generative mechanisms in his theory of causation and the criticism he levels against positivism lead him to reject induction and deduction in favour of retrodiction as the most appropriate method of inference. Retroduction is a form of transcendental argument, i.e. an argument based on the (assumed) existence of mechanisms and structures in the domain of the real, which are inaccessible to experience and manifest themselves only indirectly through the phenomena they cause. So retrodiction depends on a transcendental ontology such as Bhaskar's. 'When retrodicing, researchers ask "why things appear as they do". Thus critical realists try to move from what is experienced toward knowledge of what is really there' (Olsen 2010, xxv).

Bhaskar (1998) suggests that ‘the construction of an explanation for [...] some identified phenomenon will involve the building of a model [...] of a mechanism, which *if* it were to exist and act in the postulated way would account for the phenomenon in question’ (5, original emphasis). Thus retrodiction is no formalised logic of inference (Danermark et al. 2002; Downward & Mearman 2007, 12), since it includes a major creative aspect. It is rather a systematic thought process, which starts from the knowledge of a phenomenon, moves to the knowledge of the causes of somehow ‘analogous’ phenomena and culminates in the theorising of possible causes and alternative explanations of the phenomenon in question. According to Bhaskar & Lawson (1998), retroductive reasoning is central for an

‘explanatory science [...] which] seeks to account for some phenomenon of interest [...] in terms of a (set of) mechanism(s) most directly responsible. Producing this explanation will involve drawing upon existing cognitive material, and operating under the control of something like a logic of analogy and metaphor, to construct a theory of a mechanism that, if it were to work in the postulated way, could account for the phenomenon in question. The reality of the mechanism so retrodicated is subsequently subjected to empirical scrutiny, and the empirical adequacy of the hypothesis maintained compared to that of competing explanations’ (5).

This description characterises the first of two basic models of explanation which Bhaskar deems particularly suitable in the context of his transcendental realist ontology. While retrodiction is the inference from events to their underlying generative mechanisms, retrodiction is the inference from events to their antecedent conditions, i.e. to the specific circumstances under which the mechanisms are actualised. Retrodiction and retrodiction are complementary activities, but the ability to retrodict presupposes theoretical explanation and therefore retrodiction.

3.2 The realist evaluation of social programmes

Evaluation, like any kind of systematic inquiry, is framed and guided by paradigms, i.e. ‘patterns of beliefs and practices that regulate inquiry within a discipline by providing lenses, frames and processes through which investigation is accomplished’ (Weaver & Olson 2006, 460). By far the most widely used paradigms in the evaluation of social programmes are postpositivism, pragmatism, constructionism and critical theory. Critical realism still seems to be regarded as too exotic by the authors of general evaluation textbooks to even touch upon the possibility of realist evaluation. Mertens & Wilson (2012), for instance, in their 620 page ‘comprehensive guide’ to programme evaluation theory and practice make no mention of realist perspectives whilst dedicating 110 pages to the four aforementioned paradigms.

The most comprehensive treatises of realist evaluation so far have been provided by Pawson & Tilley (1997; see also Pawson 2000; Pawson & Tilley 2001; Pawson 2003; Pawson & Manzano-Santaella 2012; Pawson 2013), whose Realistic Evaluation will be outlined below and applied to the GUUGLE programme; and by Henry, Julnes, Mark and others (Henry & Julnes 1998; Henry & Rog 1998; Julnes & Mark 1998; Julnes, Mark & Henry 1998; Mark, Henry & Julnes 1998; Mark & Henry 1998; Mark, Henry & Julnes 2000) on Emergent Realist Evaluation.

Beyond that, a number of elaborations on more specific theoretical and methodological aspects of realist evaluation (e.g. Astbury & Leeuw 2010; Blamey & Mackenzie 2007; de Souza 2013; Jackson & Kolla 2012; Pedersen & Rieper 2008; van der Knaap et al. 2008) and of empirical realist evaluation studies from different disciplines (e.g. Dalkin

et al. 2012; Gill & Turbin 1999; Greenhalgh et al. 2009; Leone 2008; Marchal et al. 2012; Ranmuthugala et al. 2011) have been published.

Since Realistic Evaluation enjoys a much wider acceptance as an empirical evaluation approach than Emergent Realist Evaluation and since most publications on Emergent Realist Evaluation have been authored by the originators themselves, preference in this study is given to Pawson & Tilley's Realistic Evaluation.

3.2.1 Realist evaluation and the role of theory

Scriven's (1967) distinction between the formative evaluation of ongoing programmes and the summative evaluation of completed ones (see section 1.3) claims to be exhaustive. Yet realist evaluation goes beyond this dichotomy by asking, 'why a program works for *whom* and in *what* circumstances' (Pawson & Tilley 1997, xvi, original emphasis). Given this question, it is clear that realist evaluation is not about running down checklists of criteria or comparing actual outcomes with intended ones.

Against Scriven's (1991b) view that theories are a 'luxury for the evaluator, since they are not even essential for *explanations*, and explanations are not essential for 99% of all evaluations' (360, original emphasis), realist evaluators insist that 'empirical work in program evaluation can only be as good as the theory which underpins it' (Pawson & Tilley 1997, 83). In fact, realist evaluators regard social programmes as (manifestations of) theories on how the desired outcomes are expected to result from interventions. Programme evaluation for them is theory testing.

More specifically, the theories which social programmes embody correspond to what Merton (1968) termed theories of the middle range, i.e. 'theories that lie between the

minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory' (448) of social systems. Merton stresses that unlike all-inclusive or grand theories³, middle-range theories are 'close enough to observed data to be incorporated in propositions that permit empirical testing' (*ibid*). This implies that a social programme can be evaluated by empirically testing hypotheses or propositions derived from the (middle-range) theory on which the programme is based.

The central role of theory in realist evaluation is a direct consequence of the transcendental nature of the main objects of interest: the unobservable, hypothesised generative mechanisms underlying social programmes and the contextual conditions under which the mechanisms are actualised to co-produce observable events. It is through the acts of theorising and testing and not through induction that mechanisms behind empirical demi-regularities are identified and explanations established.

3.2.2 Pawson & Tilley's Realistic Evaluation

Pawson & Tilley (1997) call their approach Realistic Evaluation because, according to them, it comprises the real, the realist and the realistic: the approach deals with social realities, adopts a realist paradigm and claims to be realistic in the sense of practical feasibility and usefulness. Realistic Evaluation starts from the insight that the aim of every social intervention programme is to replace some prevailing social regularity (R_1)

³ Grand theories result from 'the initial choice of a level of thinking so general that its practitioners cannot logically get down to observation' (Mills 2000, 33). Examples include Parson's (1951) structural functionalism, Habermas's (1984; 1987) theory of communicative action and Luhmann's (1996) theory of social systems (see also Skinner 1985).

which is seen as problematic with another one (R_2) deemed more acceptable or desirable. The shift from R_1 to R_2 depicted in Figure 3.3 represents the successful outcome (O) of an intervention.

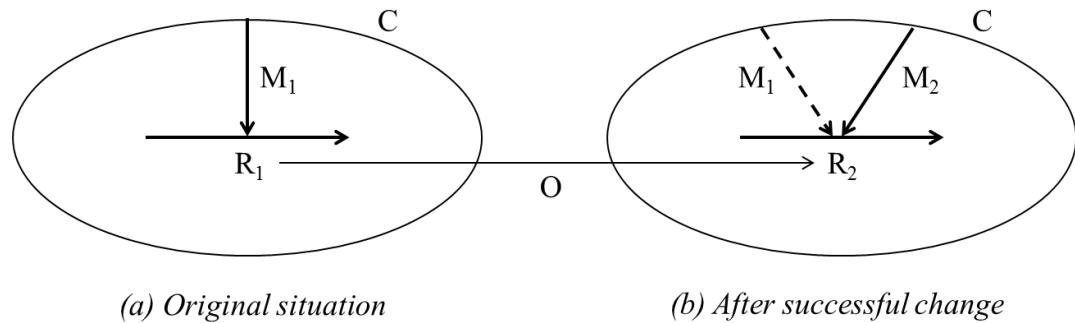


Figure 3.3. Successful programmed social change (Pawson & Tilley 1997, 74).

The initial regularity R_1 is sustained by the mechanism M_1 . In order to bring about change, a social programme must transform, counteract or overturn M_1 with the new mechanism M_2 . In Figure 3.3 (b), the problem mechanism M_1 is indicated by a dashed arrow reflecting its successful displacement by the programme mechanism M_2 . Note that the context C in which the regularities R_1 and R_2 are caused by the mechanisms M_1 and M_2 is the same in Figure 3.3(a) and (b). The assumption of a stable context is usually made when planning interventions, since the aim is not to alter the context itself. However, the same intervention in a different context can produce a different outcome: M_2 may prove unable to displace M_1 and R_1 might persist.

Pawson & Tilley summarise the realist proposition as follows: '*causal outcomes follow from mechanisms acting in contexts*' (58, original emphasis) or simply (C)ontexts + (M)echanisms = (O)utcomes. In this formula resonates the aforementioned overriding question of realist evaluation, why a programme works for whom and in what

circumstances. Why it works is a matter of its mechanisms; in what circumstances it works refers to its contexts; and for whom it works defines its outcomes.

An important aspect of Realistic Evaluation is Pawson & Tilley's understanding of social mechanism in relation to programmes. First of all, social mechanisms can originate from relationships among individuals or from relationships between individuals and social structures. This means that 'causal powers reside not in particular objects [...] or individuals [...] but in the social relations and organisational structures which they form' (64). Like Bhaskar in his TMSA, Pawson & Tilley recognise the stratified nature of social reality and ascribe distinct causal powers to each stratum.

Secondly, social mechanisms operate by affecting people's reasons and resources, altering their choices and capacities. In line with Bhaskar, Pawson & Tilley consider reasons to be the causes of intentional human behaviour. Social interventions cannot change people's behaviour; 'rather it is the underlying reasons or resources that they offer subjects that generate change' (Pawson 2002, 342). Therefore, in Realistic Evaluation, theories are especially concerned with psychological responses to incentives leading to behaviour change.

Thirdly, and again in accordance with Bhaskar, Pawson & Tilley stress the transcendental nature of social mechanisms. Since mechanisms cannot be observed, identifying them requires developing 'propositions about what it is within the program which triggers a reaction from its subjects. These hypothesised processes [...] always work in a "weaving process" which binds resources and reasoning together' (66). The procedure of generating propositions about mechanisms which, if they were to exist and act as proposed, would account for the phenomenon in question, is retrodiction.

Finally, Pawson & Tilley emphasise the role of context in activating social mechanisms, realising causal powers and turning potential events into actual ones. The relationship between mechanisms and their effects is not fixed but conditioned by context, which comprises individuals' characteristics, interpersonal relations, institutional settings and the wider social, economic and cultural setting (Pawson 2013, 37). This is crucial for empirical programme evaluation, as it makes clear that retrodiction is necessary but not sufficient to establish an explanation of why and how a programme works for certain stakeholders. Retroduction needs to be complemented by retrodiction.

A realist explanation is a proposition or theory which combines the three explanatory components context (C), mechanism (M) and outcome pattern (O). Pawson & Tilley refer to such propositions as CMO configurations or CMOCs. 'The task of realist evaluation is to find ways of identifying, articulating, testing and refining conjectured CMO configurations' (77).

3.2.3 Realist evaluation cycle

Like adaptive coping, experiential learning and OD (see Figures 2.1 and 2.2), Realistic Evaluation proceeds in an iterative, circular manner. It can be decomposed analytically into four distinct phases depicted in Figure 3.4. The starting point consists of an initial programme theory on how mechanisms and contexts may give rise to intended outcome patterns. The theory can originate from various sources, including stakeholders' understandings of how the programme works or why it fails to work ('folk theories'), research-based literature, programme planning documents or common sense.

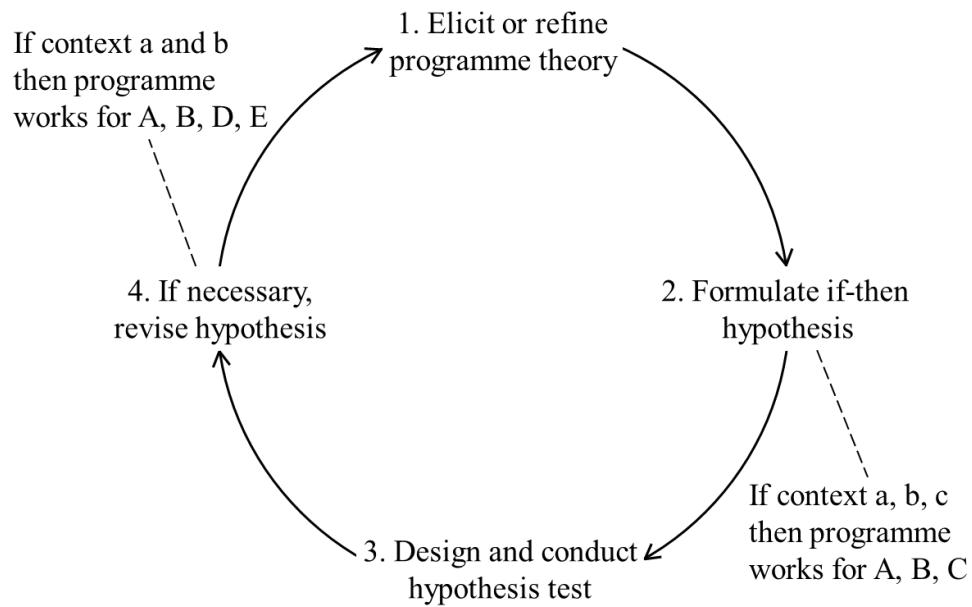


Figure 3.4. Realist evaluation cycle (adapted from Pawson 2013, 88).

In a second phase, hypotheses are derived from the initial theory, stating what kinds of observable regularities ought to be expected when and where. The idea is to empirically test the hypotheses in phase 3, in order to draw conclusions regarding the adequacy of the underlying theory. Critical realism's combination of a realist ontology with a relativist epistemology establishes the possibility of and the need for validating theories empirically, even though critical realism itself – in the face of conceptually mediated and theory-laden empirical observations (Sayer 2000, 70) and the impossibility of conclusive falsification in empirical social research (Archer 1998; Sayer 1992) – ‘takes a balanced and modest stance regarding the prospects for affirming and rejecting theories based on empirical evidence’ (K. D. Miller & Tsang 2011, 144).

The empirical testing of hypotheses, even if some of the results are inconclusive, contributes to a better understanding of a programme, allowing in phase 4 to specify with increasing accuracy which of its parts work for whom in what circumstances. The

programme specification thus obtained may or may not conform to what has been expected. The evaluation cycle is closed by feeding back the insights thus gained into further theory development. The realist evaluation cycle can be repeated as often as necessary to establish a robust programme theory. There is no rule that the same kinds of data and analytical methods should be used each time; on the contrary, realist evaluation researchers advise against it (Mark et al. 2000; Olsen 2010). Pawson & Tilley advocate the theory-guided use of quantitative and qualitative research strategies, contemporaneous and historical time scales, cross-sectional and longitudinal viewpoints, small and large samples etc. (85).

3.2.4 Theory-driven methodical pluralism

In general, methodical pluralism, or mixed methods research, refers to the application of different methods within the same study. Often, the term is meant to denote a combination of quantitative and qualitative research, implying the use of different methods on different kinds of data. For Creswell & Clark (2007), proper pluralism needs to go beyond methodical eclecticism:

‘Mixed methods research is a research design with philosophical assumptions as well as methods of inquiry. As a methodology, it involves philosophical assumptions that guide the direction of the collection and analysis of data and the mixture of qualitative and quantitative approaches in many phases of the research process. As a method, it focuses on collecting, analysing, and mixing both quantitative and qualitative data in a single study or series of studies. Its central premise is that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone.’ (5, original emphasis)

Methodical pluralism is espoused not only by the realist paradigm, but also by evaluators working in the pragmatist and the transformative traditions (Mertens &

Wilson 2012). Within each paradigm, the use of mixed methods has a different purpose and follows a different logic. In transformative or critical evaluation different methods are often combined in a dialectical manner not only to expose issues of power and justice from different angles but also to create awareness of the historical and contextual factors responsible for them. The pragmatist evaluation paradigm stresses the importance of getting evaluation results used (Patton 2012). Methods are chosen and mixed in such a way that the intended users of the results will understand, appreciate and actually apply them. In Realistic Evaluation, the mixing of different methods serves the sole purpose of confirming, falsifying and especially refining theory. Therefore, methodical decisions are not led by data or by pragmatic considerations but by the theory to be evaluated.

In single-cycle studies, methodical pluralism is largely constrained to concurrent mixed methods designs such as triangulation. Sequential mixed methods designs, where the use of one method informs the use of another method applied subsequently (and potentially to different data), are unsuitable for single-cycle evaluation studies. Mixing methods sequentially is better done across different cycles, so that the findings of an earlier cycle may not only inform the choice of methods and data in the next cycle but also the further development of programme theory and the formulation of hypotheses.

3.3 Design of the GUUGLE programme evaluation

This section outlines the design of the GUUGLE programme evaluation, building on elements of Realistic Evaluation and using Robson's (2002) framework for research design in Figure 3.5 as a blueprint.

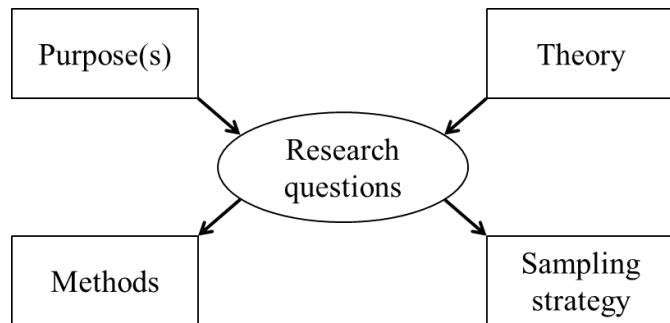


Figure 3.5. Framework for research design (Robson 2002, 82).

The purpose of this study is to conduct a formative evaluation of GUUGLE. The theoretical framework, including theories of adaptive coping, OD and organisational learning as well as critical realism and the theoretical foundations of Realistic Evaluation, has largely been expounded in the present and the previous chapters. What remains to be introduced once the GUUGLE programme itself has been outlined, is its programme theory. The research question of this thesis requires evaluating GUUGLE with regard to its fitness for the purpose of putting an end to pedagogical solitude.

Different methods of data analysis will be embedded in three interconnected realist evaluation cycles, which are depicted in Figure 3.6. The first cycle takes the GUUGLE programme theory as its starting point, breaks it down into a set of hypotheses and applies statistical path analysis to quantitative participant and survey data to test and refine the hypotheses. Each subsequent cycle builds on the findings and insights of the previous one. The second cycle uses various kinds of content analysis on GUUGLE-related documents. The third cycle is based on realist interviews with participants and non-participants of GUUGLE who are chosen on the basis of observations made in the first two cycles.

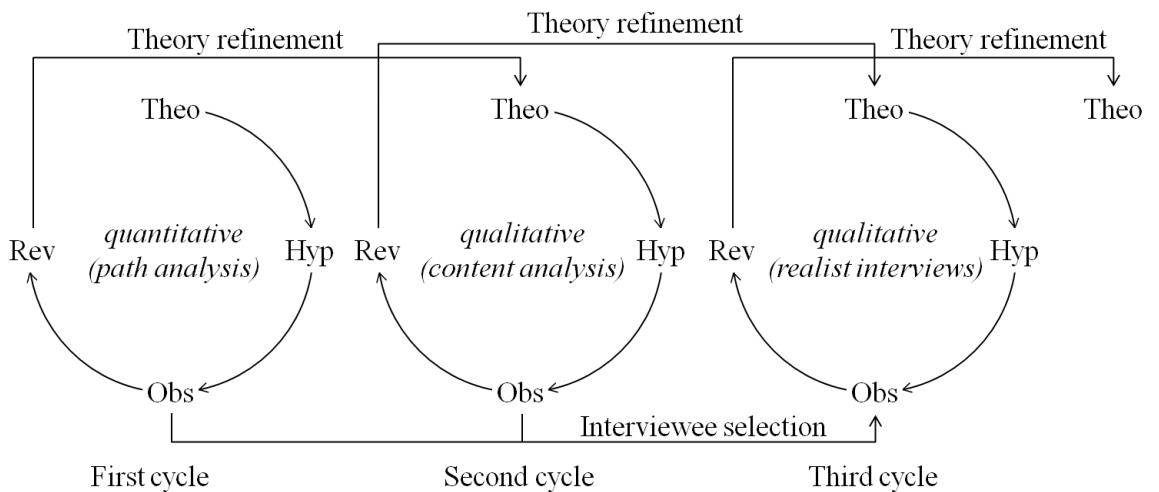


Figure 3.6. Explanatory sequential evaluation design.

Together, the three cycles form an explanatory sequential design. Explanatory designs are characterised by a quantitative research phase, which serves to identify significant, surprising or otherwise relevant patterns in the data, and a second phase of qualitative research aimed at explaining those patterns (Creswell & Clark 2007, 71-75).

The sampling strategy, finally, is fairly straightforward. Most of the data take the form of documents which have been published online, distributed on paper, circulated by email or presented at GUUGLE events. These documents have been or are still available to all members of my home institution. Material used for the evaluation which has not been made accessible to a wider audience consists of data gathered for the purpose of programme management: GUUGLE participant lists and workshop minutes, the notes of 18 interviews conducted in 2011 with colleagues to obtain feedback on GUUGLE, the responses of an anonymous online survey about GUUGLE of 2012 and survey data which I collected from colleagues in 2008 for a social network analysis. All interviewees and the participants of the social network analysis had given their

voluntary informed consent and were guaranteed anonymity. The participant lists and workshop minutes contain people's names, but they will only be analysed quantitatively with no references to specific individuals or small groups.

So most of the data used for the evaluation are by-products of running the GUUGLE programme. The use of an existing data pool has the distinct advantage of excluding strategic or otherwise biased behaviour on the part of the research subjects who might want to influence certain research outcomes. Moreover, since each analysis will be based on all available data of a certain type, the researcher's selection bias can be avoided. And given the kinds of data used, ethical issues are limited to the question as to whether data gathered for the purpose of programme management can be used for programme evaluation.

3.4 Summary

Critical realism posits that events are generated by unobservable mechanisms which have the tendency to be triggered in certain contexts. The effectiveness of a social programme depends on its ability to trigger those mechanisms that are necessary to achieve the desired outcome. Implicit in a programme is a theory of what those mechanisms may be, why they are believed to be necessary and how they are to be actualised. In Realistic Evaluation, programme theory consists of statements about the configuration of three elements: context (C), mechanism (M) and outcome pattern (O). The evaluation of a social programme is therefore the empirical testing (and refining) of the CMO configurations on which the programme is based. Programme theory also guides methodical choices including the combined use of different methods and types of data. For the GUUGLE programme evaluation, a explanatory sequential mixed-methods

research design will be adopted, which starts with a quantitative evaluation cycle to identify potentially significant patterns, followed by two qualitative cycles aimed at explaining those patterns.

Chapter 4 The GUUGLE Programme and Its Context

This chapter finally introduces the GUUGLE programme and its context. Both will be given equal attention, since mechanism and context are equally important for generative explanation. Unfortunately, ‘contextual layers are infinitely complicated, intertwined and in motion’ (Pawson 2013, 37) and there is no way of knowing *a priori* which contextual conditions are decisive for triggering a hypothesised mechanism or for the success of GUUGLE. The best this chapter can do when introducing context is to highlight aspects which appear to be directly relevant to the programme and distinctive about its setting. Making an educated guess about relevant context features is not dissimilar to retrodiction, the inference from events with respect to their antecedent conditions (see section 3.1.4).

4.1 A context marked by pedagogical solitude

GUUGLE emerged from, and in response to, a context marked by pedagogical solitude. Being a cultural condition of organisation and not an empirical phenomenon, pedagogical solitude can only be detected indirectly by means of indicators suggesting its presence. Two indicators will be used below which also provide insights into the social context of GUUGLE: social network structure and discourse.

Pedagogical solitude arises in certain institutional contexts. Reports about pedagogical solitude have come from HE institutions in Germany and elsewhere (see chapter 1), but so far no research into the institutional properties conducive to it seems to have taken place. The presentation of GUUGLE’s institutional context will therefore be guided by conjectures about what those properties may be.

4.1.1 Institutional context: a German Fachhochschule

The GUUGLE programme is at home at the *Hochschule Bremerhaven*, a medium-sized public *Fachhochschule* with over 3,000 students enrolled in 30 undergraduate and graduate degree programmes taught by 69 professors and a triple-digit number of part-time lecturers. All non-administrative senior managers are also professors. There are two faculties, one for the engineering sciences and one for the economic sciences.

Fachhochschulen are non-university HE institutions which Klumpp & Teichler (2008) view ‘in comparative perspective, as one of the most ambitious alternative types of higher education institutions to the universities in Europe’ (119). Most *Fachhochschulen* were established in West Germany between 1969 and 1975 by upgrading existing technical or vocational colleges. The *Fachhochschulen* concentrate on engineering and other technical disciplines, business studies, social work and design. Unlike universities, *Fachhochschulen* focus on teaching rather than research. Teaching facilities tend to be better and classes are typically much smaller than at the universities. Research at *Fachhochschulen* is applied rather than basic and often takes place in collaboration with regional businesses.

Between 1991 and 2013, the *Fachhochschulen* absorbed the bulk of the increase in student numbers in Germany. Today, about 800,000 students are enrolled at 215 *Fachhochschulen*, representing 32% of the German student population (Destatis 2013). The Bologna process has strengthened the *Fachhochschulen* in relation to the universities. They now award the same undergraduate and graduate degrees. Doctorates, however, may still only be awarded by universities.

At the *Fachhochschulen* practically all full-time teaching staff are professors. In Germany, professors are senior academics with civil servant status and life tenure. Formally, they are all on the same hierarchical level and no professor who is not a rector or dean has the authority to give instructions to other professors (see section 2.1.3). The civil servant status in combination with the constitutionally guaranteed academic freedom makes it difficult even for rectors and deans to enforce instructions or sanctions on professors. Also the possibilities to reward professors are very limited. Leadership therefore depends on consensus and tends to be weak. Together, these factors are at least partly responsible for the *Hochschule Bremerhaven*'s nature as a loosely coupled system, professional bureaucracy and organised anarchy, as described in section 2.3.1.

Applicants for university professorships must hold a doctorate and have demonstrated their special aptitude for research at post-doctoral level over several years. At the *Fachhochschulen*, the post-doc research requirement is replaced by the requirement that applicants must exhibit particular achievements with regard to the application of scientific findings and methods in several years of professional activity outside the university (Article 44, HRG 2007).

While university professors in Germany teach about nine hours per week and can revert to research assistants and secretarial services, most professors at *Fachhochschulen* have teaching obligations amounting to 18-19 hours per week with virtually no academic personnel supporting them. Teaching loads can only be reduced after sufficient external funding from research grants or contracts has been secured. A survey I conducted among 259 professors at *Fachhochschulen* (Vogel 2009b, see section 1.1) showed that they regarded the high teaching obligation as the biggest threat to their professionalism,

as it renders research difficult and, especially at smaller *Fachhochschulen*, forces them to teach subjects outside of their fields of expertise.

4.1.2 Social context I: network structure and collegial exchange

For a social network analysis I conducted in 2008 at the *Hochschule Bremerhaven* to study the extent of collegial exchange and pedagogical solitude, I mapped various communication and influence networks among professors (Vogel 2009a). 40 of them completed a questionnaire in which they stated how often they normally communicated with each of the other 39 participants in general, about the contents and the methods of their teaching and how strongly their 39 colleagues had influenced their teaching, respectively. The analysis showed that the *Hochschule Bremerhaven*'s teaching focus was hardly reflected by its professors' communication patterns.

As argued in section 1.1, pedagogical solitude is a learning obstacle for individuals and entire organisations. For a relation between teaching staff to be conducive to collegial learning it must go beyond an occasional chat. In the social network analysis I therefore concentrated on relations which promised to foster learning. Based on Granovetter's (1973) distinction between strong and weak ties, I defined as a strong tie a relation between two professors which involved at least weekly communication, at least monthly communication about teaching contents and methods and at least some influence of one professor on the other's teaching in the past. By contrast, I considered a relation to be a weak tie if it involved at least monthly communication about teaching contents or methods or both.

Figure 4.1 depicts the graph of the combined strong-tie and weak-tie networks. Bold lines indicate strong ties. Thin solid lines represent weak ties between professors communicating about the contents *and* the methods of their teaching. Dashed lines and dotted lines depict weak ties focusing *either* on teaching contents *or* on teaching methods but not both. Little circles and triangles represent members of the *Hochschule Bremerhaven*'s engineering faculty and economic faculty, respectively. The spatial distribution of nodes in the graph results from a scaling algorithm. Proximity of two nodes reflects their similarity in the sense that they share similar shortest paths to all other nodes. The large red ovals, finally, highlight the existence of four network clusters, i.e. sub-graphs of significantly higher density than their environment.

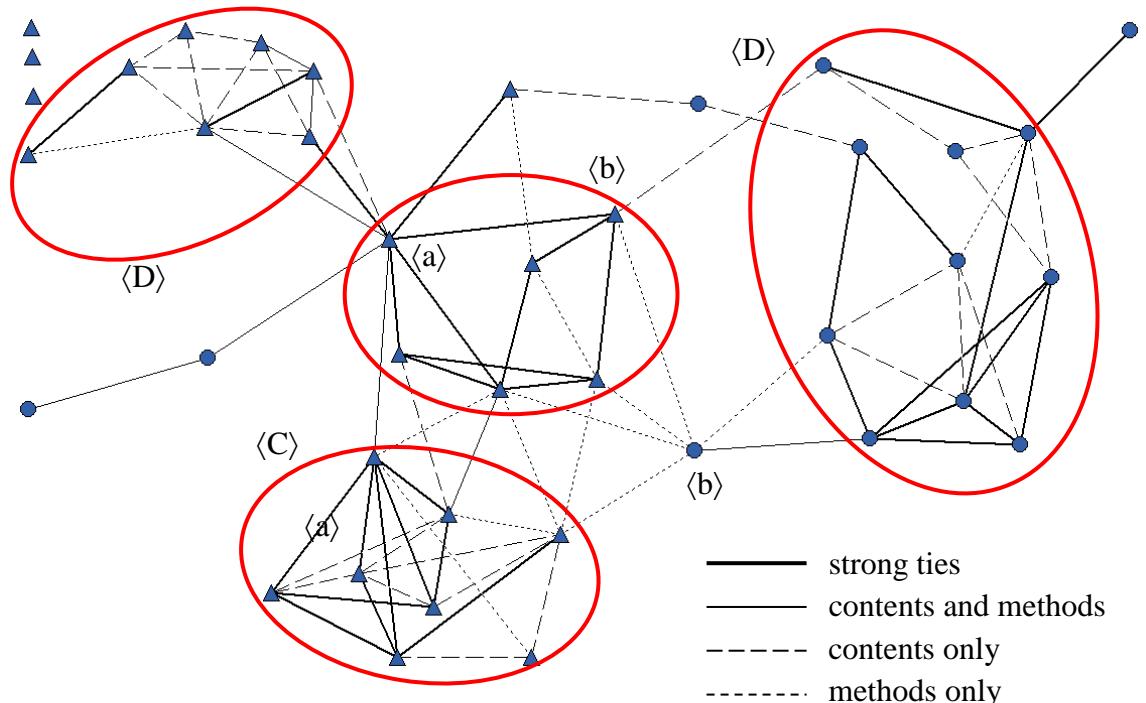


Figure 4.1. Strong-tie and weak-tie networks of 40 professors (Vogel 2009a, 28).

Of the 780 possible ties, 34 or 4% were strong and 51 or 7% were weak. The combined density of 11% was rather low considering the fairly small number of professors at the

Hochschule Bremerhaven. Yet one might argue that pedagogical solitude only concerned a minority of professors. For example, only 20% of the survey respondents had two or less ties within the sample. But the picture changes once the network structure is taken into account. 38% of the respondents shared no ties with members of other clusters, confining them to local knowledge exchange, which tends to be characterised by limited cognitive distance and informational redundancy. According to section 1.1, the limitation of CETL to direct colleagues within a small, tight-knit group is a form of pedagogical solitude. Moreover, by removing ⟨a⟩ and ⟨b⟩ from Figure 4.1, the network would lose five strong and ten weak ties and disintegrate into the three effectively sealed clusters ⟨A⟩, ⟨B⟩+⟨C⟩ and ⟨D⟩. Hence much of the network-based potential for cross-fertilisation, innovation and thus learning within this sample depended on just two professors.

4.1.3 Social context II: discourses about teaching and learning

I became interested in the causes of pedagogical solitude and the factors shaping the social network structure among the professors at the *Hochschule Bremerhaven*. To my surprise, with only few exceptions, colleagues and also academics from elsewhere tried to explain and rationalise the status quo when I involved them in conversations about the (from my perspective) underdeveloped culture of CETL. It seemed normal to them that academic teaching was a relatively isolated, privatised activity. Even those who responded to my questions with cynicism or resignation had accepted the normality of pedagogical solitude.

Critical theory warns that the perception of normality may result from a dominant ideology which maintains and protects a particular status quo. ‘Ideology does this by

convincing people that existing social arrangements are naturally ordained and obviously work for the good of all' (Brookfield 2005, 41). This insight induced me to carry out a critical analysis of my colleagues' discourses on teaching, learning and CETL in order to identify possible ideological and power-related causes of pedagogical solitude (Vogel 2011). Based on in-depth interviews and observations I identified two categories of discourses which appeared to sustain the normality of pedagogical solitude at the *Hochschule Bremerhaven*: working conditions discourses and professional identity discourses. Figure 4.2 depicts both categories and their subcategories.

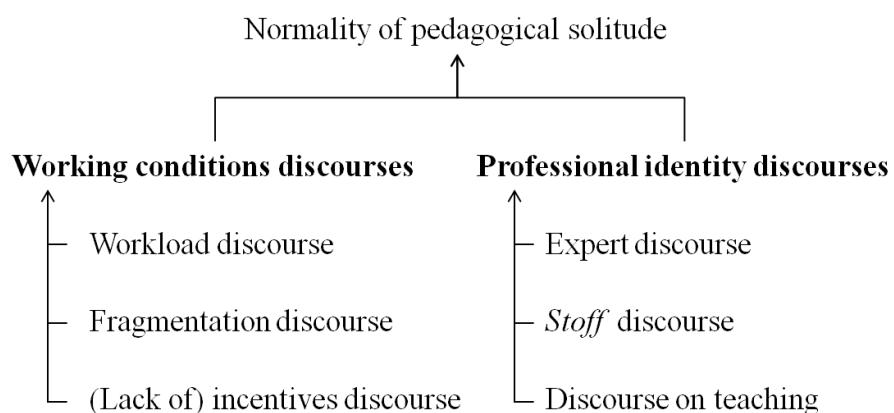


Figure 4.2. Discourses sustaining the normality of pedagogical solitude (ibid, 35).

By stressing excessive workloads, the fragmentation of academic working life and the lack of incentives, working conditions discourses represented the professors' working conditions as limiting rather than enabling and as *incompatible with CETL*, thus providing socially acceptable explanations and justifications of pedagogical solitude. Professional identity discourses, on the other hand, conveyed the impression that *CETL was not necessary*. The professors' dual professional identity as subject experts and academic teachers was clearly dominated by the former. For successful teaching, the

content aspect (in German: *Stoff*) was regarded as far more important than the pedagogical aspect, reflecting an orientation towards teaching rather than learning (see Samuelowicz & Bain 2001 for a characterisation of both orientations). Due to a fairly simplistic understanding of the teaching-learning nexus as a matter of information transmission, the professors showed a tendency to trivialise pedagogical issues and thus to devalue CETL.

Jointly the two categories of discourses established CETL as neither possible nor necessary, representing pedagogical solitude at least near-ideologically as naturally ordained (see Brookfield's above quotation) and perfectly acceptable. I interpreted pedagogical solitude as an expression of the prevailing social order in German HE, which is characterised by a division of responsibilities, according to which professors are responsible for teaching and students for learning. Unlike their UK counterparts, German professors, historically, have never had a responsibility for student learning. Until today, the failure of the student is the failure of the student, not the failure of the teacher. An emerging culture of CETL would undermine this social order by creating transparency and making questionable practices visible, by shifting professors' attention from teaching to learning, and by exposing discrepancies between naïve understandings and actual complexities of the teaching-learning nexus, revealing the traditional division of responsibility between professors and students as untenable.

4.2 The GUUGLE programme

My own early perceptions of pedagogical solitude at the *Hochschule Bremerhaven* and my research findings led me to consider possibilities for changing the status quo. An effective strategy for transforming it, I thought, would be to actively involve influential

colleagues in conversations on teaching and learning in order to challenge their beliefs about the possibility and importance of CETL, whilst simultaneously ‘injecting’ a continuous stream of contemporary ideas and controversies about teaching and learning into the organisation to disrupt its self-referentiality and show the need for CETL. So I started searching for approaches which would enable me to implement this strategy.

4.2.1 Professional learning communities

A promising approach I came across in the literature on school reform and teachers’ professional learning was the formation of professional learning communities (PLCs). Stoll et al. (2006) define a PLC as ‘a group of people sharing and critically interrogating their practice in an ongoing, reflective, collaborative, inclusive, learning-oriented, growth-promoting way’ (223). Stoll & Louis (2007) add that PLCs focus:

‘not just on individual teachers’ learning but on (1) professional learning; (2) within the context of a cohesive group; (3) that focuses on collective knowledge, and (4) occurs within an ethic of interpersonal caring that permeates the life of teachers, students and school leaders.’ (3)

The emphasis on teachers’ own practice, on collective knowledge and interpersonal caring seemed quite suitable to stir up a culture of pedagogical solitude. Conversations about one’s own practice reach deeper and promise a more lasting impact than conversations about teaching generally. Collective knowledge implies a set of shared concepts and a shared language to express them, both of which appeared necessary for fruitful CETL. And the concern for interpersonal caring reflects a more comprehensive, complex and human understanding of the teacher’s role than the strongly teaching-oriented, technicist understanding expressed by professors at the *Hochschule Bremerhaven* in interviews for the discourse analysis.

The size of PLCs varies between a handful of participants and over 15 (Richlin & Essington 2004). The first PLCs are typically initiated, convened and led voluntarily and informally by interested individuals. Positive learning experiences may encourage colleagues to start their own PLC, possibly with a different focus. Over time, many PLCs, especially in North America, became institutionalised as parts of teacher training or academic development programmes, with compulsory participation for new teaching staff (Cox & Richlin 2004).

But it was the personal, engaged and elaborate report of a small, isolated PLC at the University of Glasgow, calling itself the critical professionalism project and comprising only six university teachers from different disciplines (Walker 2001), which showed me how a PLC at the *Hochschule Bremerhaven* could set an example of CETL for others. Upon the initiative of one of its members, and inspired by Barnett's (1997) concept of critical professionalism, this PLC had:

‘set out to explore our professional activity in education framed by an overlapping, interwoven patchwork of layers of discourse and practices:

- a framing context of higher education policy and discourses;
- collaborative and reflective professional dialogue in an interdisciplinary group of lecturers;
- action in classrooms;
- action research processes which sought to understand and improve curriculum practice;
- the development of a reflexive “model” of continuing professional development in higher education.’ (7)

The range of activities undertaken and documented by this group shows that what happens in a PLC is entirely up to its members’ own needs and ambitions. Given this openness, it is important that PLCs set themselves clearly defined goals, agree on a work plan and keep discipline.

4.2.2 Pilot PLC 2009-2010

In April 2009, my colleague Wolfgang Lukas and I announced a first PLC. We invited all professors of the *Hochschule Bremerhaven* to come, but we were particularly keen on getting certain informal multipliers and gatekeepers into the group. The social network analysis and Figure 4.1 proved extremely useful for identifying these key people who we then approached personally to convince them to join in. Although our Rector supported the PLC by granting small teaching load reductions to participants, from the start it was perceived as a ‘bottom-up’ self-help initiative by professors for professors.

I came up with the name GUUGLE, an acronym for *gut und gerne lehren & lernen*, which may be translated as ‘successful and enjoyable teaching and learning’. The memorable acronym and the colourful logo I designed stood out, attracted attention, won us sympathies and were probably instrumental in the further development of the initiative. In May 2009 the PLC kicked off with twelve participants. We put together a work plan for the following months. Until February 2010 the group met eight times for three hours to talk about aspects of our jobs, our teaching and the outcomes of classroom experiments and small student surveys we had previously agreed to perform. We documented our work in minutes and photo protocols.

The participant feedback on this pilot PLC was very positive. The atmosphere was relaxed and informal, conversations were fairly open and instructive, and the fact that no decisions were made allowed everyone to act less cautiously and strategically than in formal committee meetings. The only criticism participants uttered concerned the PLC’s work plan, which had been far too ambitious.

4.2.3 GUUGLE programme design

Encouraged by this first experience with a PLC, Wolfgang and I attempted to acquire external funding for a scaled-up version of GUUGLE. To counter the professional identity discourses ('CETL is not necessary') and working conditions discourses ('CETL is not possible'), our aim was to offer reasons (making CETL necessary) and resources (making CETL possible) for conversations on teaching and learning. We felt that students had an important perspective to contribute and should be accepted as conversation partners. Moreover, since we were critical of the usual approaches of academic development through teacher training and pre-packaged programmes, and because we wanted GUUGLE to be owned by the teaching staff and students rather than by some administrative unit, we defined four guiding principles for GUUGLE. All activities had to be (a) self-directed by the participants; (b) cooperative, involving at least three people; (c) reflexive, dealing with the teaching and learning of those involved; and (d) situated in the normal work and study context of the *Hochschule Bremerhaven*.

By chance, a major industry-backed German HE and science funding organisation issued a competitive call for proposals for change initiatives aimed at improving teaching and learning in HE. Under time pressure, Wolfgang and I wrote down our ideas and entered them for the competition. We proposed to:

- 1) launch and facilitate several PLCs with different foci;
- 2) encourage and support projects initiated and realised by student teams who wish to enhance the study and learning conditions at the *Hochschule Bremerhaven*;

- 3) organise an annual GUUGLE Forum, i.e. a conference on teaching and learning for staff and students;
- 4) communicate about teaching, learning and GUUGLE to internal audiences;
- 5) document and evaluate our activities.

The first three elements meet the above guiding principles (a)-(d). The first four elements were a direct response to the findings of my critical discourse analysis. With those elements we intended to counteract the dominant discourses which seemed to sustain the normality of pedagogical solitude. In particular, we wanted GUUGLE to challenge our colleagues' professional identities by confronting them with, and involving them in exchange about, their students' views on teaching and learning, alternative concepts of professionalism of teaching and the shift from the teaching paradigm to the learning paradigm (Barr & Tagg 1995).

The jury liked our proposal and awarded us half a million euros (£420,000) to run the programme for five years, starting in 2010. Through synergies with existing services and other programmes we later found a way to extend GUUGLE until 2016 without requiring extra resources. Had GUUGLE not already commenced in 2009, it could have been regarded as a direct response to Palmer & Zajonc (2010) whose call to renewal of HE carries the subtitle 'Transforming the Academy through Collegial Conversations'.

4.2.4 *GUUGLE in practice*

Between March 2010 and March 2013, eight PLCs took place, which were joined by a total of 36 of the 69 professors at the *Hochschule Bremerhaven*. Many of them participated in more than one PLC and several in more than three. Chapter 6 gives

details about the PLCs' participation patterns and possible impacts. All PLCs were facilitated by Wolfgang or me. The two most popular PLCs concentrated on teaching and assessment methods. One PLC looked at the changing teaching conditions at our home institution, and one focused on professorial self-concepts, autonomy and development. One PLC was interested in the challenges of supervising and evaluating student projects. One worked on a definition of good teaching and a teaching-related mission statement for the *Hochschule Bremerhaven*. Another one developed a new concept and process for the student evaluation of teaching. And one PLC was launched specifically for newly appointed professors. In response to demand arising from different PLCs, a large number of teacher training seminars and workshops with external experts on a range of topics were organised by the GUUGLE team which had grown once funding had become available. Figures A1 a-d in Appendix A1 show impressions from PLC and workshop sessions.

Student projects were more difficult to get going than PLCs. Assuming students to have at least an intuition about how their own learning and study conditions might be enhanced, we invited student teams to submit proposals for improvements they wished to bring about. The students were offered funding, advice and credit points upon completion of their projects. However, only 14 projects took place between March 2010 and March 2013. The students' perspective change from 'recipient' to 'actor' in their education proved more challenging than anticipated. All completed projects were publicised and some projects were also presented at GUUGLE Forums, triggering further discussions.

Every autumn, the GUUGLE Forum conference brings together teaching staff, students and external guests to learn about, share, reflect on and discuss a particular aspect of teaching and learning. The conference agenda combines project presentations, workshops, exhibitions, guest lectures, panel discussions, training sessions and ample opportunity for informal social interaction. Each GUUGLE Forum is preceded by weeks of extensive communication. Between 2010 and 2013, participant numbers receded from 160 to 100, probably owing to habituation effects. Figures A2 a-d in Appendix A1 depict scenes from GUUGLE Forums.

Communication about teaching, learning and GUUGLE-related topics has been an essential part of the programme from the beginning. The purpose is not only information and promotion, but also the professors' continuous exposure to ideas about and controversies around teaching and learning, in order to disrupt their dominant professional identity discourses (see section 4.1.3) and demonstrate the need for CETL.

The GUUGLE programme's communication media include a printed newsletter, circular emails, a website, posters, flyers, postcards, roll-up displays, presentations by the GUUGLE team as well as announcements by GUUGLE ambassadors (students and teaching staff acting as multipliers) in class. For external communication, members of the GUUGLE team presented the programme at over a dozen conferences and published various articles about the programme (e.g. Lukas et al. 2013; Ritzenhoff & Vogel 2012).

Figures A3 a-c in Appendix A1 show examples of GUUGLE communication media.

For documentation and evaluation purposes, we collected written feedback from participants of most GUUGLE events immediately after they had taken place, conducted a series of interviews in 2011 with participants and non-participants on their

perceptions of GUUGLE and administered an online survey on GUUGLE to all teaching staff in 2012. The summative evaluation of individual activities and the gathering of formative feedback have been an integral part of GUUGLE from the beginning.

4.2.5 GUUGLE-X

In response to demands and suggestions which had been articulated in PLCs and at the first GUUGLE Forum, colleagues and I put together a further funding proposal one year after we had received the money for GUUGLE. We wanted to use the momentum GUUGLE had created as an opportunity to launch further teaching-related projects and merge this new project portfolio with the GUUGLE programme. The projects we proposed concerned infrastructure (e.g. upgrades for e-learning and e-assessment systems and the development of a diagnostic self-assessment tool for prospective students), programmes (e.g. an enhanced induction and training programme for new teaching staff; a prolonged, practice-oriented induction phase for all new students; systematic preparation and training of teaching assistants) and management (e.g. implementation of a data-based quality assurance cycle at the level of individual degree programmes). Again our proposal happened to be successful, and we received 2.1 million euros (£1.8 million).

To distinguish GUUGLE, which started in 2009, from the portfolio of projects added in 2011, the latter will be referred to as GUUGLE extension or GUUGLE-X. For students and most staff, GUUGLE and GUUGLE-X are indistinguishable since both sail under the GUUGLE flag. However, for the present study the distinction and relationship between them matter for three reasons. Firstly, the present evaluation focuses on

GUUGLE only. Secondly, the additional projects of GUUGLE-X are putting a substantial strain on management and key staff and compete with GUUGLE for time, energy and attention. And thirdly, while GUUGLE is funded by a private-sector organisation which grants significant freedom regarding the use of the resources provided, GUUGLE-X is paid for by taxpayers' money and must comply with rigid bureaucratic and accounting regulations. The perception of GUUGLE-X as a slow tanker has spilled over to the agile speedboat GUUGLE, as interviews and surveys have shown. In other words, GUUGLE-X is changing the social context of GUUGLE.

4.3 GUUGLE as adaptive coping facilitator

At a more abstract level, the intended impact of GUUGLE as an OD programme which stimulates and sustains CETL can be illustrated by means of Schein's adaptive coping cycle from section 2.1. The cycle is depicted again in Figure 4.3, but now boxes indicate the cycle phases in which GUUGLE is intended to intervene and make organisational coping with environmental change more effective.

The sensing of changes in the external environment (phase 1) is supported by the GUUGLE Forum conferences, occasional seminars and regular teacher training workshops run by external experts. These events inject information about recent developments in the HE sector, research-based perspectives on learning, new teaching-related ideas etc. into the organisation. GUUGLE also supports the sensing of changes in the internal environment of the *Hochschule Bremerhaven*, including those triggered by GUUGLE activities (phase 5), by registering them when they come up as topics in PLCs or are addressed by students in their self-directed projects. Due to the institution's

selective sensing and blind spots (see section 2.3.2), many of these ideas, topics and developments might otherwise have gone unnoticed.

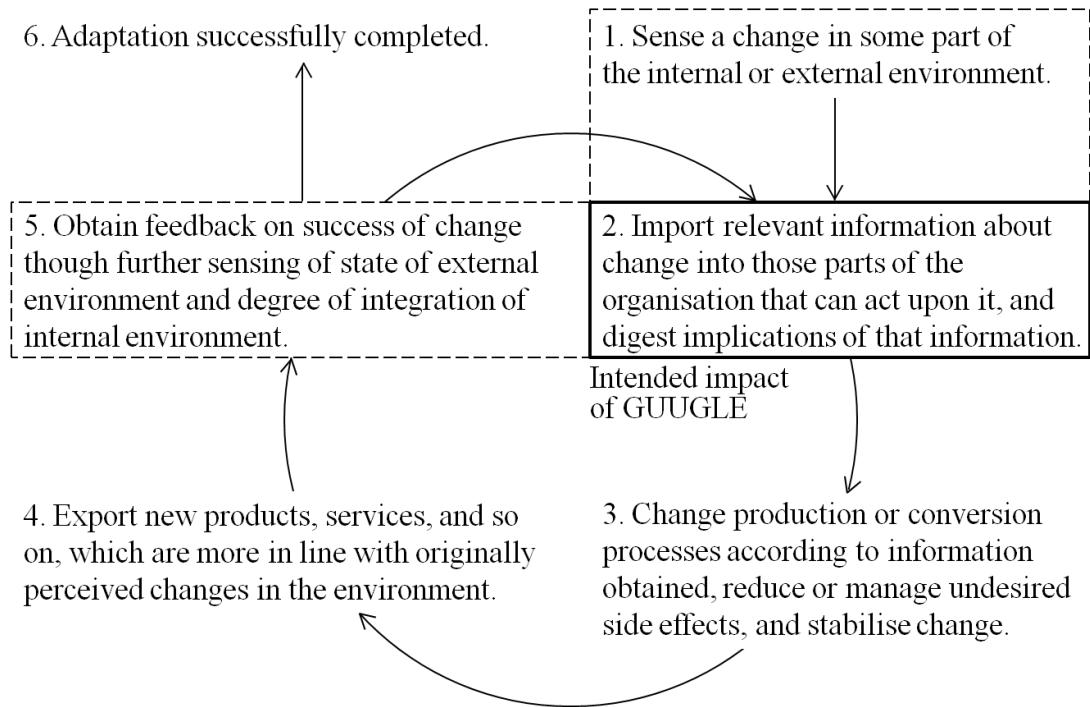


Figure 4.3. Intended impacts of GUUGLE on the adaptive coping cycle.

With its particular emphasis on fostering the professors' CETL, facilitating phase 2 of the adaptive coping cycle is the key concern of GUUGLE. Sharing, interpreting and reflecting on information about external and internal environmental changes constitutes an important part of what happens in the PLCs. The GUUGLE Forum conferences serve a similar purpose by internally disseminating information and providing a platform for conversation and reflection. And students' project initiatives and outcomes stimulate collegial exchange about student perceptions, needs and capabilities. Moreover, as argued by Schein (1993), the main coordination problem in complex organisations is integration across subcultures through the development and learning of an overarching

common language and mental model (see section 2.2.2). This problem is evidenced at the *Hochschule Bremerhaven* by the professors' fragmentation discourse (see Figure 4.2). With their cross-disciplinary composition of participants and choice of topics, the GUUGLE Forums and PLCs contribute to the organisation's integration across subcultures.

Phase 3 of the adaptive coping cycle, which involves deciding on how to respond to environmental changes and implementing those decisions, is the main concern of GUUGLE-X. So far, GUUGLE activities have only rarely resulted in new processes in phase 3. In the formative evaluation of GUUGLE, this apparent lack of problem-solving capacity and operational impact will emerge as a potentially critical issue for the sustainability of the GUUGLE programme.

Finally, even though the ultimate aim of GUUGLE is to improve the quality of teaching and learning, the programme does not (and cannot) intervene in phase 4 of the adaptive coping cycle where the actual 'teaching product' and 'learning outcome' are generated. While GUUGLE-X provides various new services directly to prospective and current students, GUUGLE adheres to the OD tradition of intervening only indirectly by targeting communication habits, belief systems and culture.

4.4 Summary

The purpose of this chapter was to introduce the institutional and social context of GUUGLE and to outline the evolution, structure and intended organisational impact of the programme itself. The institutional context is shaped by the professors' high degree of professional autonomy, their life tenure as civil servants and the associated

limitations for leadership and enforcing cooperation. The social context before GUUGLE was characterised by fairly sparsely connected and fragile networks of strong and weak ties among the professors as well as by a set of discourses sustaining a normality of pedagogical solitude.

Starting as a small, informal PLC for professors, GUUGLE evolved into an OD programme comprising several PLCs, GUUGLE Forum conferences, student projects and extensive internal and external communication. By targeting the attitudes and communication habits of individual professors, GUUGLE addresses the organisational culture of pedagogical solitude which limits the *Hochschule Bremerhaven*'s capacity for adaptive coping with environmental change. The seeming success of GUUGLE formed the basis of GUUGLE-X, a much larger portfolio of projects which have started impinging upon the contextual conditions of GUUGLE.

Chapter 5 GUUGLE Programme Theory

Normative programme theory clarifies the assumptions about how and why an intervention programme is expected to work. Ideally, programme design, implementation and evaluation are guided by the same theory to ensure the consistency of those activities and help programme management and stakeholders develop shared mental models. The programme theory outlined in this chapter, however, is the result of theorising GUUGLE *ex post*, as its design was guided more by tacit knowledge and intuition about what was necessary for and acceptable within the institutional and social context of the *Hochschule Bremerhaven* than by formal theory or standard OD programme designs.

The challenge of the present chapter is to verbalise and visualise the implicit GUUGLE programme theory, including its assumptions about the contexts, mechanisms and outcomes, in such a way that it can subsequently be used for the realist evaluation of GUUGLE without further adjustments. The following sections are therefore dedicated to the development of a suitable framework and its adaptation to GUUGLE. The framework itself will be quite general and involve a number of modifications to the Realistic Evaluation approach.

5.1 Choosing a realist framework

Pedagogical solitude has earlier been characterised as an organisation cultural condition. As such, it forms part of the social context in which individuals decide and act. GUUGLE is intended to transform this cultural condition into a different one, a condition under which it is normal to have collegial conversations about teaching and

learning, where ideas and experiences are shared and collectively reflected upon and where teaching is a ‘community property’ (Shulman 1993) rather than private affair. So addressing pedagogical solitude could be seen as a pure macro-level social problem.

5.1.1 Coleman’s social macro-micro-macro scheme

Coleman (1987; 1990) agrees that social outcomes may result from social conditions, but he stresses that since social conditions do not have the agency to bring about change, the causal chain passes through the micro level of individual conditions and individual outcomes. In other words, there can be no change at the macrosocial level which is not caused by change in the microsocial level of individuals’ choices and actions. To illustrate this view, Coleman devised a scheme of the social macro-micro-macro link, nicknamed ‘bathtub’, which is depicted in Figure 5.1. Bhaskar’s TMSA (see section 3.1.3) addresses the same issue of macro-micro-macro causal chains and offers a similar answer. Even his diagrammatic representation (see Figure 3.2) resembles Coleman’s.

The Coleman bathtub involves three mechanisms, which have been referred to as situational, action-forming and transformational (Hedström & Swedberg 1998). Situational mechanisms operate from the macro to the micro level. Through them, specific social conditions or events shape the conditions under which individuals decide and act. Individual conditions forming under the combined influence of social conditions, individual disposition and biography include values, beliefs, orientations, motives, resources, capabilities and perceived opportunities. Action-formation mechanisms are located at the micro level of the individual and describe the psychic processes leading from combinations of individual conditions to specific actions.

Transformational mechanisms, finally, operate from the micro to the macro level through the interaction of a number of individuals, generating collective or social outcomes.

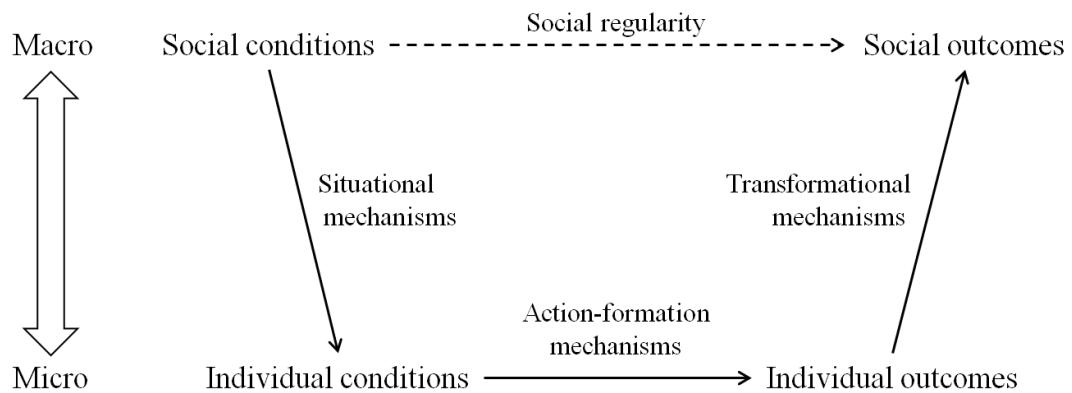


Figure 5.1. Coleman's bathtub (adapted from Coleman 1986, 1322; Hedström & Swedberg 1998, 22).

On the basis of this scheme, the impact of a social programme such as GUUGLE can be modelled by assuming that the programme alters the subjects' social conditions, for instance by offering new resources, prospects or other incentives. In the light of their individual conditions, the subjects consider the new social conditions and choose to react to them in a certain way. The pattern emerging from the typically interdependent individual outcomes forms the programme's social outcome.

5.1.2 *Bathtub versus rugby ball*

In many ways, Coleman's bathtub and Pawson & Tilley's (1997) context-mechanism-outcome 'rugby ball' (74) diagram depicted in Figure 3.3(a) resemble each other. Both serve to explain a social regularity which is caused by generative mechanisms operating

in different strata of reality⁴, rooted in individuals' choices and triggered under certain contextual conditions. But there are also two major differences between the schemes.

A comparison of Figure 3.3(a) and Figure 5.1 shows that the social regularity in the former is depicted as separate from its context, whereas in the latter the regularity emerges from the context (i.e. of the social conditions). Pawson & Tilley's distinction between context and regularity is not ontological; it serves the expositional and analytical purpose of foregrounding a regularity of interest. Yet it is only through the separation of regularity and context that Pawson & Tilley can maintain that 'in most social programs there is no significant expectation that the prevailing contextual conditions will be transformed' and that change occurs '*within* broadly the same contextual conditions' (76, original emphasis). In Coleman's bathtub scheme, by contrast, change does not occur *within the same*, but *due to changing* contextual conditions which trigger individual reactions.

This difference results from different concepts of context. For Pawson & Tilley, context refers to the circumstances under which a certain programme operates. Context interacts with the programme by enabling or disabling mechanisms introduced by it, but it does not comprise the programme, its mechanisms and the regularities they cause. In short, context in the rugby ball diagram is the *context of a programme*. Coleman, on the other hand, understands context as the social and psychic conditions shaping individuals' action-formation processes. A programme's effectiveness depends on its ability to alter

⁴ In Figure 5.1 the micro and macro levels correspond to Bhaskar's strata of reality (see section 3.1.3). Figure 3.3(a) does not explicitly distinguish between social and psychic mechanisms but elsewhere, Pawson & Tilley refer to micro mechanisms triggered by reasoning and to macro mechanisms actualised by the availability of resources (67).

those conditions so as to trigger certain responses. In other words, context in the Coleman bathtub is the *context of programme subjects*. Since the programme subjects' context includes the programme context as well as the programme, Coleman's concept is more comprehensive.

The second major aspect that distinguishes the Coleman bathtub in Figure 5.1 from the rugby ball diagram in Figure 3.3(a) is the degree of visual detail. Where the latter summarises context by a single line⁵, the former depicts social and individual conditions separately and defines a mechanism connecting them. Where the latter limits itself to a single mechanism that gives rise to a regularity, the former analytically decomposes the same mechanism into three and assigns them specific functions. Where the latter needs words to describe the ontic level (micro/psychic, macro/social) on which a mechanism operates, the former speaks through its design. And finally, although Pawson & Tilley claim that 'generative mechanisms [...] *constitute* the regularity; they *are* the regularity' (67, original emphasis), it is not their rugby ball diagram but Coleman's bathtub that visualises the coincidence of mechanisms and regularity.

To conclude, the main advantage of Pawson & Tilley's scheme is its very high degree of abstraction and the (apparent) simplicity that comes with it. The scheme allows formulating numerous alternative CMO configurations rather quickly because not much time needs to be spent on working out detailed causal chains. Coleman's scheme, on the other hand, offers the possibility to model programme theories with more complex internal structures. It also provides more guidance to evaluators by forcing them to think

⁵ In Pawson et al. (2005, 23), the oval is depicted as comprising four layers of context: individual, interpersonal-relational, institutional and infrastructural. See also Pawson (2013, 37).

through the macro-micro-macro link. Finally, by adopting a programme subject's perspective on context, the difficult separation of the programme and its context is avoided. In the following, Coleman's bathtub will be adopted as a framework for the GUUGLE programme theory.

5.2 The GUUGLE bathtub

The bathtub diagram can be used to depict the pre-GUUGLE situation at the *Hochschule Bremerhaven* as well as the GUUGLE programme theory, corresponding to the two diagrams in Figure 3.3. In terms of this figure, the purpose of GUUGLE is to counteract mechanism M_1 with mechanism M_2 in order to replace the regularity R_1 with R_2 and thus to manage the transition from Figure 3.3(a) to (b).

Figure 5.2 depicts the pre-GUUGLE status. Certain HE traditions and working conditions have created social conditions at the *Hochschule Bremerhaven* which lead professors to believe that CETL is difficult (due to the working conditions) and unnecessary (due to the HE traditions). As my 2009 critical discourse analysis (see section 4.3) has shown, these beliefs were reflected in and reinforced by the working conditions discourse and the professional identity discourse.

Beliefs⁶ are individual conditions of action. As pointed out in section 3.1.3, individuals' belief systems are a source of reasons and, as Bhaskar stresses, reasons are causes (see section 3.1.2). The same idea is expressed by the action-formation mechanisms linking individual conditions with individual outcomes in Coleman's model. Therefore the

⁶ In Figure 5.2 and all following bathtub-type diagrams, individual conditions such as beliefs are printed in italics to indicate their location in the psychic stratum of reality.

professors' belief that CETL is difficult and unnecessary induced them to make no effort to get it started (self-fulfilling prophecy).

When such behaviour is not displayed occasionally by a few but regularly by many, it becomes a cultural trait. Culture is not only 'something that people do' (Alvesson 2013, 3) but also something that people refrain from doing. The transformational mechanism linking individual lack of effort to engage in collegial exchange (individual outcome) and the organisation cultural trait of pedagogical solitude (social outcome), finally, is straight-forward.

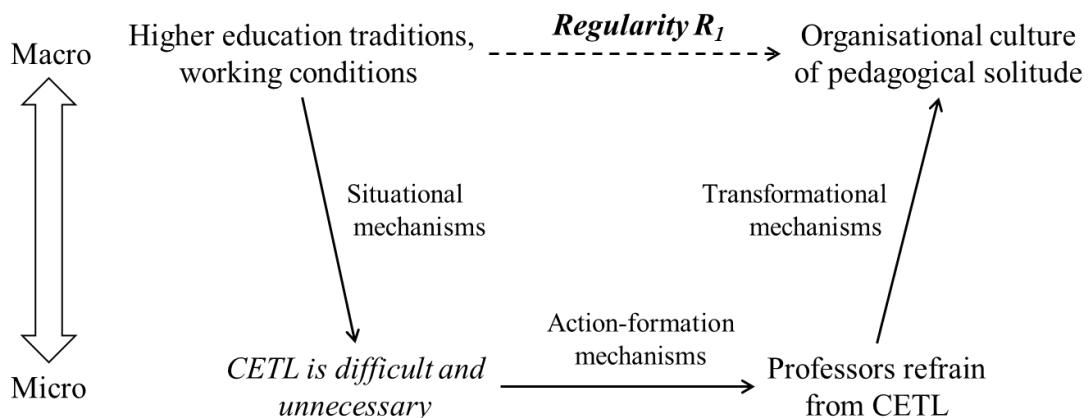


Figure 5.2. Coleman bathtub depicting the situation before GUUGLE.

The purpose of GUUGLE is to change the social conditions at the *Hochschule Bremerhaven* in such a way that the HE traditions and the working conditions which have been shaping the professors' individual conditions of action are counteracted and ideally superseded by alternative social conditions which are conducive to CETL. This is illustrated in Figure 5.3, which summarises the normative programme theory of GUUGLE. The programme itself is depicted as the new social condition, providing the

resources and reasons necessary for CETL. As Pawson (2002) notes, it is not the social interventions which change people's behaviour; 'rather it is the underlying reasons or resources that they offer subjects that generate change' (342; see also section 3.1.2).

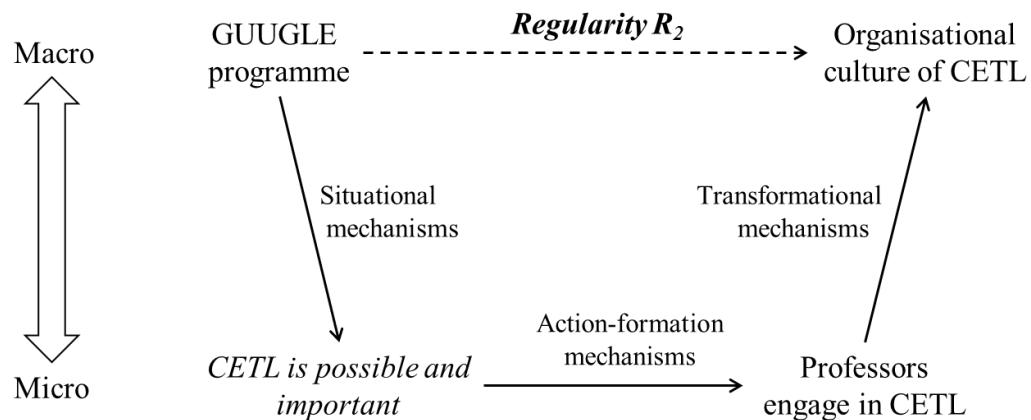


Figure 5.3. Coleman bathtub depicting the intended effect of GUUGLE.

Over time the professors learn that the exchange with colleagues is no longer difficult and unnecessary but has become possible and important. This new condition leads more and more of them to change their behaviour, seek opportunities for CETL and gradually turn such behaviour into a new norm, anchoring collegial dialogue and discussion about teaching and learning in the organisational culture.

Coleman's scheme has the advantage of summarising in a very abstract fashion the macro-micro-macro link involved in macro-level social change. However, as a model of OD and indeed of practically any programme theory, the Coleman bathtub has one major shortcoming. It includes no mechanism which would ensure the sustainability of the intended individual and social outcomes once the programme activities are phased out. The consequence of this omission becomes clear when the GUUGLE programme is

removed from Figure 5.3. All mechanisms collapse, all effects of GUUGLE vanish and the *Hochschule Bremerhaven* returns to its pre-GUUGLE state depicted in Figure 5.2.

5.3 From bathtub to whirlpool

The sustainability of the intended outcomes of GUUGLE (or any OD programme for that matter) requires homeostatic mechanisms, making their outcomes independent of the continuation of the GUUGLE programme. Homeostasis is the tendency of complex dynamic systems to maintain a relatively stable condition of their internal environment despite disturbances or variations in their external environments (Simon 1999). Homeostatic control is achieved through reinforcing and balancing feedback, i.e. circular mechanisms which counteract forces driving a system away from its homeostatic equilibrium. Since all arrows in Coleman's original scheme point from left to right, there can be no circular mechanisms. So to be usable as an OD framework, the bathtub needs to be modified.

Theoretically, GUUGLE ought to benefit from two feedback loops, which contribute to first raising and later stabilising participation rates and frequencies of CETL among the professors above the pre-intervention levels. The first feedback loop may be called 'CETL as an important practice'. Participation in GUUGLE-mediated CETL may be perceived as a personally beneficial or important practice for a variety of reasons, including the topics addressed, insights gained, quality of relationships developed with colleagues, communication skills acquired and the positive experience in general. Professors who come to regard CETL as an important practice are likely to engage in further collegial exchange, to talk about their experience (positive word of mouth) and encourage colleagues to join in. Once a critical mass of professors engages in CETL, a

bandwagon effect sets in and induces others to follow their example. The bandwagon effect is a social mechanism which describes people's behaviour in situations in which their participation decisions are influenced positively by the number of people already engaging in an activity (Granovetter 1978; Schelling 1978). But with a growing share of professors already engaging in CETL, it gets increasingly difficult to win over the remaining ones, and so the bandwagon effect dwindles.

This is when the second feedback loop kicks in to establish 'CETL as a cultural norm'. While word-of-mouth promotion of GUUGLE and CETL and the bandwagon effect involve conscious decision-making, the anchoring of CETL in the organisational culture of the *Hochschule Bremerhaven* through continued practice implies an internalisation of CETL, i.e. its acceptance as an integral part of a professor's role and professional identity. Behaving according to this norm is associated with positive social sanctions (e.g. esteem, respect, trust and regard as equal) by those adhering to the same norm. Feedback loop 2 is the mechanism through which CETL stabilises and reproduces itself in the long run. For the GUUGLE programme, normalising the professors' behaviour in such a way is certainly a stretch goal, not least because cultural change in organisations, and especially public HE institutions, is slow and perhaps far too slow for the limited duration of the GUUGLE programme.

Figure 5.4, which is an extension of Figure 5.3, visualises the refined normative GUUGLE programme theory. Its shape is a reminiscence of the Coleman bathtub, but the diagram now contains the two feedback loops, turning it into a whirlpool. The circular arrow symbols indicate directions of the feedback mechanisms. The idea is that once the two mechanisms have been activated, the GUUGLE programme can be phased

out, as CETL will have become self-sustaining. Word of mouth is part of feedback loop 1, but to highlight the loop's dual (individual and social) nature, a social link is indicated and labelled 'word of mouth'. Also the resources (time, space, expertise, coordination, facilitation, teaching load reductions and money) and reasons for CETL the GUUGLE programme needs to provide are stated explicitly to emphasise them: resources make CETL possible and reasons make it important.

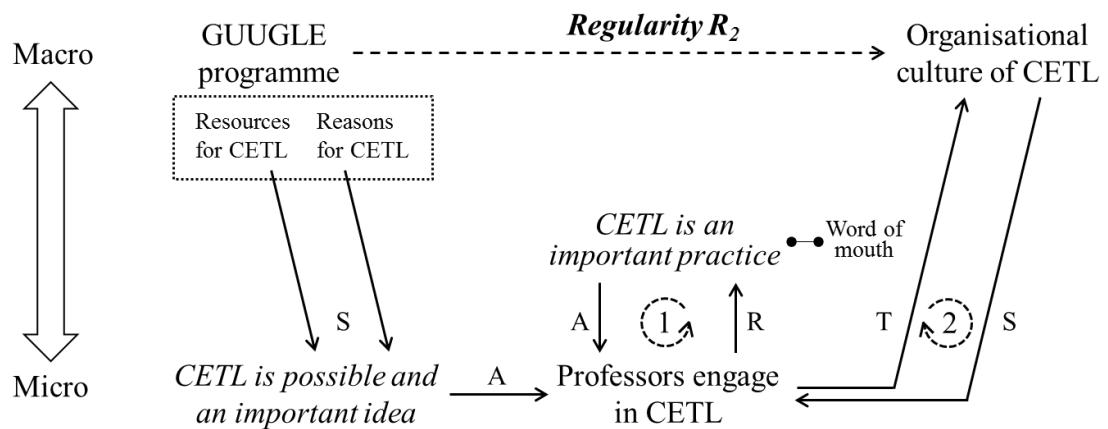


Figure 5.4. Whirlpool diagram of the GUUGLE programme theory.

The diagram distinguishes between CETL as an *important idea* and as an *important practice*. The latter is part of feedback loop 1 and characterises the individual condition (Coleman's terminology; see Figure 5.1) of professors actually engaging in CETL, whereas the former refers to an attitude based on abstract reasons for exchange provided and communicated by the GUUGLE programme (see section 4.2.4). Once professors realise the potential of CETL to drive professional and organisational development, they are much more likely to participate in PLCs and other GUUGLE activities.

Each solid arrow in the whirlpool scheme represents a situational mechanism (S), an action-formation mechanism (A), a transformational mechanism (T) or a reflexive mechanism (R). The latter is not an element of Coleman's original scheme, which has no need for it. Like the action-formation mechanism, the reflexive mechanism is located at the micro level and describes the psychic process through which people make sense of their own actions and experiences and evaluate them. The outcome of an activated reflexive mechanism is a new individual condition, which may or may not trigger further action.

Although the whirlpool scheme has been developed specifically as a framework for the GUUGLE programme theory, its applicability appears much more general. GUUGLE shares with other OD programmes the mode of action of interventions. As outlined in section 2.2, OD is concerned with system-wide deliberate change in organisations, i.e. with change at the social macro level, which OD attempts to achieve by altering the context in which organisation members operate. Dialogic OD, for example, aims to transform the social conditions that shape everyday thinking and behaviour 'by involving more and different voices, altering how and which people engage with each other' (Marshak & Bushe 2013, 1. See also section 2.2.2.). To anchor any new behavioural, attitudinal or other patterns in the organisation, homeostatic control mechanisms are necessary to ensure that the organisation does not return to its pre-intervention state as soon as an intervention is phased out.

These common features of GUUGLE and OD in general – altering the social (macro) conditions under which organisation members (micro) act in order to achieve organisational (macro) change and making this change sustainable through social

feedback – suggest that the whirlpool scheme in Figure 5.4 can be used as a universally applicable framework for theorising, planning, visualising and evaluating OD and similar types of social interventions.

5.4 Reconciling the GUUGLE whirlpool with CMO configurations

The GUUGLE programme theory as depicted by the whirlpool diagram differs in an important way from the programme theories that Pawson & Tilley use as examples for their evaluation approach. The standard CMO configuration has a one-shot structure with a single mechanism, whereas the whirlpool's internal structure involves a sequence of mechanisms and outcomes. It resembles the 'sequence of events that is expected to lead to a particular desired outcome' (Davies 2012) on which the Theories of Change approach to programme evaluation is based (Connell et al. 1995; Connell & Kubisch 1998; Weiss 1995. See also section 1.3.). Blamey & Mackenzie (2007) suggest that 'there is no obvious reason for believing that Theories of Change and Realistic Evaluation could not coexist within the one programme evaluation' (451) and Pawson (2003) acknowledges that

'If outcomes [of Theories-of-Change evaluations] are as intended, we have a very good idea that the programme was responsible, because we have tracked its inner *workings*. In action mode, we may observe a failure in a particular programme assumption [...]. If so, it can be identified, amended and re-tested for its ability to improve the programme.' (473, original emphasis)

Especially after Pawson's endorsement, borrowing from Theories of Change is the obvious strategy for reconciling the whirlpool scheme with the CMO configuration. For every mechanism in Figure 5.4, a separate CMO configuration can be defined. The outcome of each CMO configuration becomes part of the context of the subsequent

CMO configuration. This is made possible by the bathtub/whirlpool scheme for which context is that of the programme subjects, and not the programme's context (see section 5.1.2). To stress this aspect and make it transparent, the programme subjects' context can be decomposed into the *programme context* (C), which is familiar from the CMO configuration, and the *delta context* (D), which comprises the changes (=delta) introduced by a programme. Each mechanism along the whirlpool scheme's macro-micro-macro pathway thus operates within a different delta context, whereas the programme context C is assumed to stay broadly the same.

Table 5.1 is a variation of Pawson & Tilley's CMO configuration table. It shows the sequential DMO configurations or DMOCs representing Coleman's bathtub in Figure 5.1 and can easily be amended to match all kinds of whirlpool diagrams, including Figure 5.4.

Table 5.1. Sequential DMOCs representing the Coleman bathtub in Figure 5.1.

Delta Context (D)	+ Mechanism (M)	= Outcome (O)
Social context	Situational mechanism	Individual condition
↓		
Individual condition	Action-formation mechanism	Individual outcome
↓		
Individual outcome	Transformational mechanism	Social outcome

It should be noted that although the DMOCs form a sequence, the delta contexts are neither necessary nor sufficient to trigger the respective mechanisms; they only make them more likely to operate. This is consistent with Bhaskar's view that all mechanisms in open systems are only tendencies (see section 3.1.2).

5.5 Rival programme theories

The proposed GUUGLE programme theory in Figure 5.4 may be plausible, turn out to be empirically compatible with GUUGLE and able to explain all observations made in connection with it, but chances are that some other theory could do the same and better.

As Pawson (2012) himself stresses,

‘programmes never offer up a single theory. In realist terminology, there will always be multiple Ms – a proliferation of ideas within a programme, creating different resources that trigger different reactions among participants. There will always be multiple Cs – a huge range of individual, institutional and infrastructural features that condition the action of the assorted mechanisms. There will always be multiple Os – an extensive footprint of hits and misses, an uneven pattern of success and failure associated with the underlying causal dynamics.’ (184)

The trouble is that ‘the finiteness of our imaginative procedures always leaves an infinity of alternative conceivable hypotheses never even thought of by us’ (Sklar 1981, 17). So, in addition to the creativity of the inquirer, retrodiction involves the selection of a theory from an almost random subset of all conceivable theories. Moreover, since theory selection is based on a limited number of empirical observations, all of which can normally be accounted for by more than one theory (underdetermination thesis; see Quine (1975)), the retroductive process can turn out inconclusive.⁷

In the face of these complications, the space restrictions of this thesis and the availability of a programme theory which has been guiding the planning, decision-making, coordination and communication activities of the GUUGLE team for five years

⁷ The Duhem-Quine underdetermination thesis does not only pose a challenge to realist evaluation but to Bhaskar’s transcendental concept of causal power in general (e.g. Kaidesoja 2007).

as a shared mental model, an alternative way forward seems preferable. Instead of probing a range of rival theories, only the proposed theory will be tested. Observations not accounted for by it will be used as opportunities to critically review and refine the theory, rather than rejecting it in favour of another.

5.6 Summary

Using Coleman's macro-micro-macro 'bathtub' scheme as a starting point and adding homeostatic control mechanisms in the form of feedback loops, this chapter developed a normative programme theory for the GUUGLE programme. The feedback loops which turn the bathtub into a whirlpool are the precondition for the sustainability of GUUGLE's impact. The chapter suggested that the whirlpool scheme could be generalised and used as model for theorising, planning, visualising and evaluating any kind of OD and similar social interventions. To reconcile Pawson & Tilley's Realistic Evaluation with the whirlpool-based GUUGLE programme theory, their concept of context was decomposed into programme context (assumed fixed) and delta context (altered through programme interventions), and their CMO configuration was replaced by a sequence of DMO configurations. Finally, the chapter provided a justification for limiting the realist evaluation of GUUGLE to a single programme theory rather than considering rival theories.

Chapter 6 First Evaluation Cycle: Quantitative Analysis

The purpose of this and the next two chapters is to empirically test and refine the GUUGLE programme theory by following the four phases of the realist evaluation cycle in Figure 3.4 and applying different analytical methods in each cycle. Phase 1 of the first cycle (theory building) was completed in the last chapter, so this chapter continues with phases 2-4 (hypothesis generation, hypothesis testing and programme specification) and phase 1 of the second cycle.

6.1 Deriving the hypotheses

Figure 5.4 specifies the following four outcomes:

- (a) Professors regard CETL as an important idea.
- (b) Professors engage in CETL.
- (c) Professors regard CETL as an important practice.
- (d) CETL forms part of the organisational culture.

The outcomes form a sequence, which is defined by the order in which they are connected by the six solid arrows in Figure 5.4: (a) precedes (b) and (b) precedes (c) but (c) also precedes (b) because together they constitute feedback loop 1. And all of them precede the final outcome (d), which, in turn, is part of feedback loop 2 and hence precedes (b). Given this sequence, the achievement of normative outcomes can serve as GUUGLE implementation progress indicator. Table 6.1 shows the sequence of the six DMOCs corresponding to the six mechanisms which connect the four normative

outcomes (a)-(d). The programme context (C) has been characterised in section 4.1 and is omitted here.

Table 6.1. Normative GUUGLE programme theory translated into DMOCs.

#	Delta context	+	Mechanism	=	Outcome
1	Communication of reasons for CETL		Reasons trigger learning processes in professors.	=	(a) Professors regard CETL as an important idea.
2	Outcome (a) + resources for CETL	+	Professors decide to use the resources for CETL.	=	(b) Professors engage in CETL.
3	Outcome (b)	+	Professors experience CETL as an opportunity for collaborative reflection and learning.	=	(c) Professors regard CETL as an important practice (and tell their colleagues: positive word of mouth).
4	Outcome (c) + resources for CETL	+	Professors continue using the resources.	=	(b) (More and more) professors engage in CETL.
5	Outcome (b) + much practice of CETL + resources	+	Professors internalise CETL as part of their professional role.	=	(d) CETL forms part of the organisational culture.
6	Outcome (d) + resources	+	Respecting the cultural norm of CETL is associated with positive social sanctions.	=	(b) Professors engage in CETL.

The Realistic Evaluation of GUUGLE aims to answer the question why the programme works for whom and in what circumstances. For whom it works is an aspect of programme outcomes. A person for whom GUUGLE works shall be defined behaviourally as somebody who participated in at least one PLC and who intends to get involved in further CETL within or outside of GUUGLE.

The GUUGLE programme theory does not suggest that one group of professors is more likely to benefit from GUUGLE than another, so determining *for whom* the programme works will require exploratory rather than confirmatory data analysis. *Why* GUUGLE

works refers to the causal mechanisms introduced by the programme and constitutes the object of interest of retrodiction. There is, of course, no guarantee that the normative programme theory is correct and the mechanisms in Table 6.1 are accurate or complete. Ad hoc hypotheses may need to be developed and tested as part of the evaluation.

Not all DMOCs in Table 6.1 are testable by means of the same data sets and methods. In most cases, mixed-method triangulation will be applied. Table 6.2 gives an overview of the evaluation cycles in which the different DMOCs will be addressed.

Table 6.2. Coverage of DMOCs in different evaluation cycles.

DMOC #	1 st cycle	2 nd cycle	3 rd cycle
1	x	x	x
2	x		x
3	x	x	x
4	x	x	x
5		x	x
6			x

6.2 Validating the normative outcomes (a)-(c)

The quantitative data used in this chapter originate from two sources: the participant lists of all PLCs, GUUGLE Forum conferences and teacher training seminars between May 2009 and February 2013 as well as an anonymous survey conducted in October 2012 among all teaching staff. The survey data set used here comprises only the responses from the 27 professors (39% of all professors at the *Hochschule Bremerhaven*) who completed the online questionnaire. Responses from non-professorial teaching staff were eliminated from the data set. In the following, the first three normative outcomes of the GUUGLE programme theory will be addressed. For outcome (d), no quantitative data is available yet.

6.2.1 Validating outcome (b)

The most important normative outcome of GUUGLE, the starting point of both of its feedback loops in Figure 5.4 and, ultimately, its *raison d'être* is (b): 'professors engage in CETL'. Establishing this outcome empirically is the first step of testing the DMOCs 2 and 4, both of which involve this outcome. In the present evaluation cycle, the quantitative dimension of CETL will be the focus. In the second and third evaluation cycle, the nature of the CETL – dialogue, discussion, problem-solving etc. – will be analysed.

A look at PLC participant numbers and repeater rates gives an impression of the extent to which the GUUGLE programme has managed to mobilise professors for CETL and to thus achieve the normative outcome (b). As Table 6.3 demonstrates, every second professor at the *Hochschule Bremerhaven* has taken part in at least one PLC and almost every second professorial participant has joined two or more PLCs.

Table 6.3. PLC participation by professors 2009-2013.

No. of PLCs joined by individual professors	Number of professors	Share of PLC participants	Share of all 69 professors
≥ 1 PLC	36	100%	52%
≥ 2 PLCs	17	47%	25%
≥ 3 PLCs	13	36%	19%
> 3 PLCs	8	22%	12%

Considering further that PLCs extend over one or two semesters, comprise between four and eight sessions of three hours each, and in some cases require the participants to do research, write input papers, try out new teaching methods and collect student feedback, it is fair to say that GUUGLE has made a notable difference at the *Hochschule*

Bremerhaven where, apart from formal committee meetings, no forums for CETL existed prior to GUUGLE.

Moreover, 14 professors who never joined a PLC attended between one and three GUUGLE Forum conferences, in some cases as presenters or podium discussants. All professors who ever attended a GUUGLE-run teacher training, guest lecture or seminar were also active in PLCs and/or turned up at a GUUGLE Forum. The number of professorial GUUGLE participants thus amounts to 50 out of 69.

The 2012 survey responses project a similar image. The respondents were asked to state the number of GUUGLE events they had joined in the past ('Past GUUGLE participation' or PastGP in Table 6.4) and their intention to join future GUUGLE events ('Future GUUGLE participation' or FutuGP in Table 6.4). Among the GUUGLE activities, the PLCs are the most dialogic ones, yet also the annual GUUGLE Forum conferences provide good opportunities for CETL. The two relative frequency distributions show a very high repeater rate among the respondents and also a clear intention by over three quarters to participate in (further) GUUGLE events in the future.

Table 6.4. Professors' past and intended future participation in GUUGLE.

Past GUUGLE participation, PastGP (n=27)		Future GUUGLE participation, FutuGP (n=27)	
Never	7%	No	7%
Once	19%	Rather not	7%
Twice	7%	Possibly	7%
Three times	15%	Rather yes	26%
More than three times	52%	Yes	52%
	100%		100%

6.2.2 *Validating outcome (a)*

The next step is to validate the normative outcome (a): ‘Professors regard CETL as possible and as an important idea’. The possibility of CETL refers to the resources it requires, especially time, space and organisation. The belief in the possibility of CETL is implicit in the professors’ decision to join PLCs (outcome (b)) and needs no further evidence. As for the importance of the idea of CETL, the respondents of the 2012 survey were asked to what extent they agreed with the statement, ‘The aim of GUUGLE to stimulate collegial exchange about teaching and learning is important to me’.

The relative frequency distribution of the respondents’ answers can be seen in Table 6.5 (‘Importance of CETL as idea’ or ICETL). Over 70% of the respondents felt that the idea of CETL was indeed important to them. Given that three and a half years earlier, just before the start of GUUGLE, I had identified a professional identity discourse which asserted that CETL was unnecessary (see section 4.1.3), a significant change in attitude seems to have taken place in the meantime.

Table 6.5. Professors’ views regarding the importance of the idea of CETL.

Importance of CETL as idea, ICETL (n = 27)	
Disagree	7%
Tend to disagree	7%
Neutral	15%
Tend to agree	41%
Agree	30%
	100%

6.2.3 *Validating outcome (c)*

With the available quantitative data, normative outcome (c) – ‘Professors regard CETL as an important practice’ – cannot be validated in a straightforward manner. The GUUGLE participant data contain no such information and the 2012 survey did not include a question about the respondents’ actual participation in CETL or PLCs. In the absence of first-best data, a second-best item needs to be used as substitute.

In the survey, the respondents were asked to rate the importance which 15 different actual and potential GUUGLE activities and service offers should be given within the programme in the future. One of the proposed activities was PLCs. Using the item ‘Importance of PLCs’ or IPLC as a substitute for the (unobserved) importance attributed to the practice of CETL may be justified on the grounds that PLCs are a format of CETL and that those professors who attach great importance to PLCs would probably do the same with CETL in general.

The relative frequency distribution of the survey item IPLC is shown in Table 6.6. Almost two-thirds of the professorial survey respondents rated the PLCs as important or rather important, thus validating the normative outcome (c) within the limitations inherent in the chosen substitute item.

Table 6.6. Professors’ views regarding the importance of PLCs.

Importance of PLCs, IPLC (n = 27)	
Unimportant	7%
Rather unimportant	11%
Neither nor	19%
Rather important	30%
Important	33%
	100%

6.3 Testing the hypothesised causal model

After establishing the outcomes (a)-(c) of the GUUGLE programme theory, the hypothesised regularities between them (DMOCs 2-4) can be tested. The tests applied in this section quantify the strength of associations between observed variables but do not permit drawing conclusions about the directedness of associations. To this end, the quantitative data and analyses in this chapter are complemented with qualitative data and analyses in the next two chapters.

6.3.1 Path model

Figure 6.1 depicts a linearised version of the micro level of Figure 5.4. Linearisation means that feedback loop 1 has been ‘unlooped’ and stretched out. The three micro-level outcomes of Figure 5.4 have increased to four, since ‘CETL takes place’ is now shown twice, before and after ‘CETL is an important practice’. This way, Figure 6.1 matches the equally linearised structure of the DMOC sequence in Table 6.1 and simplifies the testing procedure substantially.⁸

The arrows connecting the outcome boxes indicate the direction of hypothesised causal mechanisms. Together, outcomes and mechanisms form a causal chain or a pathway leading from the exogenously determined outcome (a) – note that in the full model in Figure 5.4 this outcome is endogenous – via the intermediate outcomes (b) and (c) to

⁸ Path models without feedback loops are called recursive. Nonrecursive models can also be constructed but ‘The problem is that not only would we be going around in circles in the diagram, but we would also be going around in circles trying to interpret the results. The best advice we can offer about postulating a nonrecursive model is *don’t try this at home.*’ (Norman & Streiner 2003, 161, original emphasis)

the final outcome (b). The purpose of this section is to test whether the available data support the hypothesis that the professors' view of CETL as an important idea could have caused, via the depicted path, their repeated participation in such activities.

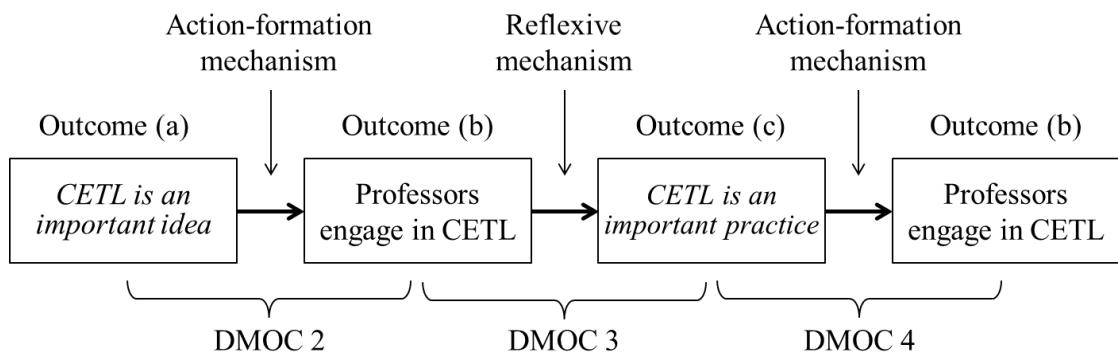


Figure 6.1. Linearised causal pathway described by the DMOCs 2-4.

To turn Figure 6.1 into a statistically testable path model, the outcomes must be replaced by observed variables. As before, the survey item ICETL (see Table 6.5) represents the professors' views about whether CETL is an important idea and the survey item IPLC (see Table 6.6) serves again as substitute for the importance the professors attribute to the practice of CETL. Participation in CETL appears twice in Figure 6.1 and is substituted by the two survey items PastGP and FutuGP (see Table 6.4).

To complete the path model, a disturbance term ε_i is added to each of the three endogenous variables. It captures that share of a variable's variation which is not explained by the variable preceding it along the causal path. Path models are statistically estimated and tested using path analysis, a method closely related to multiple regression analysis. The path analytical computations have been performed using SPSS 20 in

combination with AMOS 20. The data have been checked for compatibility with the requirements of path analysis.⁹

Figure 6.2 presents some of the results of estimating the model as described. Each arrow is labelled with a standardised path coefficient. These coefficients indicate the relative strength and sign of the direct influence of a variable assumed to be a cause on another variable assumed to be an effect. In the case of a single-path model like the present one, the path coefficients are identical with Pearson correlation coefficients. The numbers written on top of the variable boxes are the squared multiple correlations R^2 which measure the portion of the variance of a variable, which is explained by the model. An R^2 close to 1 means that a variable's variance can largely be explained by the variable preceding it on the path, whereas an R^2 close to 0 states that the model contributes only little to the explanation of the corresponding variance.

Thanks to its simplicity, Figure 6.2 is easily interpreted. Starting from the left with the only exogenous variable of the model and assuming the directions of all arrows to be correctly chosen, the path coefficients state that the professors' view that CETL is an important idea had a positive influence on the frequency of their past participation in

⁹ The standard assumption is that only multivariate normal data should be used in structural equation models (SEM), of which path models are a special case. However, all four variables fail the Shapiro-Wilk test for univariate normality even after a logarithmic transformation of the reflected data. Yet non-normal data are per se no obstacle to SEM, as the effect of non-normality depends on its extent. Various studies suggest that 'problems may occur when univariate skewness and univariate kurtosis approach values of 2 and 7, respectively [...] and Mardia's normalised multivariate kurtosis [...] greater than 3 could produce inaccurate results' (Finney & DiStefano (2006), 272-273; see also Byrne (2009), 102-105). For the untransformed variables, univariate skewness reported by AMOS ranges between -.70 and -.13 and univariate kurtosis between -.95 and .52. Mardia's normalised multivariate kurtosis value amounts to 1.05. This means that the extent of non-normality in the original data is fully acceptable for the use in path analysis.

GUUGLE events. The more GUUGLE events they attended, the more important they found PLCs, which, in turn, strengthened their intention to participate in future GUUGLE events. All three path coefficients are statistically highly significant ($p=.000$).

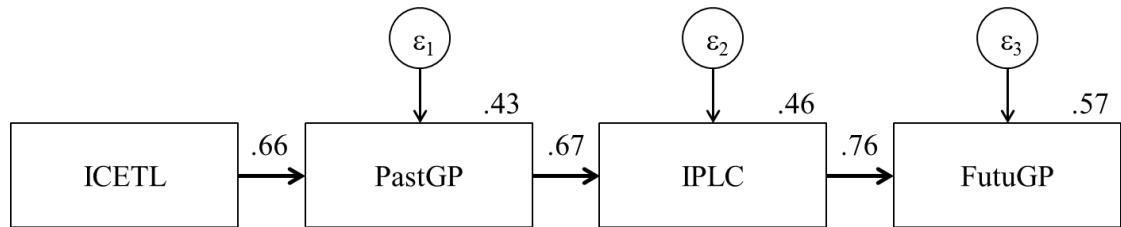


Figure 6.2. Estimated path model with standardised coefficients.

So far, the model seems to confirm the microsocial level of the GUUGLE programme theory shown in Figure 5.4. However, the fairly low R^2 values give reason to worry. The exogenous variable ICETL only accounts for 43% of the variance of PastGP which it is expected to determine, leaving 57% of the variance unexplained, and the R^2 of IPLC is almost as low. In fact, the path model fails eight standard tests of goodness of fit with the data (see model 1, Table A1, Appendix A2). Thus it must be rejected and replaced by a better one.

6.3.2 Extending the path model

The goodness of fit of path models increases with every additional connection. The aim is to explain a satisfactorily large portion of the variables' variation with the smallest possible number of connections. As Figure 6.2 shows, the path model so far comprises the smallest possible number of connections any model with $k=4$ variables can have, while the maximum is $k(k-1)/2=6$ connections. Hence the poor model fit is not very surprising.

The model in Figure 6.2 relies on the supposition that the only direct effect of ICETL is on PastGP. The model further assumes that PastGP is the only factor influencing IPLC. Both assumptions seem overly (and unnecessarily) restrictive and could be responsible for the low R^2 associated with IPLC. It is quite conceivable that professors who believe that CETL is an important idea will support the concept of PLCs independently of their own experience with them. In terms of the path model, this hypothesis translates into an additional causal link from ICETL to IPLC.

Estimating the revised model produces the standardised path coefficients and R^2 values shown in Figure 6.3. The first thing to note is that the coefficient associated with the new path is highly significant ($r=.63$; $p=.000$) whilst the coefficient of the connection between PastGP and IPLC has dropped from .67 to .26, losing its statistical significance ($p=.076$). The model now suggests that it was primarily the professors' belief in the idea of CETL and not their actual past experience with CETL as GUUGLE participants, that induced them to rate PLCs as important in the survey.

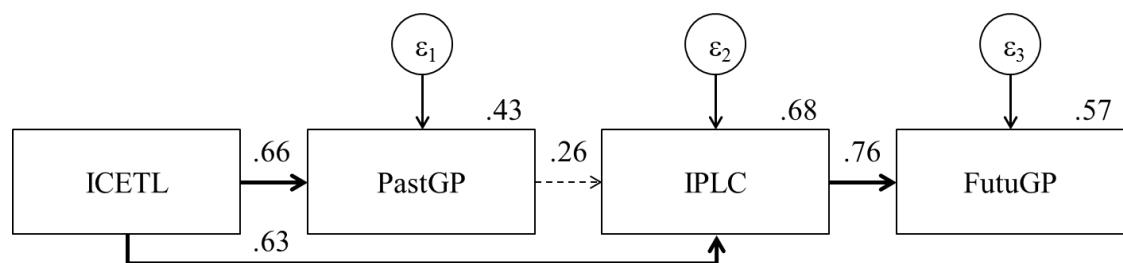


Figure 6.3. Estimated extended path model with standardised coefficients.

A second observation is that the R^2 of IPLC has increased from .46 to .68, that is, the portion of the variable's variance explained by the model has gone up by 22 percentage points due to the new connection. The impact on model fit is also positive: the revised model passes all goodness-of-fit tests (see model 2, Table A1, Appendix A2).

6.3.3 *Beneficiaries of GUUGLE and the path model*

Its fit to the data may be sufficient and its interpretation plausible, but the revised path model still suffers from a low R^2 of PastGP. This time no help can be expected from an additional connection if the model is to remain recursive and its basic structure intact. So the question is, if ICETL only accounts for 43% of the variance of PastGP, what else may have influenced this variable? Here is a link to the guiding question of Realistic Evaluation: why does a programme work for whom and in what circumstances? So far, this study was concerned with why GUUGLE might work and in what circumstances, but not with its beneficiaries. For whom GUUGLE might work does not follow from its programme theory and has therefore also not been taken into account in the path model. This omission could explain the relatively low R^2 of PastGP.

For whom GUUGLE works is not about specific professors but about the shared characteristics or common factors motivating them to participate. Due to the absence of a theory predicting the professors' participation choices, pragmatic considerations (such as the availability of quantitative data) and ad hoc hypotheses need to be used to narrow down the search for causes. The 2012 survey data set includes nine items about respondent attitudes and experiences, each of which is exogenous to the path model in Figure 6.3, unrelated to GUUGLE and might have given rise to the observed PastGP pattern: self-concept (teacher/expert), professional responsibility (for learning/teaching), perception of students (as kids/adults), energy impact of teaching (takes/gives energy), perception of feedback from students (not/very important), typical size of classes taught (large/small), time available to cover course contents (time pressure/Enough time), results of own teaching evaluations by students (not so good/very good) and discussion of teaching evaluation results with colleagues (never/always). However, correlation

analysis reveals not a single even remotely significant relationship (see Table A2, Appendix A2). Perhaps professors from the engineering faculty were less attracted by the ‘soft’ dialogic approach of GUUGLE events than other professors. But also this hypothesis is not supported by the data, as the independent samples t-test, with engineering faculty membership as independent binary variable and PastGP as dependent variable, does not indicate a significant effect ($p=.870$).

After exhausting all possibilities of finding in the 2012 survey data a theoretically justified explanatory variable for PastGP that is not already included in the path model, the search continues in the 2009-2013 GUUGLE participation data. A repetition of the independent samples t-test with engineering faculty membership as independent binary variable and the number of GUUGLE events actually attended as dependent variable again does not yield a significant result ($p=.210$). But this time statistical significance is not decisive, as the data set comprises the entire population of professors of the *Hochschule Bremerhaven*. The hypothesised pattern is clearly discernible in the data: professors from the engineering faculty were underrepresented in GUUGLE events. But is the reason necessarily their alleged scepticism towards dialogical methods?

An alternative explanation offers itself: the two initiators of GUUGLE, Wolfgang and I, are both business professors with limited contacts to the engineering faculty. Given the ‘bottom-up’ evolution of the GUUGLE programme, our own social networks may have influenced who joined the first PLCs. Through word of mouth, the early participants, whose social networks overlapped with ours, attracted further participants. This way, professors from the engineering faculty would almost certainly have been underrepresented among the GUUGLE participants.

To test this new hypothesis, the GUUGLE participant data can be combined with data from my social network analysis (see section 4.1.2) using the participants' names as unique identifiers. The social network analysis took place just before the beginning of GUUGLE and provides a snapshot of the teaching-related communication patterns among 40 professors including Wolfgang and me. The number of ties along the shortest path between any of the other 38 participants of the social network analysis and either Wolfgang or me (whichever path is shorter) is called degrees of separation. It can serve as a measure of a professor's network proximity to the GUUGLE team (S. Cohen, Kimelfeld & Koutrika 2012). Correlating the number of PLCs joined by all 40 professors with their respective degrees of separation from the GUUGLE team yields a highly significant Spearman rank coefficient of $\rho=-.561$ ($p=.000$). Eliminating Wolfgang and me from the sample makes no fundamental difference: $\rho=-.488$ ($p=.002$). In plain language this means that, on average, the lower the degrees of separation between a professor and the GUUGLE team, the greater is the number of PLCs she or he joined in the past. Figure 6.4 visualises this correlation for all 40 professors who are represented by the horizontal axis and ordered by the number of PLCs joined. Since Wolfgang and I are the GUUGLE team, our degrees of separation are zero.

This is evidence that the beneficiaries of GUUGLE have been especially those professors who happened to be part of Wolfgang's and my social networks at the *Hochschule Bremerhaven*. It means that its own history has been an influential contextual condition of the GUUGLE programme. The effect of network proximity on participation choices may be able to explain the fairly low R^2 of PastGP in Figure 6.3. This, in turn, suggests that the characterisation of educational institutions as loosely coupled systems (see section 2.3.1) does not necessarily extend to all of their elements.

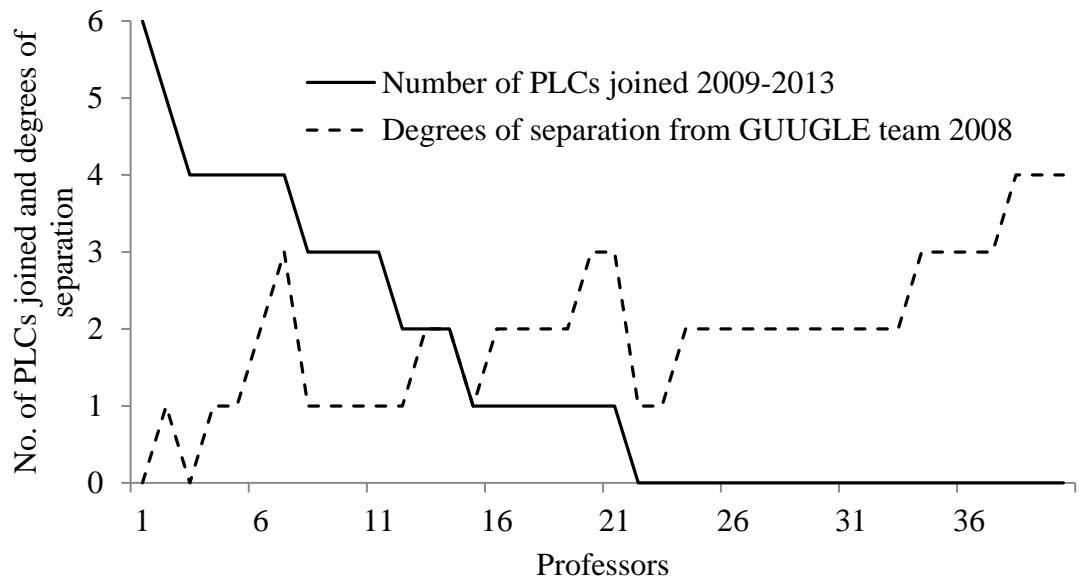


Figure 6.4. PLC participation frequency and proximity to the GUUGLE team

As the 2012 online survey was anonymous and thus no identifier is available to connect the degrees of separation data with the survey data, network proximity cannot be included as an explanatory variable in the path model. Its influence on the professors' participation in past GUUGLE events can only be captured indirectly by assuming that PastGP's low R^2 in the path model is largely due to the strong exogenous impact of network proximity. Technically speaking, this means that the disturbance term ε_1 in Figure 6.3 comprises an important component representing the influence of the PLC participants' network proximity to the GUUGLE team.

To verify whether the close association between network proximity and PLC participation choices was only an initial phenomenon which vanished over time as GUUGLE became better known and publicised, rank-biserial correlations between participation data and the participants' degrees of separation from the GUUGLE team could be calculated for each PLC separately and then checked for a time trend.

However, since the PLCs had only between six and 14 participants and many of them were not included in the social network analysis, the sample sizes would be too small for any meaningful and robust results.

6.4 Specifying the GUUGLE programme

After completing phase 3 of the first realist evaluation cycle, the re-specification of the GUUGLE programme, i.e. the re-statement which parts of the programme work for whom and in what circumstances, would be next. But since all of this has already been said in sections 6.2 and 6.3, a brief summary should suffice.

The focus of this evaluation cycle has been on the DMOCs 1-4. All outcomes specified by these configurations have been confirmed on the basis of the available quantitative data: (a) professors regard CETL as possible and as an important idea; (b) professors engage in CETL; and (c) professors regard CETL as an important practice. Not only reasons in favour of CETL seem to have enabled the GUUGLE programme to achieve its outcomes (a)-(c), but also the social network structure: professors' network proximity to the GUUGLE team may have strongly influenced their PLC participation decisions. This aspect is probably due to the circumstances under which GUUGLE came into existence. Its informal 'bottom-up' origins offer a good explanation of how the personal relationships of the GUUGLE team could have been so influential in the recruitment of GUUGLE participants.

6.5 Refining the GUUGLE programme theory

To close the realist evaluation cycle, this section proposes a refinement of the GUUGLE programme theory, taking the empirical findings of this cycle into account. Actually it is

not a refinement but the transition from a normative to an empirically grounded, explanatory programme theory which results from integrating the estimated and tested path model of this chapter into the normative programme theory. Diagrammatically speaking, this requires first restoring the original circularity of the model. In section 6.3.1, feedback loop 1 of Figure 5.4 was ‘unlooped’ to obtain a recursive path model. To reverse this transformation, PastGP and FutuGP which both represent CETL (see Figures 6.1 and 6.2) and form the beginning and the end of the feedback loop are merged. Figure 6.5 shows step by step how the path diagram is ‘re-looped’ and rearranged.

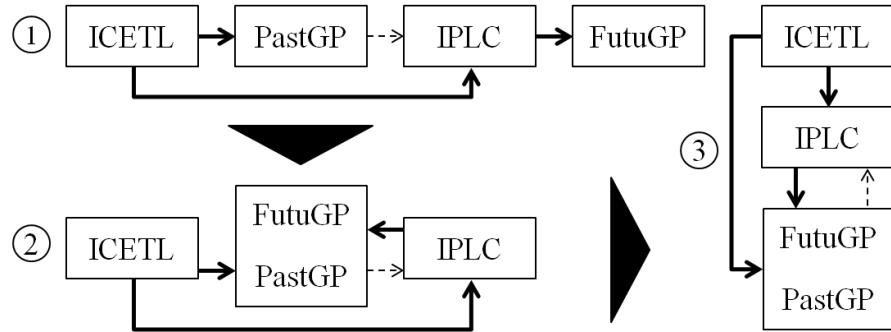


Figure 6.5. Re-looping the path model.

After replacing the variable names by the respective outcomes they represent, adding the path coefficients and plugging everything into the slightly adjusted Figure 5.4, the result is Figure 6.6.

The whirlpool structure of the model is preserved. As in Figure 5.4, the GUUGLE programme provides resources and reasons for CETL and CETL itself is the starting point of two feedback loops. But now the entire micro level of the diagram is represented by the path model in which feedback loop 1 is embedded. The professors’

engagement in CETL is exogenously fuelled by GUUGLE resources, by the reasons for CETL provided by GUUGLE, by feedback loop 2 and also by professors' network proximity to the GUUGLE team.

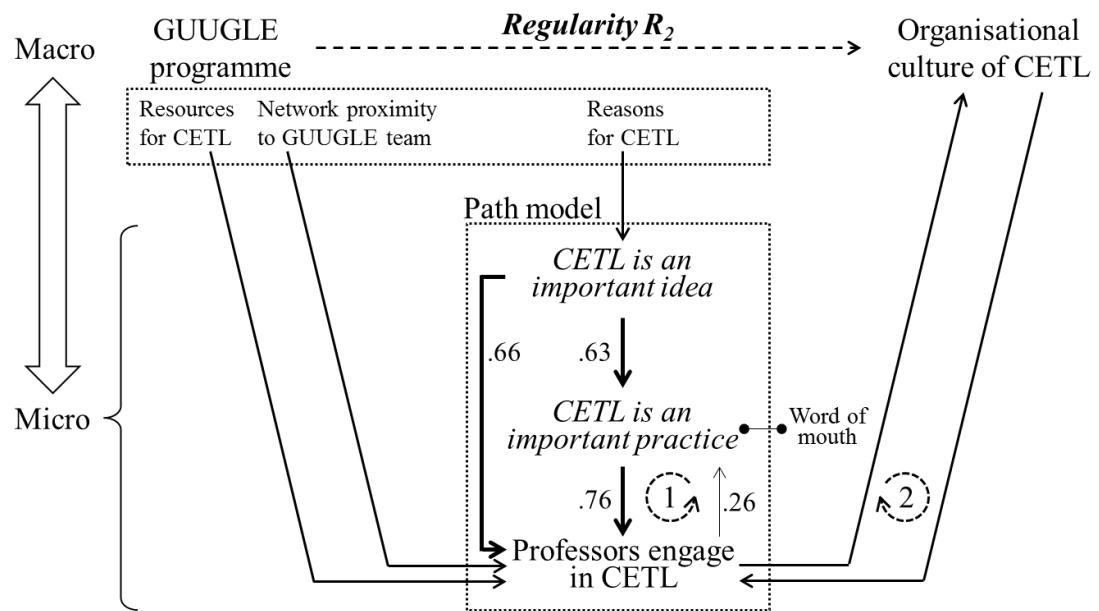


Figure 6.6. Whirlpool diagram of GUUGLE after the first evaluation cycle.

Chapter 7 Second Evaluation Cycle: Content Analysis

The search for generative social mechanisms may involve, but does not end with, the establishment of systematic covariation between variables or events. As Hedström & Swedberg (1996) point out, the only ‘mechanism’ which causal modelling is able to identify is a regression coefficient statistically quantifying the association between two measures. ‘But if there is no underlying model with a reasonable substantive interpretation, little has been gained by such curve fitting’ (Coleman 1964, 518). The causal modelling performed in the first evaluation cycle goes beyond the criticised ‘black-box’ explanations in that its analyses and findings are theory-guided and embedded in a model. Nevertheless, the explanatory power of path models is limited. Therefore, the second and third evaluation cycle will complement and enrich the quantitative results.

7.1 Deriving the hypotheses

After refining the GUUGLE programme theory to match the empirical findings of the first evaluation cycle, also the corresponding DMOCs need to be adjusted. Fortunately, the normative outcomes (a)-(d) stay the same as in section 6.1. Only the order of the first two DMOCs must be reversed, some mechanism descriptions slightly altered and the GUUGLE history added as a delta context. Table 7.1 contains the updated set of DMOCs which represent the hypotheses to be tested subsequently.

The distinction between ‘CETL as an important idea’ and ‘CETL as an important practice’ merits clarification. In section 5.2, the former was said to refer to an attitude based on abstract reasons for CETL and the latter to the specific individual condition of

professors actually engaging in CETL. But after adjusting the GUUGLE programme theory, personal experience with CETL is no longer a precondition for it to be considered an important practice. DMOC 2 in Table 7.1 hypothesises that professors can also come to regard the practice of CETL as important on the basis of expectations and not only of personal experience.

Table 7.1. Refined GUUGLE programme theory represented by DMOCs.

#	Delta context	+	Mechanism	=	Outcome
1	Communication of reasons for CETL		Reasons trigger learning processes in professors.	=	(a) Professors regard CETL as an important idea.
2	Outcome (a)	+	Professors form expectations about how CETL may benefit them in practice.	=	(c) Professors regard CETL as a (potentially) important practice.
3	Outcome (a) or (c) + 'bottom-up' evolution of GUUGLE + resources for CETL	+	Professors, especially those close to the GUUGLE team decide to use the resources offered.	=	(b) Professors engage in CETL.
4	Outcome (b) + resources for CETL	+	Professors experience CETL as an opportunity for collaborative reflection and learning.	=	(c) Professors regard CETL as an important practice (and tell their colleagues: positive word of mouth).
5	Outcomes (c) + much practice of CETL + resources	+	Professors internalise CETL as part of their professional role.	=	(d) CETL forms part of the organisational culture.
6	Outcome (d) + resources	+	Respecting the cultural norm of CETL is associated with positive social sanctions.	=	(b) Professors engage in CETL.

The argument here is that professors who believe that CETL is an important idea also ask themselves how they might personally benefit from engaging in it. The answer to this question differs from person to person, since it is influenced by the GUUGLE communication as well as by individual values, aspirations, circumstances, biographical

characteristics etc. The totality of these influences may or may not lead a professor to conclude that for her or him, CETL would probably be an important practice. Once this view is adopted, actual CETL is mainly a matter of opportunity, i.e. of the right people (apparently including the right mediators, such as the GUUGLE team) and the necessary resources (time, space, organisation). The next section will take a closer look at the professors' expectations of CETL, which served as their reasons for participating in PLCs.

7.2 DMOC 2: professors' expectations of CETL

In the 2011 interviews, 18 professors were asked an open-ended question about their main reasons or motives for (not) joining PLCs. The answers can be grouped as shown on the vertical axis of Figure 7.1. The column of figures to the right of the diagram contains the number of times each reason was stated. If a respondent gave two reasons, each one is counted only half. The horizontal axis shows the number of PLCs the interviewees joined. The bubble positions in the matrix represent combinations of reasons and numbers of PLC participations. Bigger bubbles indicate that a particular combination occurred twice. Bubbles located between two horizontal lines mean that respondents stated two reasons: the one above and the one below the respective bubble position.

According to this diagram, the issues addressed in PLCs were the main reason for seven of the 14 PLC-experienced interviewees to join. As hypothesised by the mechanism of DMOC 2, the respondents seemed to have formed more or less specific expectations about how their participation in a certain PLC could help them identify solutions, win allies or at least find open ears. Collegial exchange (CETL) was named by four

professors as a reason for joining a PLC. Two of them expressed their expectation that PLCs would allow them to get to know colleagues (better) and one professor said that for him, the topic of a PLC was almost irrelevant, as long as it stimulated conversations with colleagues. Curiosity about the format and interest in the process of a PLC was mentioned three times as a reason for joining. One respondent had only recently been hired and was obliged to attend the PLC for new teaching staff.

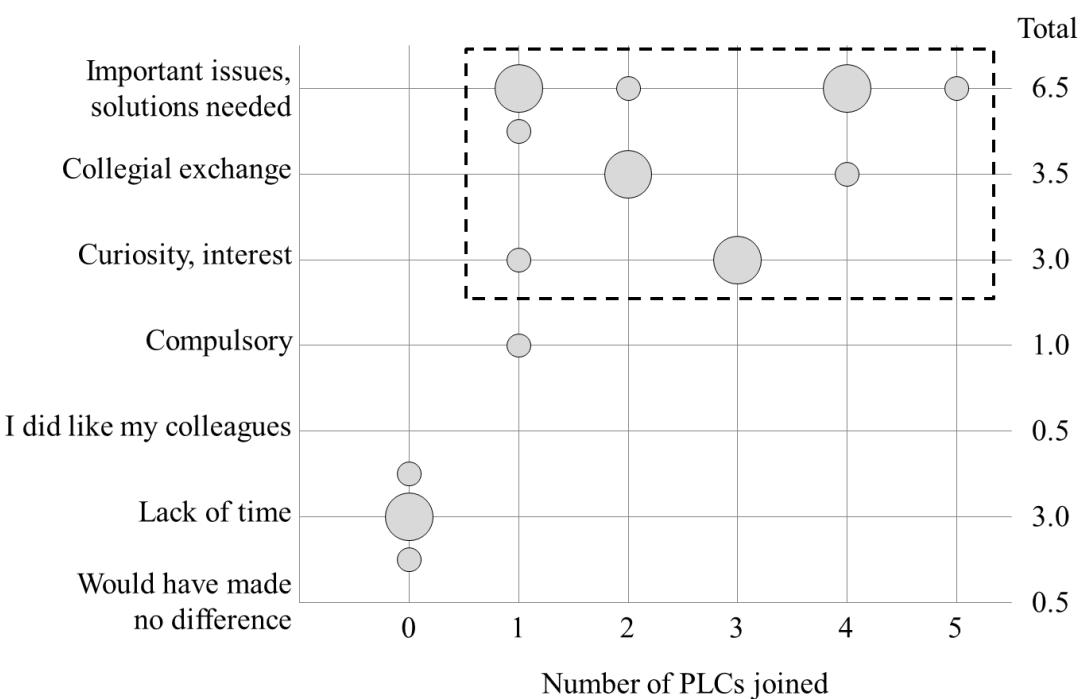


Figure 7.1. Reasons stated for PLC (non-) participation (basis: 18 interviews).

Lack of time was the reason given by all four respondents who had never participated in a PLC. But since the time invested in a PLC is compensated through teaching load reductions, it can be assumed that their stated lack of time was a polite way of expressing their lack of interest. In addition, one of these four professors spoke about his frustration with his job and felt that no PLC would make any difference. Another said that he might have participated if any of his direct colleagues had done so.

These insights shed some light on the reasons underlying PLC participation decisions. Personal relationships with members of the GUUGLE team do not seem to be one of them. But network proximity, which was identified in the first evaluation cycle as an important influence on professors' decision to join a PLC, should not be interpreted as a reason or even the only reason. Rather, professors with direct personal contacts to the GUUGLE team may have paid more attention to invitations to join PLCs than other professors who received the same invitations but did not know the senders well. In this sense, network proximity acted as a facilitator in the professors' decision-making process, not as a reason.

In the following section, the respondents' reasons and motives for (not) participating will be compared against the reasons for CETL which were promoted in GUUGLE's own information and promotional material.

7.3 DMOC 1: communicated reasons for CETL

The hypothesis expressed by DMOC 1 is that the GUUGLE programme, through its communication activities, provides reasons for CETL which induce or encourage professors to regard CETL as important or confirm them in this. If the hypothesis is correct, the reasons for CETL should be reflected by the professors' expectations of how CETL may benefit them in practice, as stated by DMOC 2. To test this hypothesis empirically, a modified content analysis is conducted with the aim to categorise GUUGLE's communication in a way that allows a comparison with the professors' stated reasons depicted in Figure 7.1. The media taken into account comprise all information and promotion materials targeted at professors and produced by the GUUGLE team between 2009 and March 2013 (newsletters, circular emails, reports,

posters, flyers, roll-up displays, Powerpoint documents, website content). The media analysed do not include emails to individuals and materials created by PLCs.

The content analysis proceeds in four steps. In a first step, a small number of *a priori* codes are defined to categorise the available material by topic ('what is it about?') and purpose ('what is it for?'). The topic codes are chosen so as to be compatible with the professors' stated reasons for joining PLCs. The purpose codes capture GUUGE's main communication purposes, namely information, activation of participants, accountability and conveying a positive image.

The actual coding takes place in the second step. The unit of analysis, i.e. the basic unit of text to be classified during content analysis, is either the whole document or, in the case of GUUGLE newsletters and the website, the individual article or webpage. Each of the 324 units of analysis is coded with respect to topic and purpose.

The third step involves a weighting procedure. Every unit of analysis is assigned two percentages, one reflecting the estimated share of readers or recipients of that unit (69 professors=100%) and the other reflecting the estimated attention given to it by a recipient or reader (highest attention=100%). Multiplication of the two percentages yields an impact factor for each unit of analysis. For example, a particular Powerpoint presentation may have enjoyed the attendees' full attention (100%) but if the meeting during which it was given was attended by only seven professors (10%), its overall weight amounts to only 10%. The GUUGLE newsletter, on the other hand, is distributed to all professors (100%). The attention it receives is only fair (e.g. 50%), as feedback suggests. The impact factor thus equals 50%. However, the unit of analysis is not the newsletter itself but the articles it contains. Each article covers a fraction of a

newsletter. This fraction is multiplied by the newsletter's overall impact factor to obtain the individual article's impact factor.¹⁰

In the fourth and final step of the content analysis, a matrix is constructed with the topic codes on the vertical and the purpose codes on the horizontal axis. For each topic-purpose combination all occurrences are identified and their respective impact factors added up. The resulting topic-purpose impact factors are normalised so that their total equals 100%. Figure 7.2 depicts the matrix. The percentages on its border are the vertical and horizontal sums.

The content analysis of the GUUGLE media reveals that the topic of teaching and learning at the *Hochschule Bremerhaven* had by far the highest normalised impact factor (54%). More than half of it was associated with the purpose of activating teaching staff to participate in GUUGLE events. In this respect, Figure 7.2 corresponds very closely to Figure 7.1, according to which specific issues in need of solutions (for academic teachers, these are primarily issues of teaching and learning) were the main reason for several professors to join a PLC. The second most important topic of the GUUGLE communication was GUUGLE itself (Figure 7.2). This information supply was met by the demand of three interviewed professors who stated that they had been curious about GUUGLE and the format and process of the PLCs (Figure 7.1).

¹⁰ Naturally, the impact factors can only be very rough indicators of the influence which the underlying information and promotion materials may have exerted on the professors. While the number of recipients of a circular email or a printed newsletter which is put in every professor's pigeon hole can be estimated with some accuracy, estimating the number of professorial website visitors, even if it is based on personal feedback and web analytics, involves great uncertainty. The same applies to the estimation of professors' attention attracted by posters on the wall, whereas the attention given to a presentation by the attendees can be estimated fairly well.

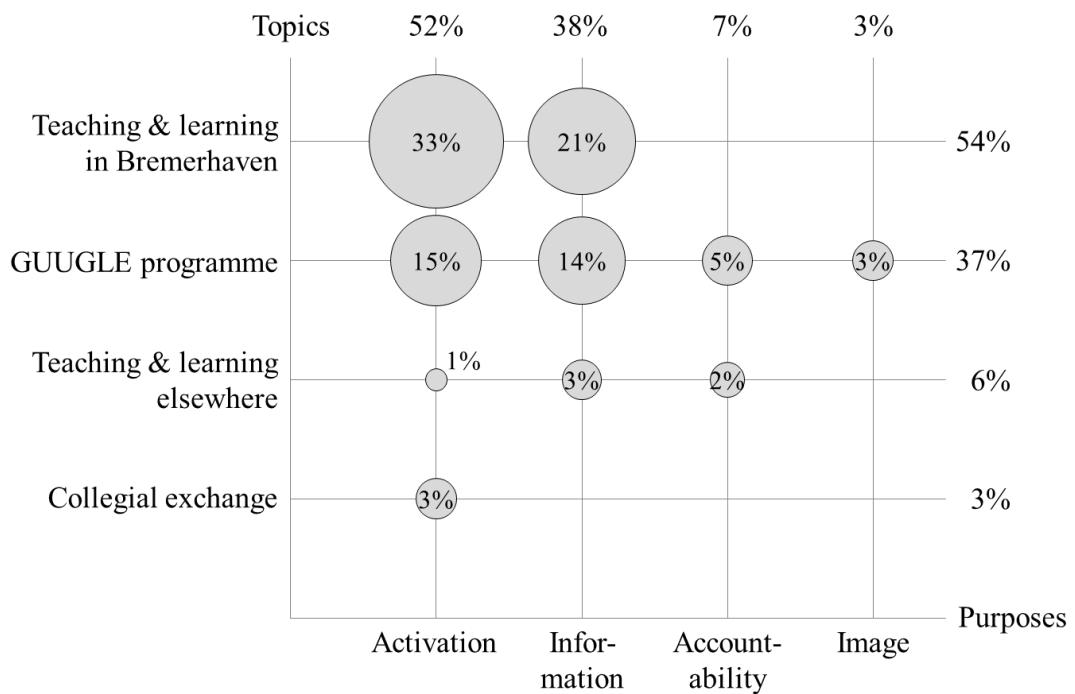


Figure 7.2. Normalised topic-purpose impact factors.

Four interviewed professors joined PLCs because they were keen on CETL (Figure 7.1). The specific issues addressed by the PLCs were less relevant to them. The importance of CETL as such, however, was deliberately de-emphasised by the GUUGLE communication (Figure 7.2). It would have been odd, and might have been received badly, if GUUGLE's key message to the professors had been that they needed to talk to each other more often. Instead, issues of professional relevance were put forward almost as a Trojan horse to get the professors to engage in CETL.

Figure 7.2 highlights the communicative emphasis the GUUGLE team put on activating and mobilising teaching staff. Activating communication consisted mostly of invitations to attend events, join PLCs, submit ideas and give feedback, but also of suggestions to try out this method and to look at that book or website. The main purposes of GUUGLE's informative communication, on the other hand, were agenda-setting

(bringing up and pushing new issues relevant to teaching and learning) and momentum-building (showing what was happening at the *Hochschule Bremerhaven* through GUUGLE; who was participating; what progress had been made; what others were doing elsewhere etc.).

7.4 DMOC 4: the nature of CETL in PLCs

According to Figure 6.6, the professors' continued willingness to engage in CETL depends on their view that CETL is an important practice. While this view is encouraged by GUUGLE's communication of reasons for CETL, it also depends on CETL being a positive and beneficial experience for the participating professors. In fact, the social mechanism hypothesised by DMOC 4 – professors experience CETL as an opportunity for collaborative reflection and learning – is essential for the closure of feedback loop 1, one of the two pillars of sustainable CETL. The present section will test the validity of DMOC 4 by analysing the nature of CETL that occurred in PLCs.

To this end, a combination of manifest and latent content analysis is applied to the minutes, photo protocols, written inputs and outputs of PLCs between 2009 and 2013, followed by a cluster analysis. The contents of interest concern the social processes and interactions characterising and shaping CETL in the PLCs. While some of these contents are manifest in the PLC documentations, clearly visible und unmistakably recognisable for the intended readership, most are only present latently. The distinction between manifest and latent content originates in Freudian dream analysis. Latent content requires an interpretive reading of texts. It 'cannot be measured directly but can be represented or measured by one or more [...] indicators' (Hair et al. 1998, 581). For example, minutes of a PLC session might not state explicitly that the participants

reflected on their own practice, yet a sentence like ‘The way we handle student projects is problematic’ implies that reflection on practice took place.

The PLC documents are analysed with respect to 12 questions which can be answered on the basis of (largely) manifest content. The answers, in turn, serve as indicators of (largely) latent aspects of CETL, including how formal or informal, planned or emergent, participative, cooperative and result-oriented the conversations during PLC sessions have been:

- 1) Was the session facilitated by a participant?
- 2) Was there an agenda for the session?
- 3) Was the session dedicated to specific questions or problems? (C1)
- 4) Did participants prepare inputs for this session? (C2)
- 5) Did the participants generate and collect ideas or aspects? (C3)
- 6) Was a photo protocol produced? (C4)
- 7) Did the participants discuss the issue(s) at hand?
- 8) Did the participants reflect on their own practice? (C5)
- 9) Did the participants inform, teach or counsel each other? (C6)
- 10) Did the participants rate or select ideas or aspects? (C7)
- 11) Did the participants make formal decisions? (C8)
- 12) Was there a clear result of the session? (C9)

The documentations of 40 PLC sessions are suitable for analysis. 11 sessions have not been documented to the extent necessary to answer all questions and are therefore excluded from the sample.

To keep things simple, each question is answered with either yes (=1) or no (=0). Questions 1, 2 and 7 are all answered affirmatively. The remaining nine questions are given the variable names C1-C9 (see above) and included in a cluster analysis to identify data patterns. A cluster is a group of objects (i.e. observations or cases) which are similar to each other and at the same time dissimilar to objects not belonging to that cluster. Cluster analysis is the umbrella term for a range of statistical methods to identify clusters in data. The method chosen here is two-step clustering because of its ability to automatically determine the number of clusters based on given statistical evaluation criteria and to calculate each variable's importance for the construction of the clusters (Mooi & Sarstedt 2011, 259-260).

Table 7.2 visualises the entire sample and the results of the cluster analysis. The first column identifies PLCs (letter) and individual sessions (number), followed by the nine columns containing the values of C1-C9 (i.e. the answers to the corresponding questions) for each PLC session. Cells with positive values (=yes) are shaded grey to make data patterns stand out. The bottom row shows the share of cells in the respective column with a positive value (1s). The percentages in the top row result from the two-step clustering procedure and indicate each variable's importance (VI) for the clustering solution, with 100% being the maximum. According to VI, the variables C1-C4 have little influence on the construction of clusters, whose composition hence depends largely on C5-C9.

Table 7.2. Cluster analysis of CETL taking place in PLCs.

VI =	23%	1%	18%	17%	74%	71%	64%	100%	91%
PLC	C1	C2	C3	C4	C5	C6	C7	C8	C9
P6	1	1	0	0	0	0	0	0	1
R3	1	1	0	0	0	0	0	1	0
L2	1	0	1	1	0	0	0	1	0
M2	1	1	1	1	0	0	0	1	0
A2	0	0	1	0	1	0	1	1	0
A5	0	0	1	0	0	0	0	1	1
P4	1	0	0	0	0	0	0	1	1
L5	1	0	1	1	0	0	0	1	1
R1	0	0	1	0	0	0	1	1	1
P5	0	0	1	0	0	0	1	1	1
P1	0	0	1	1	0	0	1	1	1
L1	0	0	1	1	0	0	1	1	1
E4	1	1	0	0	0	0	1	1	1
E3	1	1	0	1	0	0	1	1	1
R5	1	1	1	0	0	0	1	1	1
E1	1	1	1	1	0	0	1	1	1
R8	1	0	0	0	0	0	1	0	0
R2	1	1	0	0	0	0	1	0	0
R7	1	1	0	0	0	0	1	0	0
L7	1	1	0	0	0	0	1	0	0
P2	1	1	0	1	0	0	1	0	0
A1	0	0	1	0	1	0	0	0	0
A3	1	0	0	0	1	0	0	0	0
R6	1	0	0	0	1	0	0	0	0
A6	1	0	0	0	1	0	0	0	0
A4	1	0	1	0	1	0	0	0	0
A7	1	0	1	0	1	0	0	0	0
A8	0	1	0	0	1	1	0	0	0
L8	1	0	0	0	1	1	0	0	0
L4	1	1	0	0	1	1	0	0	0
P7	1	1	0	0	1	1	0	0	0
P8	1	1	0	0	1	1	0	0	0
L2	1	0	1	0	1	1	0	0	0
L3	1	0	1	0	1	1	0	0	0
L6	1	0	1	0	1	1	0	0	0
S1	1	1	1	1	1	1	0	0	0
P3	1	1	1	1	1	1	0	0	0
M3	1	1	1	1	1	1	0	0	0
E2	1	0	0	0	1	0	0	1	0
R4	1	0	0	0	1	1	0	1	0

1s = 80% 45% 50% 28% 50% 43% 23% 43% 30%

Cluster 1: focus on problem-solving

Cluster 2: focus on reflection and learning

Two-step clustering divides the data set into two clusters, which in Table 7.2 are separated by a horizontal line. A look at the most influential variables C8 (Did the participants make formal decisions?) and C9 (Was there a clear result of the session?) immediately suggests an interpretation of the two clusters. Cluster 1 is characterised by CETL involving formal decision-making and entailing clear results. So cluster 1 seems to comprise primarily those sessions where CETL was focused on problem-solving. This interpretation is supported by C7 (Did the participants rate or select ideas or aspects?): evaluating and eliminating alternatives is a typical step in problem-solving. It took place quite often in cluster 1 sessions but never in cluster 2.

The nature of CETL in cluster 2, on the other hand, is revealed by the positive values of the variables C5 (Did the participants reflect on their own practice?) and C6 (Did the participants inform, teach or counsel each other?). Obviously, an emphasis on reflection and learning characterises CETL in this cluster.

What do these findings imply for the validation of DMOC 4? The hypothesised social mechanism asserts that CETL is experienced as an opportunity for collaborative reflection and learning and therefore regarded as an important practice by professors. Cluster 2, representing 60% of the analysed cases, substantiates this hypothesis. However, cluster 1 suggests the existence of a further mechanism. Professors may view CETL as an important practice because it allows them to tackle shared professional issues outside of the regular committee structure.

Committees are bodies of elected representatives which can make binding decisions. Committee meetings need to be prepared beforehand, but unless preparatory work is formally delegated to a subcommittee or working group, which rarely happens, it tends

to be done in isolation by individuals who have a particular interest in the decisions to be prepared. Still more often, issues perceived as important by many professors do not get addressed at all, because nobody wants to do the work on behalf of everybody else, so everybody waits for others to take the initiative. PLCs appear to be seen and used as new opportunities to overcome the deadlock and prevent garbage can processes of decision-making (see section 2.3.1). Problems of general concern have been brought up, analysed and discussed and solutions proposed by PLCs with no mandate and no special personal concernment of their participants.

In the face of the findings of the cluster analysis, it seems necessary to distinguish between two mechanisms in DMOC 4. Professors see CETL as an important practice, not only for the collaborative reflection and learning it makes possible, but also because it is an opportunity for tackling shared professional issues. This finding is fully supported by the analysis in section 7.2 and Figure 7.1.

7.5 Critical voices about the GUUGLE programme

For the GUUGLE team, and especially me as one of the programme's initiators, it is much easier to collect data about why GUUGLE works for whom and in which circumstances, as suggested by Pawson & Tilley (1997), than it is to collect data about why GUUGLE does not work for whom and in which circumstances. Since participation in GUUGLE is voluntary, those who are not interested in it or have been disappointed simply stay away and, when asked for the reason, blame their lack of time (see section 7.2). Moreover, GUUGLE was one of the winners of a nationwide competition of change initiatives in HE institutions and received praise for its 'bottom-up' approach and its ambition to stimulate CETL. Speaking up against GUUGLE under

these conditions means going against the grain of the official discourse and risking being called a whiner. And since the money paying for GUUGLE has been raised exclusively for this purpose and cannot be put to other ends anyway, why would any professor bother criticising openly the way it is being spent?

This section brings together, contextualises and comments on some of the critical remarks about GUUGLE present in the available data. Its purpose is to highlight potentially problematic aspects of the programme which may have escaped the analysis so far, to challenge the current state of the GUUGLE programme theory and to collect further issues to be included in the realist interviews in the third evaluation cycle. The critical comments can be grouped into five thematic categories: preaching to the converted; deficient interfaces; insufficient substitute for leadership and structures; disparagement; and little perceived cultural impact.

7.5.1 Preaching to the converted

A PLC participant who was interviewed in 2011 questioned the ability of the GUUGLE programme to reach and activate professors who are generally less involved: ‘The active ones have been encouraged, but the passive ones remain passive.’ In the 2012 survey, one professor wrote: ‘It seems to me that there is a vast discrepancy between the ambition of GUUGLE and its approaches on one side and the teaching staff on the other. I’m not sure that those who should be reached actually wish to be reached and can be reached.’ Another respondent of the same survey agreed: ‘Unfortunately GUUGLE only reaches the committed. The less committed carry on as before.’

The possibility that GUUGLE might only preach to the converted was identified early on. The first three GUUGLE Forum conferences were preceded by lengthy discussions about making participation compulsory for full-time teaching staff. But each time, the GUUGLE team refrained from recommending this step to senior management, in order not to jeopardise the conference climate. Instead the strategy was to make GUUGLE events as attractive as possible and to rely on the positive word-of-mouth mechanism of feedback loop 1.

After one-third of the professors had joined PLCs in 2010, it was believed that a critical mass of participants might already have been reached and a tipping point passed, beyond which the social dynamics associated with feedback loop 2 would work in favour of CETL. However, this turned out to be a misjudgement. Continued efforts were and are necessary to activate professors to join PLCs. The vast majority of PLC participants in 2012 and 2013 had taken part in previous PLCs. The marginal effort of recruiting a ‘first-timer’ has grown exponentially.

7.5.2 Deficient interfaces

In the 2012 online survey, one respondent wrote: ‘After good discussions in our PLC the question was: what now?’ Similarly, two interviewees in 2011 expressed their satisfaction with the quality of CETL in PLCs and the GUUGLE Forum conferences but deplored the lack of consequence of the conversations and the work that had taken place, blaming deficient interfaces between GUUGLE and the committee structure. This criticism could explain the low path coefficient of .26 between PastGP and IPLC in the Figures 6.3 and 6.6: professors who join PLCs in order to bring about concrete change become disappointed and revise their opinion about CETL being an important practice.

Such criticism is rooted in the participants' understanding of GUUGLE as a platform for problem-solving and results-oriented teamwork. This role of GUUGLE, which is reflected by cluster 1 in Table 7.2, was never really intended by the programme initiators and is therefore not included in the normative programme theory in Figure 5.4. It is an emergent function of GUUGLE, which affects the way PLCs work and need to be managed and facilitated.

The GUUGLE team has been rather slow in recognising and responding to the PLC participants' wish to produce concrete, demonstrable results. Perhaps for too long, the definition of a PLC as 'a group of people sharing and critically interrogating their practice in an on-going, reflective, collaborative, inclusive, learning-oriented, growth-promoting way' (Stoll et al. 2006, 223. See section 4.2.1.) has served as the only benchmark of success. It was only in 2012-2013 that a PLC set itself the goal of actually changing the way teaching is evaluated at the *Hochschule Bremerhaven*. The PLC succeeded because its members worked closely with the committee in charge.

7.5.3 Insufficient substitute for leadership and structures

In his 2011 interview, a professor who had already joined three PLCs and two GUUGLE Forum conferences criticised senior management for using GUUGLE as a 'most welcome opportunity to pass the buck' [to GUUGLE]. Another interviewee could not see any active support for GUUGLE by senior management. Obviously, both comments were critical of the institution's leadership but implicitly, they were also critical of GUUGLE by suggesting that the programme was being instrumentalised or abused.

Another criticism which concerned GUUGLE, even though it was mainly directed at the teaching conditions at the *Hochschule Bremerhaven*, is expressed by the following statement made by a survey respondent in 2012. ‘I regard GUUGLE as a Potemkin village. The conditions for teaching and studying, group sizes, courses which actually permit interaction [with students], the resources available per student etc. have worsened so much over the past 20 years that no superstructure project can mitigate the consequences.’¹¹ Another survey participant wrote a similar comment: ‘Structural and resource limitations cannot be overcome by discussing them [in PLCs]’.

The critical voices effectively assert that the GUUGLE team is compelled to act as a substitute for the educational leadership which senior management fails to provide and to compensate structural deficits resulting from years of funding cuts. These assertions are not entirely unfounded. Inspired by Walker’s (2001) critical professionalism project, my colleague Wolfgang and I saw the pilot PLC (see sections 4.2.1-4.2.2) not only as a means of addressing pedagogical solitude but also as a form of protest against the lack of academic development programme or effort at the *Hochschule Bremerhaven*. This lack may have been regarded by the quoted survey respondents as a sign of missing leadership and resources. Moreover, when in late 2010 GUUGLE-X took off and the combined budget of GUUGLE and GUUGLE-X reached 2.6 million euros, the impression may have been created that GUUGLE had taken over the responsibility for all teaching-related infrastructure and support activities.

¹¹ The term ‘superstructure project’ is the literal translation of *Überbauprojekt*, a German word the commentator seems to have made up to refer to something that is built on top of something else to make it look more impressive or to hide what lies underneath.

Yet if the impression persisted until the online survey in late 2012, something must have gone wrong with the communication of the purpose of GUUGLE. Another explanation of the above comments is that the survey respondents' misunderstanding was deliberate and an expression of their mistrust of senior management, underlining that GUUGLE operates in a political environment.

7.5.4 Disparagement

In this political environment, GUUGLE takes sides, but not with or against individuals, groups, committees or senior management. Through its approach, its events and choice of topics, GUUGLE advocates and promotes the shift from the teaching paradigm to the learning paradigm (Barr & Tagg 1995). In the past, this led to a number of minor clashes between GUUGLE team members and professors who felt provoked in their professional identity as subject experts, and some saw the need to disparage GUUGLE.

For example, during a committee meeting in 2011 one colleague dismissed GUUGLE as a 'Mickey Mouse event producing nothing but hot air'. Another colleague who had received an email with information on upcoming GUUGLE events replied by saying that 'While you waste our time talking about good teaching, I spend my time doing it.' The sender copied this reply to all teaching staff of the *Hochschule Bremerhaven*. In 2012, a survey respondent characterised GUUGLE as 'An empty prestige project which has not achieved any verifiable improvement in teaching.' All three examples involved professors who had never participated in GUUGLE.

But not every derogatory or cynical statement was necessarily intended this way. A professor and one-time PLC participant who was interviewed in 2011 replied to the

question as to what he associated with GUUGLE: 'It stands for money. Better teaching conditions. And a lot of money that is spent on padding such as coffee and cakes.' The fact that three-hour PLC sessions included coffee and sandwiches for the participants, whereas committee meetings of similar length did not, apparently dominated his impression of GUUGLE. Whilst his statement can be read as a pejorative remark about GUUGLE, it can equally be read as a criticism of the *Hochschule Bremerhaven*'s lack of consideration for committee members.

Occasional disparagement, as a reaction to challenges to the status quo, must be anticipated by the challengers. It may be celebrated as a sign that the challenge is successful and taken seriously, but disparaging voices also need to be contained to prevent them from gaining momentum. In the case of GUUGLE, openly cynical or derogatory reactions have been rare. Behind the scenes and off the record, however, more GUUGLE bashing may be going on.

7.5.5 *Little perceived cultural impact*

According to the normative GUUGLE programme theory, over time and with growing experience and practice, professors should begin internalising CETL as part of their professional role. The idea is that, as more and more professors get involved, CETL gradually becomes part of the *Hochschule Bremerhaven*'s organisational culture (see feedback loop 2 in Figure 6.6 and DMOC 5 in Table 7.1). As mentioned before, achieving this kind of cultural change within the limited duration of the GUUGLE programme is a stretch target.

Some first feedback on the progress made was collected during the interviews in 2011. The 18 interviewees were asked what impacts of GUUGLE on the *Hochschule Bremerhaven* they had noticed. At that stage, five of the 14 GUUGLE-experienced interviewees and the four non-participants negated a wider organisational impact. Two further interviewees preferred not to make a judgement, as they had only recently joined the *Hochschule Bremerhaven* and did not have a comparison with the pre-GUUGLE situation.

On the other hand, one interviewee felt that her colleagues had become more aware of the importance of didactical concepts in teaching, that teaching had become more visible, and that teaching staff were actually talking with each other. Another interviewee stated that ‘GUUGLE is affecting the *Hochschule* positively. Group dynamics have improved and non-participants are beginning to justify themselves.’ A third interviewee pointed out that GUUGLE-X would not have been possible without its preparation by GUUGLE, and a fourth interviewee had registered a growing willingness to take seriously issues that do not concern one’s own area of responsibility.

In a second attempt to detect signs of cultural change through GUUGLE, the participants of the online survey in 2012 were asked to what extent they agreed with the statement, ‘We have come closer to the aim of GUUGLE to establish regular collegial exchange about teaching and learning.’ The results for the professorial survey participants are shown in Table 7.3.

Table 7.3. Professors' views regarding the progress made towards regular CETL.

Progress made towards regular CETL (n = 27)	
Disagree	19%
Tend to disagree	11%
Neutral	48%
Tend to agree	22%
Agree	0%
	100%

More professors disagreed than agreed with this statement, almost half the respondents took a neutral position and not one respondent agreed fully. Moreover, correlation analysis reveals that those who agreed the most that CETL was a good idea (see Table 6.5) also tended to agree the most that progress towards regular CETL had been made ($r=.62$; $p=.001$), which suggests that the distribution of answers in Table 7.3 probably reflects a mix of perception and wishful thinking. It has to be concluded, therefore, that until autumn 2012, only little if any cultural impact of GUUGLE was perceived.

7.6 Re-specifying the GUUGLE programme

Based on the insights gained in the second evaluation cycle, the question what works for whom and in what circumstances can be answered with greater confidence and in greater detail than after the first cycle. The DMOCs 1, 2 and 4 of the GUUGLE programme theory (see Table 7.1) were confirmed. However, DMOC 4 seems to be incomplete. Professors regarded CETL as an important practice not only for the collaborative reflection and learning it enables but also, or even primarily, for the opportunity it provides to tackle shared professional issues. The mechanism of DMOC 4 needs to be complemented by this aspect.

The critical voices about GUUGLE suggest that further refinements of the GUUGLE programme may be necessary. Without doubt, GUUGLE fails to appeal to all professors at the *Hochschule Bremerhaven*. Although it is not clear at this stage what causes some professors to be more amenable than others to the messages and approaches of GUUGLE, the fact needs to be reflected by the programme theory. Similarly, if PLC participants have the impression that their work is in vain because of deficient interfaces between PLCs and the institutional leadership and committee structures, then this must be captured by the programme theory and tested in the third evaluation cycle.

DMOC 5 was not confirmed. But since cultural change takes long to materialise, it is not clear whether the lack of perceived cultural impact of GUUGLE is a sign that the programme theory's feedback loop 2 is fundamentally unrealistic or that it only requires more time to kick in. Hence the DMOCs 5 and 6 will be retained in the programme theory and come under further scrutiny in the third evaluation cycle.

7.7 Refining the GUUGLE programme theory

Refining the programme theory does, unfortunately, not make it simpler. Figure 7.3 is an extension of Figure 6.6 and includes the modifications outlined in section 7.6. The path coefficients are omitted. The newly added leakage symbol (↑) indicates causes triggering unintended mechanisms which induce professors not to engage in CETL or to withdraw from it. In the first case, the reasons for CETL provided by GUUGLE do not speak to certain professors (GUUGLE preaches only to the converted) or are even perceived as a threat and set off defensive mechanisms (e.g. disparagement). In the second case, CETL in a PLC is experienced as ineffective for solving key problems when solutions require committee decisions. In both cases, momentum for change gets

lost and the old status quo is strengthened. Moreover, feedback loop 2 is now depicted as dashed arrows owing to the lack of evidence of its existence.

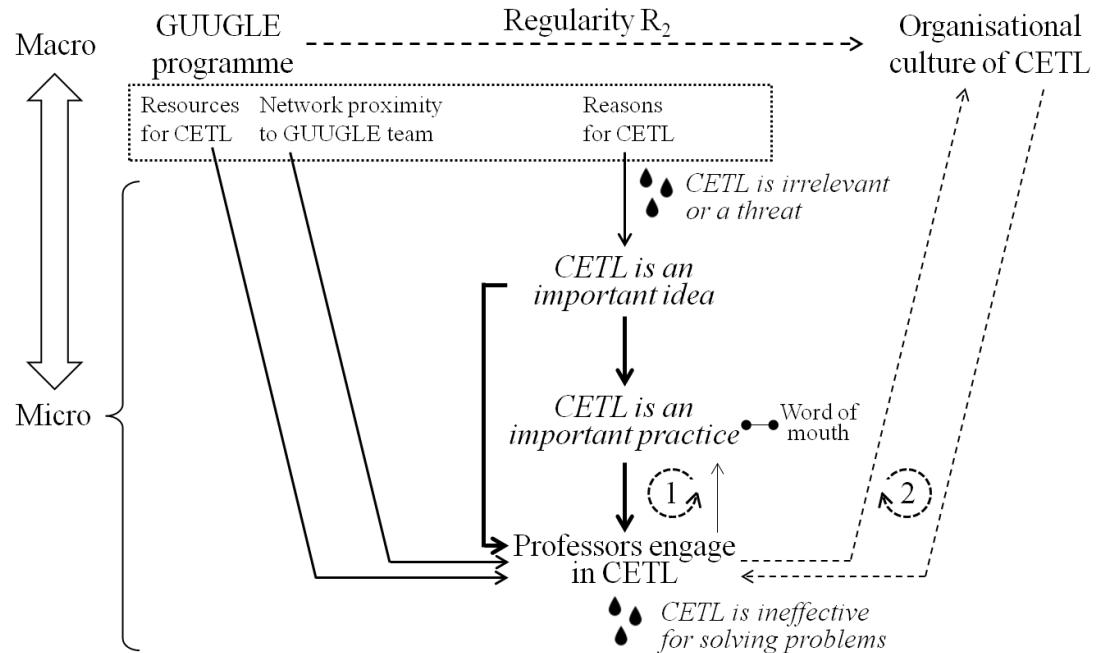


Figure 7.3. Whirlpool diagram of GUUGLE after the second evaluation cycle.

Chapter 8 Third Evaluation Cycle: Realist Interviews

The testing of DMOCs in the first two evaluation cycles had to rely on data gathered prior to the formulation of the GUUGLE programme theory. In some cases, first-best data was unavailable and needed to be substituted by second-best alternatives (see sections 6.2-6.3). In other cases the intended meaning of ambiguous statements uttered by respondents could not be clarified *ex post* (see section 7.5). The third evaluation cycle, which centres on a set of realist interviews conducted specifically for this evaluation study, offers the opportunity to fill some of the remaining gaps and validate or challenge earlier results.

8.1 Deriving the hypotheses

As in the two previous evaluation cycles, DMOCs are derived from the latest version of the GUUGLE programme. Table 8.1 contains the DMOCs 1-6 familiar from Table 7.1. Only the mechanism of DMOC 4 has been amended. Two newly added DMOCs, 1a and 4a, correspond to the two leakage symbols in Figure 7.3 and constitute dysfunctional alternatives to the DMOCs 1 and 4. The mechanism of DMOC 4/4a is reflexive, i.e. it describes the psychic process through which individuals make sense of their own actions and experiences and evaluate them. Therefore, whether professors choose to continue with CETL (DMOC 4) or withdraw from it (DMOC 4a) depends not only on the quality of their CETL but also on the prevailing social conditions (e.g. incentives, collegiality and social pressure) and each professor's individual conditions. Here, the influence of programme context and delta context becomes clearly noticeable.

Table 8.1. The re-refined GUUGLE programme theory in terms of DMOCs.

#	Delta context	+	Mechanism	=	Outcome
1	Communication of reasons for CETL		Reasons trigger learning processes in professors.	=	(a) Professors regard CETL as an important idea.
1a	Communication of reasons for CETL		Communicated reasons are ignored or considered a threat by professors.	=	Professors regard CETL as irrelevant or react with defensive routines.
2	Outcome (a)	+	Professors form expectations about how CETL may benefit them in practice.	=	(c) Professors regard CETL as a (potentially) important practice.
3	Outcome (a) or (c) + 'bottom-up' evolution of GUUGLE + resources for CETL	+	Professors, especially those close to the GUUGLE team decide to use the resources for CETL.	=	(b) Professors engage in CETL.
4	Outcome (b) + resources for CETL	+	Professors experience CETL as opportunity for problem solving and/or collaborative reflection and learning.	=	(c) Professors regard CETL as an important practice (and tell their colleagues: positive word of mouth).
4a	Outcome (b) + resources for CETL		Professors experience CETL as ineffective for solving problems.	=	Professors withdraw from CETL (and tell colleagues: negative word of mouth)
5	Outcomes (a)-(c) + much practice of CETL + resources	+	Professors internalise CETL as part of their professional role.	=	(d) CETL forms part of the organisational culture.
6	Outcome (d) + resources	+	Respecting the cultural norm of CETL is associated with positive social sanctions.	=	(b) Professors engage in CETL.

Also the professors' reactions to GUUGLE's attempts to promote reasons for CETL are contingent upon context. Under favourable social and individual conditions, professors react with learning to the stimuli offered by GUUGLE (DMOC 1), while less favourable conditions, such as the dominance of the professional identity discourses characterised in section 4.1.3, lack of tolerance for mistakes or a general fear of exposing oneself to ridicule, tend to lead to defensive behaviour (DMOC 1a).

8.2 Realist interviews

The data of the present third evaluation cycle were generated for the purpose of testing and refining the GUUGLE programme theory in Table 8.1. The method employed, realist interviews, was developed by Pawson (1995) and Pawson & Tilley (1997) from a critique of the traditional antagonism between structured and unstructured interviewing as a method which is consistent with the logic of realist inquiry.

8.2.1 *The method of realist interviewing*

Pawson & Tilley claim that interviewing in social research is primarily data-driven, as researchers attempt to elicit ‘information which is faithful to the subject’s thoughts and deeds’ (155). The authors criticise that researchers work under the assumption that ‘*the subject and the subject matter of an interview are one and the same thing*’ (ibid, original emphasis). Instead, Pawson & Tilley suggest that data construction should be theory-driven. ‘Thus on the realistic model, *the researcher’s theory is the subject matter of the interview, and the subject (stakeholder) is there to confirm, to falsify and, above all, to refine that theory*’ (ibid, original emphasis).

So the key feature of realist interviewing is a ‘research relationship’ between the interviewer/evaluator and the interviewee/stakeholder, in which the former introduces the latter to the theoretical postulates and conceptual structures under investigation, so that the latter can make an informed and critical contribution to them (ibid, 182). This relationship is depicted in Figure 8.1. The main information flows are indicated by four bold arrows. The upper pathway, which connects the researcher’s theory with the subject’s ideas, represents the realist interview’s distinct teacher-learner function,

whereas the lower pathway in the opposite direction represents the concept refinement function of the interview.

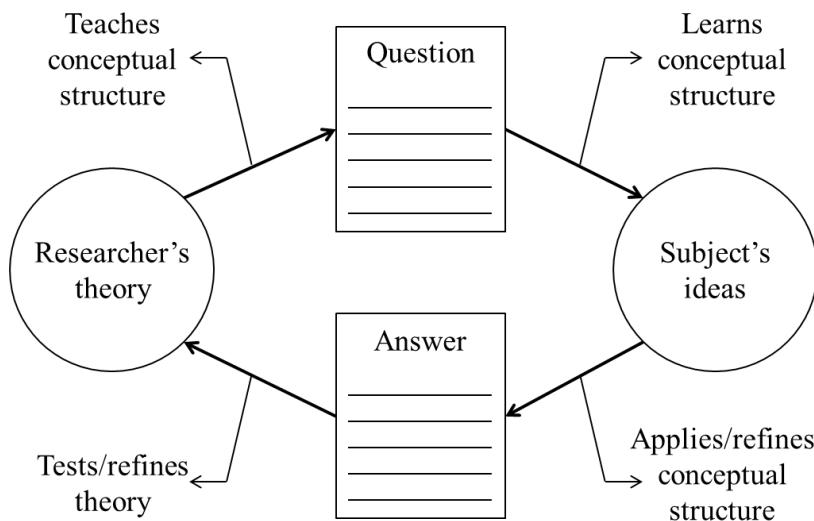


Figure 8.1. Basic structure of the realist interview (ibid, 165).

A realist interview is neither structured nor unstructured; it is both. The structure of the theory to be investigated and the way it is introduced ('taught') by the evaluator give structure to the interview, but the respondent is free to respond in any way she or he sees fit. A crucial aspect of realist interviewing is the choice of interviewees. Its importance follows from the guiding question of Realistic Evaluation: why does a programme work for whom and in what circumstances? As Pawson & Tilley emphasise, different stakeholders know different things about a programme. This 'division of expertise' follows partly from the different roles stakeholders play in or for programmes and partly from the kind of personal experience they gain with them. Programme participants, for example, tend to have a localised perspective, which is best suited to comment on mechanisms (M) hypothesised by the evaluator. Programme managers, on

the other hand, are probably most familiar with the successes and failures, i.e. the outcome patterns (O), of their programmes. They may also be able to assess the influence of contextual conditions (C and D) on programme effectiveness.

8.2.2 *Realist interviews to test the GUUGLE programme theory*

I conducted five realist interviews with professors who differed with regard to their experiences with, and attitudes towards, the programme. The choice of interviewees was guided by Figure 7.3. Four interviewees had more than six years of work experience at the *Hochschule Bremerhaven* and one had been appointed less than two years earlier. Table 8.2 summarises the respondents' relevant characteristics and the foci of the respective interviews.

Table 8.2. Subjects of realist interviews.

Interviewee	Experience	Interview focus
A	Never joined a PLC	Reasons for CETL and participation decision (DMOCs 1/1a and 3)
B	One-time PLC participant	Exit decision (DMOCs 4/4a)
C	Occasional PLC participant	Participation and exit decisions (DMOCs 1-4/4a)
D	Participant of PLC for new professors	Experience with CETL (DMOC 4)
E	Frequent PLC participant	Experience with CETL and cultural impact (DMOCs 4-6)

All interviewees knew the purpose of the interview and had been assured anonymity. Given the limited size of the *Hochschule Bremerhaven* and the information provided in Table 8.2, anonymity prohibits the use of gendered pronouns referring to the respondents. For lack of gender-neutral pronouns, A-E will be treated as female. The

interviews lasted between 40 and 60 minutes. I decided not to use a voice recorder but to take notes (for an exemplary excerpt see Appendix A3). Previous interviews with colleagues had taught me that the presence of a voice recorder could impede the conversation flow and induce interviewees to filter what they say about colleagues and students. This was the case in 2009 (see section 4.1.3) when an interviewee even asked me to stop recording before he continued. In the 2011 GUUGLE feedback interviews, which were not recorded, the respondents were generally much more outspoken.

For the teacher-learner function of the realist interviews, I had prepared hardcopies of Figure 5.2, 5.3 and 7.3, with all arrows drawn as solid lines of the same thickness, to illustrate the basic idea and development of the GUUGLE programme theory and the DMOCs. All other diagrams were available on a tablet computer but never used.

The interviews followed the structure shown in Figure 8.1. Introducing the GUUGLE programme theory by means of the three diagrams took approximately 15 minutes. Especially Figure 7.3 tended to raise questions of understanding. I avoided mentioning that the theory had already undergone empirical testing and refinement, in order not to constrain the respondents' thinking and put pressure on them to conform to what I told them. Once the interviewees signalled that they had grasped the main aspects of the programme theory, I initiated the concept refinement phase by asking them about the theory's general plausibility. This question did not presuppose the respondents' personal experience with GUUGLE and was therefore suitable also for interviewee A. From this point on I discreetly steered the conversations towards the interview foci listed in Table 8.2. At the end I asked the respondents for feedback about their interview. Three of them used the opportunity to say how much they had learned about GUUGLE.

8.3 Testing the hypotheses

The interviews' opening question about the plausibility of the GUUGLE programme theory proved to be an effective icebreaker. Based on the diagrams and my explanations, all interviewees seemed to understand and welcome the theory. Especially the whirlpool image captured their imagination. C suggested ironically that:¹²

‘Reducing GUUGLE to a set of arrows that bridge the micro-macro gap is quite thought-provoking. If the arrows are strings and I’m a micro marionette I wonder who pulls the strings [refers to situational mechanisms]. At least GUUGLE is trying to give us our own strings to pull [refers to transformational mechanisms].’

Interviewee B came up with a similarly ironical interpretation of:

‘the bathtub metaphor: “wash my fur but don’t make me wet” [¹³] could be the slogan of many of our colleagues who complain about this and that and call for change, but when GUUGLE says, let’s work together to accomplish the change, they are the first to duck.’

Further remarks about the programme theory as a whole are included in section 8.3.6 below. Next, the interviewees were asked to comment on each DMOC individually.

8.3.1 DMOC 1: Learning to regard CETL as an important idea

According to DMOC 1, GUUGLE offers reasons for CETL, triggering learning processes in professors, who come to regard CETL as an important idea. This hypothesised causal mechanism was supported by four of the five interviewees. Yet

¹² As the quotations in this chapter are based on translated interview notes, not voice recordings, they are only approximately accurate.

¹³ This is a German proverb which characterises the attitude of somebody who wants the benefits (a clean fur) without bearing the costs (getting wet).

only the recently appointed professor D agreed fully with this hypothesis, using herself as an example of someone who started teaching with a very limited idea about the teaching-learning nexus and who increasingly realised how complex good teaching actually is and how much she can benefit from sharing experiences with colleagues.

Respondent E also agreed with DMOC 1 but suggested that a positive basic attitude towards and curiosity about teaching and learning is a prerequisite for the GUUGLE communication strategy to work. So according to E, the validity of DMOC 1 depends on certain individual characteristics of the professors.

In principle, also respondent B agreed with DMOC 1. GUUGLE did teach her something that made her join a PLC, she said. But the main lesson was not that CETL is an important idea; it was that ‘the *Hochschule* does not have to be the way it is. It is ours to shape and this means we have to talk a lot more with one another’. For her, GUUGLE exemplifies how individual initiative can make a major difference even in the bureaucratic, regulated environment of a public HE institution.

A and C insisted that GUUGLE had not changed their views on CETL, yet their positions could not have been more different. Respondent A claimed that GUUGLE failed to provide a single compelling reason why CETL might be important:

‘I don't see why it matters to my teaching what others think and say about their own teaching. We teach different subjects, teach differently and are different people. Exchanging experiences and viewpoints is all very nice but it comes at the cost of mounting time pressure. [...] for me the best way to improve my teaching is to have enough time to prepare.’

So A rejected DMOC 1. Respondent C, on the other hand, did not change her mind about CETL because she had already been aware of its importance before GUUGLE.

For her, providing a complex service such as a degree programme demands substantial coordination and responsiveness, which in turn requires continuous CETL:

‘That we learn from one another is just one aspect. I see consistent service delivery as even more important. GUUGLE keeps reminding us of how demanding our job is if we want to do it well. Yes, this is what GUUGLE does: it reminds us of what we are really here for. GUUGLE is the antidote against negligence and dulling.’

So according to C, the learning triggered by GUUGLE is re-learning. The quotation is thus a confirmation of DMOC 1.

8.3.2 DMOC 1a: Perceiving GUUGLE as a threat

DMOC 1a states that the reasons for CETL which GUUGLE provides get ignored or are considered a threat by professors, leading them to regard CETL as irrelevant or to react with defensive routines. DMOC 1 and 1a are only mutually exclusive at the level of the individual professor but not at the collective level. The interviewees could agree with both hypotheses without contradicting themselves. Their reactions to DMOC 1a were mixed: A rejected it, C questioned it, B, D and E accepted it.

Respondent A, as a non-participant of GUUGLE, disliked being represented by the leakage symbol in Figure 7.3. She doubted that CETL was perceived as a threat by professors and suggested that by making this claim, the programme theory was overestimating the importance of CETL and GUUGLE for most professors. A’s main objection against DMOC 1a was the claim that GUUGLE provides reasons for CETL that get ignored:

‘What GUUGLE communicates may be reasons for some but not for others and certainly not for me. [...] if we spend time talking with our colleagues, which we do by the way, it must really be worth it and have a useful, very concrete output.’

A's point that reasons are not universally valid but individually constructed is consistent with E's view that the perceived importance of CETL depends on certain individual characteristics of the professors (see section 8.3.1).

Also C questioned the programme theory's claim that GUUGLE provides reasons for CETL. In her view, GUUGLE can only try to appeal to the professors' personal ambitions to excel in their jobs. For C, the problem is that teaching is little valued institutionally:

'I get my money whether or not I put great efforts into my teaching. Even if students complain about poor teaching it has no material consequences. All that matters for the *Hochschule* is the number of hours I stand in front of the students. [...] as long as there are no serious [extrinsic] incentives for better teaching [...] GUUGLE won't be sustainable. No incentives equal no reasons.'

B and E made only brief comments on DMOC 1a. Respondent B suggested replacing the word 'threat' by 'irritation' or 'nuisance'. And E noted a little pejoratively that times are changing and some realise it later than others. D, finally, could not imagine GUUGLE being perceived as a threat but agreed that it might make 'teaching traditionalists' feel uncomfortable and avoid GUUGLE events.

8.3.3 DMOC 2: Learning to regard CETL as an important practice

According to DMOC 2, professors who believe that CETL is an important idea form expectations about how CETL can benefit them concretely and start regarding it as a potentially important practice. This DMOC bridges the gap between the idea of CETL and the actual engagement in it. For B, D and E, this hypothesis seemed rather obvious and elicited only minor, approving remarks from them. All three respondents felt that an

important idea guides action and therefore considered the outcome of DMOC 2 and the link to DMOC 3 as natural and unproblematic. C, however, disagreed:

‘The model’s distinction between idea and practice is useful and also necessary. Many will like the idea of CETL. But what does CETL mean for my working life if I take it seriously? How much effort will it take? How will my colleagues react? Will it be worth it? I think this is where some of us say, the idea of CETL is important but it’s not for me.’

In other words, C missed a second possible outcome of DMOC 2, one which reflects professors’ hesitation or inertia. A, finally, was sceptical about DMOC 2 because of what she saw as a confusion of the abstract and the concrete. In her view, embracing an abstract idea is a different thing from acting upon it. ‘Most people like the idea of harmony, yet their behaviour towards others is often inconsiderate, self-centred and aggressive.’

8.3.4 DMOC 3: Engaging in CETL

DMOC 3 states that professors who find CETL an important idea and practice decide to use the available resources to engage in CETL, and that the propensity to do so increases with the professors’ proximity to the GUUGLE team. The respondents’ reactions to this DMOC were largely affirmative. The network proximity effect was taken for granted. Comments focused on the (in)sufficiency of the resources provided by GUUGLE. The teaching load reductions for PLC participants were generally welcomed as a good start and positive gesture, but A pointed out a complication:

‘In our faculty [...] the problem is also that we have no substitute teaching staff for many courses, especially if we need a substitute for one term only. So we can’t simply reduce our teaching hours. I don’t think this would stop me from joining a PLC if I really wanted, but it’s an issue.’

B highlighted a similar problem:

‘I assume that most of the attendees [of PLCs] belong to the group of highly committed colleagues who always do more than they must and therefore have accumulated plenty of overtime anyway. Reducing their teaching obligation probably only increases their overtime further.’

For E, the critical resource was not the reduced teaching load but the coordination of PLC members’ packed timetables and the protection of time slots for PLC sessions. D, who had participated in the PLC for newly appointed professors, generally agreed with DMOC 3 but stressed that it did not describe her own situation, since new faculty members are required to join an introductory PLC during their first year.

Only C was unsure at first whether to accept or to reject DMOC 3. She felt that if the professors had a strong incentive to join PLCs or to engage in CETL, neither resources nor the network proximity to the GUUGLE team should matter, because the professors would find ways to make it possible. But C admitted that, given the absence of an incentive scheme, DMOC 3 was probably valid.

8.3.5 DMOCs 4 and 4a: Experiencing CETL as helpful or ineffective

DMOC 4 describes CETL as an opportunity for solving shared professional problems and/or for collaborative reflection and learning. The experience of helpful and constructive conversations with colleagues confirms the participating professors’ view of CETL as an important practice. This hypothesis is essential for feedback loop 1 and the positive word-of-mouth effect. However, the low path coefficient of .26 (see Figure 6.6) suggests that in practice DMOC 4 is unstable and vulnerable.

In the realist interviews, feedback on DMOC 4 was often inseparable from comments on DMOC 4a, which is why this section treats both hypotheses together. DMOC 4a is the dysfunctional alternative to DMOC 4, in that professors experience CETL as ineffective for solving key problems because of deficient interfaces between PLCs and committees with decision power. As a result, professors withdraw from CETL and may also tell their colleagues about their experience (negative word of mouth).

None of the five interviewees rejected the two hypotheses completely, nor did anybody accept them as they were. A, for instance, who had never participated in a PLC, doubted the positive word-of-mouth effect on the grounds of her own experience, according to which 'Professors are not gregarious animals'. D, on the other hand, confirmed that her PLC participation had made a colleague on her corridor curious. They had started having conversations about topics which had come up in the PLC.

B and D had had positive learning experiences and developed deeper collegial relations through their PLCs, whereas C and E reported rather mixed PLC experiences. The consequences they had drawn for themselves, however, were completely different, and so were their views on DMOC 4/4a. D's positive experience had motivated her to actively seek advice from more seasoned colleagues on how to handle difficult students. B, by contrast, despite her similarly positive experience, had decided not to continue with another PLC because she felt she could not afford the time. The opposite decision was made by E, in spite of her mixed experiences in one PLC. She had joined another one not only for her own sake but also to set an example for others and to support GUUGLE. C, on the other hand, had quit a PLC half way through out of frustration with her fellow attendees and partly also with herself:

‘The PLCs I attended were never really a problem-solving forum. We talked about problems, and maybe some solutions were suggested, but nobody took any action to implement changes. [...] I don't think any of my PLCs had a tangible impact. [...] Yes, PLCs are an opportunity for problem solving, but the opportunity tends not to be seized. Unfortunately I have to include myself in this. My regular workload is too high to take on further tasks.’

So of the four interviewed professors who had joined PLCs in the past, D had confirmed DMOC 4 and C had confirmed DMOC 4a through their respective decisions, whereas B and E had confirmed the critical realist view that in open systems, tendencies like the ones hypothesised by the DMOCs produce demi-regularities at best (see section 3.1.2).

Three of the five interviewees felt that the mechanism of DMOC 4a was too restrictive. According to them, professors withdraw from CETL when they get the impression that their engagement is ineffective. This includes problem-solving but also collaborative reflection and learning or any other expectation participants may harbour.

8.3.6 DMOC 5 and 6: Turning CETL into culture

The DMOCs 5 and 6 constitute feedback loop 2. They hypothesise that after extensive experience with CETL, professors internalise it as part of their professional role, and CETL gradually becomes a social norm. Since social norms are generally enforced by social sanctioning mechanisms, professors would have an additional incentive (=reason) to contribute to the collegial exchange.

In the second evaluation cycle, the DMOCs 5 and 6 were disconfirmed but retained in the GUUGLE programme theory for further investigation. The realist interviews did not offer much support for them either. Respondent A, for example, felt that the whole programme, and especially feedback loop 2, rests on an unrealistically positive

conception of man, in that it assumes that professors have an inherent drive to be outstanding academic teachers. She saw the programme as a projection of the GUUGLE initiators' own particular ambitions and normative beliefs onto others. She also suggested that GUUGLE's emphasis on CETL might be an expression of Wolfgang's and my background in business studies, where 'the soft, social side of teaching is more important than in engineering, where only hard facts count'.

B, who had stopped joining PLCs for lack of time despite positive experiences, said she would not mind spending more time in conversations about teaching and learning if this was expected of her by her colleagues. So she agreed with the possibility of DMOC 5 but not with its actuality, because she did not see how the mechanism of DMOC 5 could be triggered under the current work organisation, in which no professor depends on another. C repeated her argument that the GUUGLE programme is unsustainable without extrinsic incentives for CETL. E felt that establishing CETL as a cultural norm would require continuous efforts over a very long time. Her hope was that the 'young generation', i.e. newly appointed professors, could be socialised into the habit of cooperating and conversing more with one another than the 'old generation'. Only D, as a member of the 'young generation', was somewhat more optimistic about DMOC 5, provided that at least some professors led the change by example.

8.4 Re-specifying the GUUGLE programme

Instead of reconfirming hypotheses already confirmed in the first two evaluation cycles, the realist interviews revealed an unexpected diversity of attitudes, preferences, perceptions, reasons and decisions among just five individuals. For Pawson (2013), this would come as no surprise:

‘People enter programmes at the margins or sometimes quite tangentially. They have an existing life outside programmes. There are always other programmes. Life offers many new opportunities besides programmes. And once in the ambit of a programme, there are many opportunities to quit or stay’ (116).

On the one hand, the five respondents’ answers provide a much deeper insight into the inner workings of GUUGLE, and especially into the complexity of the mechanisms involved, than the data analysed in the two previous evaluation cycles. For the GUUGLE programme management in particular, the informed and differentiated responses of the realist interviews will be of substantial value. On the other hand, concluding under these conditions what parts of GUUGLE work for whom and in what circumstances is quite daring and can hardly do justice to the nuanced accounts presented in section 8.3.

Table 8.3 is an attempt to summarise the realist interviews by crudely categorising and coding the responses (agreement=1, partial agreement=.5 and disagreement=0) and computing the shares of affirmative evaluations by DMOC and interviewee. Cells with positive values are again shaded grey to make data patterns stand out.

Table 8.3. Summary of the realist interviews.

DMOC	Interviewee					Total	Agreement
	A	B	C	D	E		
1	0	1	1	1	1	4	80%
1a	0	1	0	1	1	3	60%
2	0	1	0	1	1	3	60%
3	1	1	1	1	1	5	100%
4	0.5	0.5	0.5	1	0.5	3	60%
4a	0.5	0.5	1	0.5	0.5	3	60%
5	0	0	0	0.5	0	0.5	10%
6	0	0	0	0.5	0	0.5	10%
Total	2	5	3.5	6.5	5	22	
Agreement	25%	63%	44%	81%	63%		55%

Obviously, the mechanisms which GUUGLE aims to trigger through the resources and reasons for CTEL it offers do not work for all. The interviewees' rates of agreement with the eight DMOCs range between 25% and 81%, with B, D and E each agreeing with over 60% of them. To quantify the response pattern (dis)similarity, a hierarchical clustering procedure can be performed. This procedure joins objects together into successively larger similarity clusters. Figure 8.2 shows the dendrogram resulting from hierarchical clustering using the data in Table 8.3. From left to right, the number of clusters declines from five (each interview is a cluster) to one (all interviews form a single cluster). The further to the left clusters are merged, the more similar they are. Similarity is measured in squared Euclidean distance.

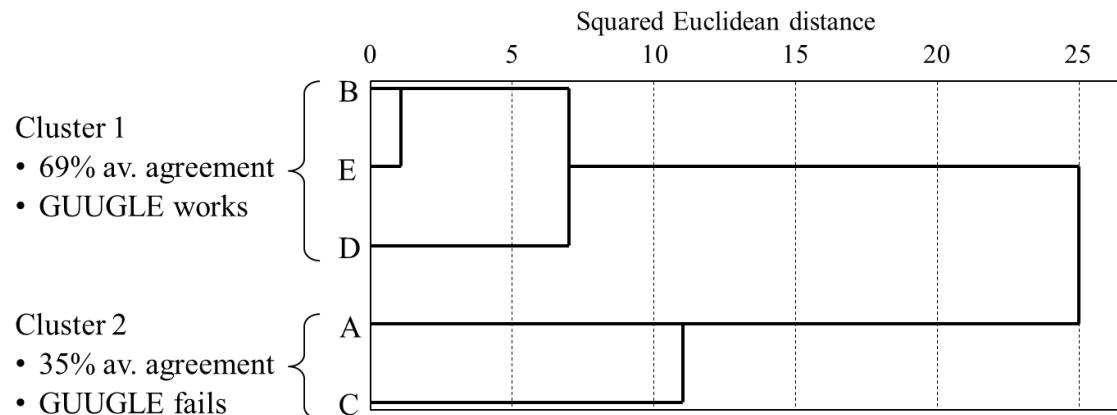


Figure 8.2. Hierarchical clustering dendrogram of the five realist interviews.

The interviews with B and E are shown to be the most similar; in fact, when reduced to their response categories in Table 8.3 they are identical. The next most similar is the interview with D. As noted above, the three respondents share a fairly high rate of agreement with the DMOCs and, as suggested by their interviews, GUUGLE is working for them. A second cluster is formed by A and C, who share a low rate of agreement

with the DMOCs and for whom GUUGLE is not working. But while C tried it out and quit, A never even gave it a try. Her attitude and beliefs led her to decline invitations to participate in a PLC. Hence, A's answers reflect her individual conditions which seem to have disabled the GUUGLE programme mechanisms.

Despite the heterogeneity of the five realist interviews, a few conclusions can be drawn about what parts of GUUGLE work for whom and in what circumstances. Feedback loop 2 (DMOCs 5+6) was disconfirmed once more, while DMOC 3 was confirmed unanimously and DMOC 1 was confirmed by all interviewees except one. Most respondents were unsure about the validity of DMOC 4 and 4a, flagging a potential problem with feedback loop 1 which includes DMOC 4. The most controversial were DMOC 1a and 2. Membership in cluster 1 and 2 coincided perfectly with the acceptance and rejection of these hypotheses.

8.5 Refining the GUUGLE programme theory

On the basis of the realist interviews, three refinements of the GUUGLE programme theory seem necessary. First of all, respondent A made the important point that GUUGLE does not provide reasons for CETL, as claimed in Figure 6.6. The GUUGLE communication raises issues in teaching and learning, addresses related problems in need of solutions etc. But whether professors take this as reasons for CETL depends on them. Replacing 'Reasons for CETL' by 'Raises issues in teaching & learning' in the GUUGLE box in Figure 8.3 stresses the subjective nature of reasons and the loose coupling between GUUGLE activities and the mechanisms they trigger in professors.

A second refinement takes A and C's criticism of DMOC 2 into account. The interviewees' arguments can be subsumed by the concept of cognitive dissonance: professors may support the idea of CETL and even deplore the lack of such conversations in their collegial environment, yet they might not want to change their own habits. According to Festinger (1957),

'1. The existence of dissonance, being psychologically uncomfortable, will motivate the person to try to reduce the dissonance and achieve consonance. 2. When dissonance is present [...] the person will actively avoid situations and information which would likely increase the dissonance.' (3)

To reduce the dissonance between their belief and their practice, professors may assert that CETL is neither possible (working conditions discourse) nor necessary (professional identify discourse). They may blind out GUUGLE's messages (selective perception), ignore invitations to GUUGLE events and evade anything that causes them further dissonance. To embed this effect in the programme theory, cognitive dissonance is added in Figure 8.3 as a further cause of 'leakage' of CETL participants.

The third refinement responds to the view of three interviewees that the mechanism of DMOC 4a is too restrictive. Professors may withdraw from CETL when they perceive it as ineffective not only in regard to problem solving but equally in regard to meeting other expectations they may have. Hence the leakage label is changed from 'CETL is ineffective for solving problems' to 'CETL is ineffective'. The low path coefficient of .26 in Figure 6.6 between the experience of CETL and the view that CETL is an important practice is probably a consequence of some professors' disappointment with CETL. This is GUUGLE's Achilles heel and puts feedback loop 1 at risk. To draw attention to it, the reflexive mechanism is marked with a warning sign in Figure 8.3.

Feedback loop 2 which received no empirical support during all three evaluation cycles will be retained as a normative component in the GUUGLE programme theory. Without anchoring CETL in the organisational culture of the *Hochschule Bremerhaven* or some of its communities of practice, it will be hard to sustain once the GUUGLE programme comes to an end.

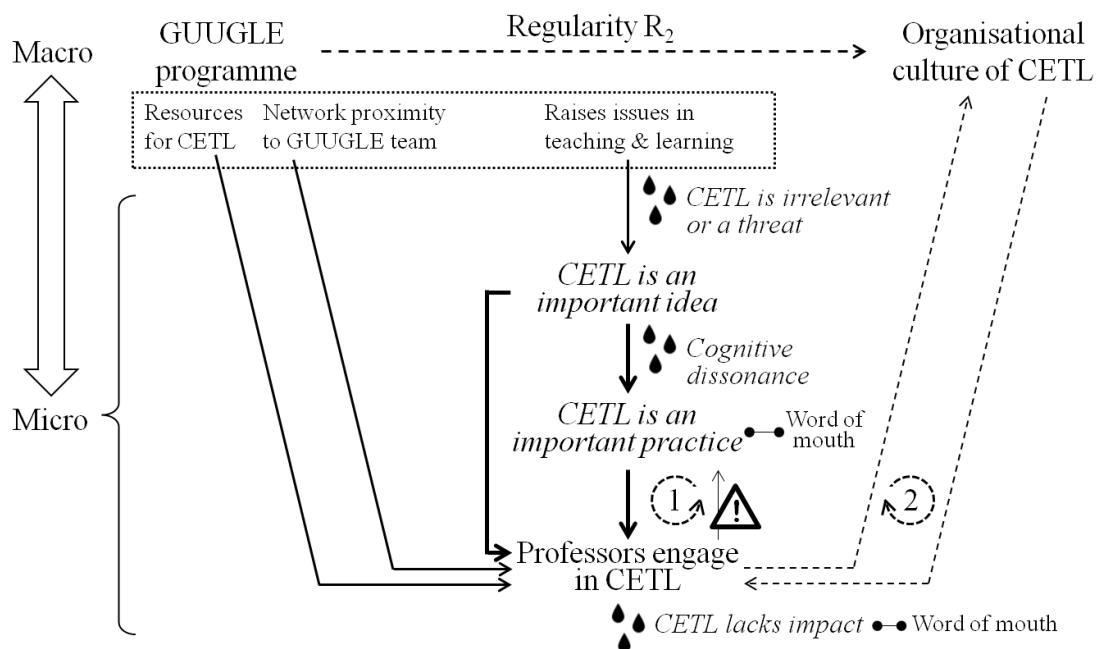


Figure 8.3. Whirlpool diagram of GUUGLE after the third evaluation cycle.

Chapter 9 Fitness for Purpose – A Discussion

The evaluative research question of this thesis formulated in section 1.2 reads: *How fit is GUUGLE for the purpose of overcoming pedagogical solitude, and how might its fitness be enhanced?* After formulating, testing and refining the GUUGLE programme theory this question can now be addressed and the formative evaluation completed.

9.1 Operationalising the concept of fitness for purpose for GUUGLE

In order to assess the GUUGLE programme's fitness for purpose, the concept of fitness for purpose first needs to be operationalised. According to the casual definition in section 1.2, it is a quality criterion which is met when something does what it is intended to do. But since GUUGLE is still in progress with possibly six more years to go, whether it does or will eventually do what it is intended to do may only be assessed *ex post*. Therefore, a substitute criterion or fitness-for-purpose predictor is needed that can already be assessed in the course of the programme.

Figure 9.1 depicts the approach taken in this chapter to construct the substitute criterion, *prospective fitness for purpose*. Its starting point is the normative GUUGLE programme theory introduced in chapter 5. If this theory does not contradict established social science theory, and if it shows how pedagogical solitude can be overcome through GUUGLE interventions, then GUUGLE can be considered *theoretically fit for purpose*. The programme's *practical fitness for purpose*, on the other hand, requires that GUUGLE has been implemented in accordance with its normative theory and that empirical data substantiate the programme's effectiveness by documenting that pedagogical solitude receded at least locally and/or temporarily following the

programme's implementation. This criterion ensures that the formative evaluation of GUUGLE complies with Scriven's (1991a) requirement that 'the formative evaluation should *at least* provide a preview of a summative evaluation' (28, original emphasis).

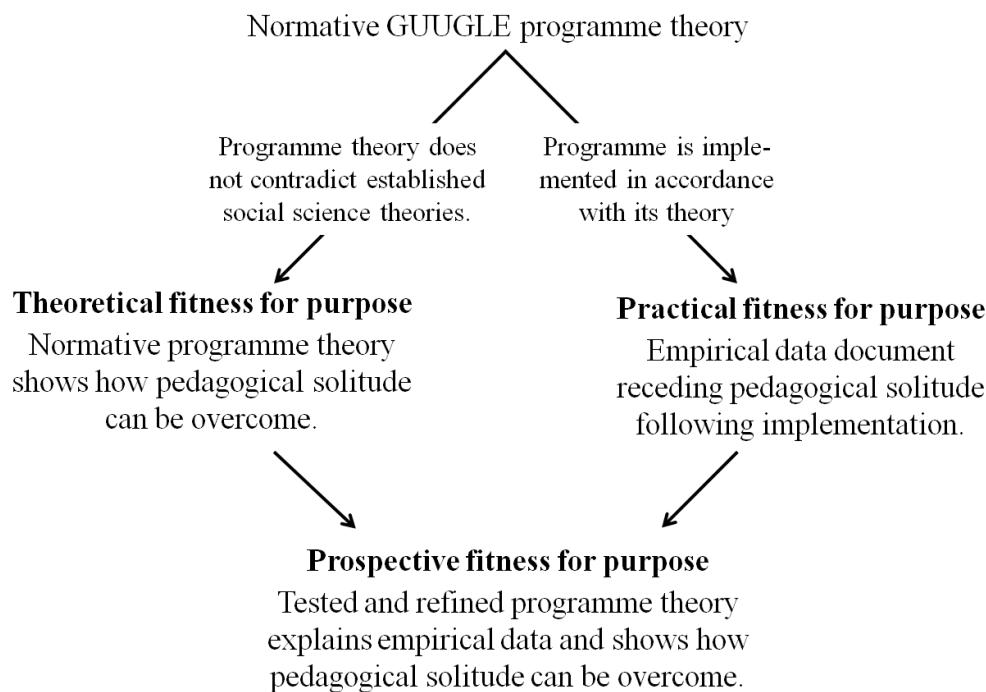


Figure 9.1. Assessing GUUGLE's prospective fitness for purpose

Finally, *prospective fitness for purpose* is given if theoretical and practical fitness have been demonstrated and if the empirically tested and refined GUUGLE programme theory is able to explain the available data on the impact of GUUGLE and shows how pedagogical solitude can be overcome through GUUGLE programme interventions.

9.2 GUUGLE's theoretical fitness for purpose

The normative GUUGLE programme theory of chapter 5 demands that resources (e.g. time, space and organisation) be made available to render CETL possible and that

reasons (e.g. arguments, incentives) be provided to position CETL as an important idea. Resources expand people's capacities to act and reasons influence their choices. As the professors make their own positive experiences with CETL, feedback loop 1 kicks in, establishing CETL as an important practice, which constitutes a reason for CETL in its own right. With some delay, the continued engagement in CETL by a growing share of professors triggers feedback loop 2, which eventually anchors CETL in the organisational culture. Once this is achieved, GUUGLE no longer needs to provide reasons for CETL because the two feedback loops (re)produce their own reasons. Resources for CETL will still be needed but not from GUUGLE because at that stage, the recognised practical importance and the cultural significance of CETL ensure that organisation-internal resources get mobilised and dedicated to CETL.

Since the normative programme theory explains plausibly and by means of established social science theories (see chapters 3 and 5) how pedagogical solitude can be overcome by GUUGLE, the programme's theoretical fitness for purpose can be regarded as given. What is more, since the normative programme theory's formulation is deliberately generic and with no restrictions to particular contextual conditions, GUUGLE's theoretical fitness for purpose also extends to HE institutions other than the *Hochschule Bremerhaven*, which is an important precondition for the transferability of the GUUGLE programme to other contexts.

9.3 GUUGLE's practical fitness for purpose

Practical fitness for purpose demands empirical evidence of the effectiveness of GUUGLE interventions to reduce pedagogical solitude. Such evidence was presented and analysed in all three evaluation cycles. The first cycle validated the achievement of

the normative outcomes (a)-(c) of the GUUGLE programme theory, namely that many professors regarded CETL as an important idea and practice and that they practiced CETL in PLCs and other GUUGLE events. The second and third evaluation cycle drew on feedback interviews from 2011, PLC minutes from 2009-2013 and the realist interviews from 2013, confirming in various ways that GUUGLE activities had been effective in stimulating and sustaining CETL.

9.3.1 A theory-implementation gap

In addition to empirical evidence, practical fitness for purpose requires that the programme's implementation be faithful to the theory on which it is based. Since the normative GUUGLE programme theory was only formulated after the programme's implementation, it may seem strange to question the faithfulness of the implementation to its theory. However, the realist interviews with A and C (see section 8.3.2) and also bits of other interviews suggest that a key requirement of the normative programme theory has not quite been met by the GUUGLE implementation, namely *to offer reasons for CETL*. In her realist interview, respondent A described the problem as follows: 'What GUUGLE communicates may be reasons for some but not for others and certainly not for me.' The consequence is depicted in Figure 8.2: whilst all professors were targeted with the same messages, arguments and offers, their subjective perception of these varied strongly, triggering different reactions which are captured exemplarily by the clusters 1 and 2 of the dendrogram.

To get to the bottom of this problem, the notion of 'reason' needs to be unpacked and disambiguated. Since GUUGLE aims to motivate certain actions on the part of the professors, the focus is on practical (i.e. action-related) reasons and not on epistemic

ones (justifying some belief).¹⁴ Practical reasons can be normative or motivating or both. A normative reason is a rational or moral argument as to why some action is justified, appropriate or required. Normative reasons are reasons *for* action. By contrast, a motivating reason is a complex of beliefs and desires that motivates action. Motivating reasons act as triggers of action-formation mechanisms in Coleman's bathtub scheme and can thus be said to possess causal power, a view shared also by Bhaskar (see section 3.1.2). Motivating reasons are reasons *of* action and therefore explanatory.

It is possible to be motivated by a normative reason, but this constellation is rare, as humans are neither perfectly rational beings nor saints. In the implementation of GUUGLE, however, this important detail may have been overlooked. That GUUGLE works for some but not for others could at least partly be the result of treating normative and motivating reasons as identical and hence professors as perfectly rational or saints. Respondent A's intuition was right when she criticised GUUGLE's conception of man as unrealistically positive (see section 8.3.6).

9.3.2 *The problem of offering motivating reasons for CETL*

Pawson (2002) stresses that change in people's behaviour results not from programmes but from 'the underlying reasons or resources that they offer subjects' (342). So what underlying *motivating* reasons does GUUGLE offer professors who are not already intrinsically motivated to participate? The programme involves no material or reputational incentives, guarantees no enhanced teaching quality, no workload

¹⁴ This section draws on Smith (1994, 94-98), Parfit & Broome (1997) and Crisp (2006, ch. 2).

reductions, no happier life nor does it threaten non-participants with sanctions. With no carrots or sticks at hand, only sermons (Bemelmans-Videc, Rist & Vedung 2010) and rational arguments remain for GUUGLE to influence the professors' behaviour. But sermons and rational arguments provide *normative* reasons and, as mentioned before, tend to work only for saints and perfectly rational individuals. Hence GUUGLE does not offer any *new* motivating reasons for CETL and can therefore only reach the already converted (see section 7.5.1) and at best some of the undecided.

Further empirical support for this critical point comes from the content analysis in chapter 7. Figure 7.1 depicts professors' stated motivating reasons for participating in GUUGLE, which are: addressing important professional issues, having conversations with colleagues and satisfying one's curiosity. The first reason is independent of GUUGLE. The second reason reflects a social need which may have lain dormant and only been activated by GUUGLE. And curiosity is a stable personal characteristic. So none of the three motivating reasons can be attributed to GUUGLE.

Also, the content analysis of the GUUGLE programme's communication, summarised in Figure 7.2, gives no clues as to what new motivating reasons it may have offered the professors to change their behaviour. Communication about teaching, learning and GUUGLE activities may, of course, direct professors' attention to certain topics and stimulate conversations – which seems to work well for a certain share of professors – but the motivating reasons to engage in conversations are rooted in the professors' own attitude towards colleagues and students, their professional ambitions and intrinsic motivation to become better academic teachers etc.

9.4 GUUGLE's prospective fitness for purpose

In section 9.1, GUUGLE's prospective fitness for purpose was defined as a substitute criterion or predictor of its actual fitness for purpose, which can only be evaluated ex post. Prospective fitness for purpose requires that the programme's theoretical and practical fitness have been established and that the tested and refined programme theory is able to explain the available data about the impact of GUUGLE and to show how pedagogical solitude can be overcome through the programmed interventions.

While GUUGLE's theoretical fitness for purpose can be regarded as given, the practical fitness for purpose of the programme's implementation has been shown to be limited to the already converted, due to GUUGLE's inability to offer new or additional motivating reasons to engage in CETL. Until this problem is resolved, it is inherited by GUUGLE's prospective fitness for purpose. But as the three evaluation cycles have shown, 'unconverted' professors were not the only ones who did not become loyal GUUGLE participants. Figure 8.3 contains three leakage symbols, each of which indicates an empirically confirmed 'point of exit' at which GUUGLE tends to lose a certain share of its potential or actual participants. Figure 9.2 depicts the sequence of exit points from Figure 8.3 as a multi-stage self-selection process.

In stage 1, the 'converted' who believe that CETL is an important idea stay in the process, whilst professors who are indifferent about CETL or who reject it cannot be reached by GUUGLE. In stage 2, the transition from idea to practice takes place. The 'activated' participate in GUUGLE events, whereas those who like the idea of CETL but fail to overcome their own barriers to action suffer from cognitive dissonance. Stage 3 is where the sustainability of CETL is decided by the share of 'repeaters' and where

those who experience CETL as lacking impact withdraw from it. Together, these three exit points pose a threat to GUUGLE's prospective fitness for purpose.

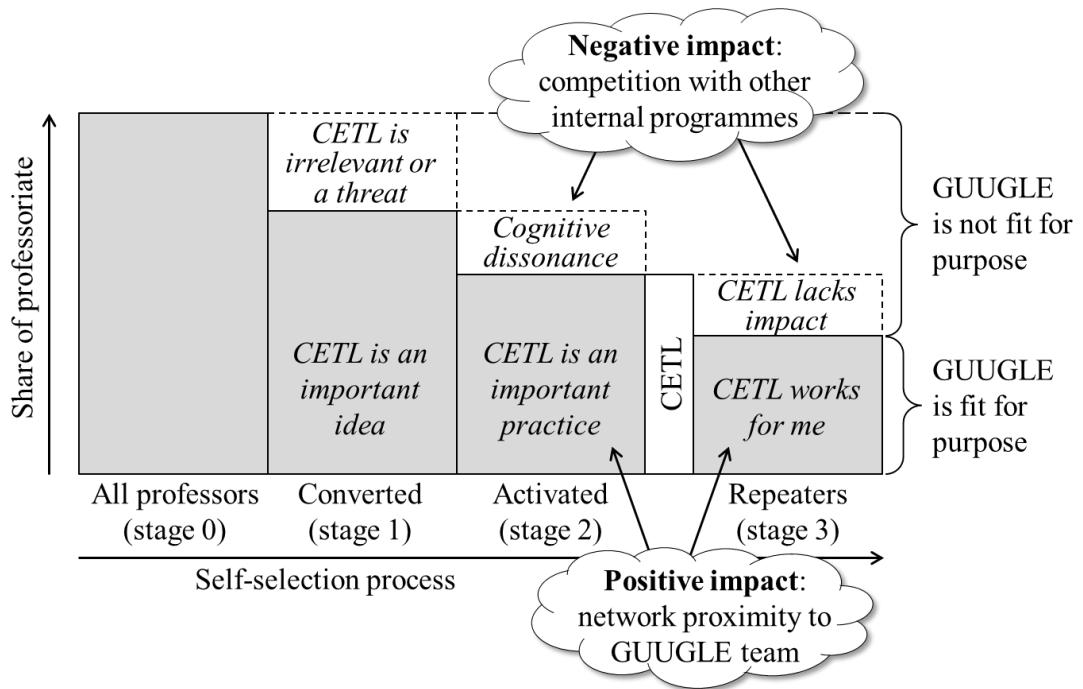


Figure 9.2. The GUUGLE funnel of self-selection.

Figure 9.2 also highlights the impact of two contextual factors on participation choices. Since the start of GUUGLE in 2009, several new change initiatives have been launched at the *Hochschule Bremerhaven*. Some of them originated from, or have been inspired by, the GUUGLE programme, the most important one being GUUGLE-X. Contrary to Pawson & Tilley's (1997) assumption, that social programmes do not significantly affect their contextual conditions, GUUGLE has triggered various changes in its setting. Ironically, GUUGLE now finds itself in the position of having to compete with its own offshoots not only for management attention but also for (potential) participants' attention, time and commitment. Interviewee B's remark that 'most of the attendees [of PLCs] belong to the group of highly committed colleagues who always do more than

they must' (see section 8.3.4) highlights a consequence of this competition: especially professors who are known to be active and supportive are being wooed and ensnared by every new initiative. Being one of them, B chose not to continue with another PLC, despite her good experiences.

The second contextual condition affecting GUUGLE participation choices, the professors' network proximity to the GUUGLE team, has already been discussed at length. Unlike the effect of GUUGLE's competition with other initiatives, the effect of network proximity on participation choices has been positive. From this effect, something can be learned for the enhancement of GUUGLE's fitness for purpose.

9.5 Enhancing GUUGLE's prospective fitness for purpose

The research question of this study asks not only for GUUGLE's fitness for purpose but also for possibilities to enhance it. However, diagnosis and treatment are two different things. The identified problems of the GUUGLE programme's implementation do not imply any particular countermeasures but rather suggest directions in which to look for them. This is why the second part of the research question has been phrased so carefully ('... and how might its fitness be enhanced?') and why the following sections will formulate no solutions but only hypotheses for solutions to be tested in practice.

9.5.1 Providing extrinsic incentives

Professors who consider CETL irrelevant lack a motivating reason to think differently about it and engage in it. As the first evaluation cycle has shown, this only applies to a fraction of the professors. Interviewee C's view that 'as long as there are no serious [extrinsic] incentives for better teaching [...] GUUGLE won't be sustainable. No

incentives equal no reasons' (section 8.3.2) is probably an exaggeration, but her insistence on extrinsic incentives has merit. There seems no other way to convince professors like respondent A ('if we spend time talking with our colleagues [...] it must really be worth it and have a useful, very concrete output'; *ibid*) to join GUUGLE or to engage in CETL more generally.

One possibility for providing extrinsic incentives for CETL consists in encouraging collaborative teaching (e.g. team teaching, rotational teaching, teaching courses linked by a common theme) or other forms of joint responsibility for student learning, giving professors not only reasons for conversations about their teaching but also shared experiences to refer to and to converse about. As Palmer (1993) observes, when academic teachers emerge from the classroom, 'we rarely talk with each other about what we have done, or need to do. After all, what would we talk about?' (8). The endorsement of collaborative teaching by senior management and the provision of additional resources for it would have the further advantages of gradually freeing the attendance of colleagues' lectures from taboo and sending a strong message of change through the *Hochschule Bremerhaven*.

A second extrinsic incentive for CETL could be provided through the inclusion of CETL-related objectives in individual performance agreements, which a growing number of professors sign every four years with the Rector. The fulfilment of the performance agreement is associated with a pay rise.¹⁵ Suitable objectives may be the organisation and facilitation of a certain number of PLCs, the use of collaborative

¹⁵ This refers to the professorial W pay scale which was introduced in 2003 and includes a performance-dependent component.

teaching in two modules with two different colleagues or the teaming up with colleagues to sit in on each other's classes, to give and receive feedback and to document the process. However, each of these objectives has a downside. Only professors with positive prior experience would agree to hosting PLCs, so no new participants can be won this way; and not all professors are equally fit to run a PLC. The feasibility of collaborative teaching in a concrete situation depends not only on one professor's willingness to do it but also on the availability of suitable modules and colleagues and their respective willingness to make this effort. Similar constraints apply to the peer review of teaching.

Also, the conditional allocation of budgets to departments, institutes, laboratories and individual professors could act as extrinsic incentives for certain types of behaviour. However, material incentives entail problems which may outweigh their benefit. Professors may get conditioned to expect material incentives also in other areas of their work, and the costs of preventing and detecting fraudulent behaviour may turn out excessive. Therefore, other mechanisms to encourage CETL seem preferable.

9.5.2 Networking and leveraging relationships

The GUUGLE programme's aim to stimulate CETL, 'de-privatise' teaching and change the organisational culture at the *Hochschule Bremerhaven* is essentially about transforming the relationships and interactions among the professors. The first evaluation cycle gave an idea of the causal power inherent in these relationships when it revealed the highly significant correlation between the professors PLC participation choices and their network proximity to the GUUGLE team. This power has been

underestimated by the GUUGLE team in the past and could be harnessed more fully through purposive internal networking and relationship building.

Networks and relationships have played an important role in the GUUGLE team's work from the beginning, but attempts to develop and leverage them were mostly unplanned and opportunity-driven. An exception was the systematic approach taken to target the participants of the pilot PLC in 2009, which was based on the social network map in Figure 4.1. But also then, the social network was taken as given, utilised only for a single occasion and not developed further. However, when networking is understood as 'creating a fabric of personal contacts who will provide support, feedback, insight, resources, and information' (Ibarra & Hunter 2007, 40), the emphasis shifts from exploiting an existing network in order to recruit GUUGLE participants to actively expanding and fostering a network which may serve a range of purposes at different points in time. Considering further that 'innovation today has become significantly more of a *networking process*' (Rothwell 1994, 43, original emphasis), the potential for rendering GUUGLE more effective, adaptive and innovative through purposive networking and relationship building seems vast.

For instance, the GUUGLE team could increase the programme's reach and impact by initiating or deepening personal relationships with professors who occupy more 'remote' positions in the GUUGLE team's social network. These relationships, or weak ties in social network terms (see section 4.1.2), should be given enough time to unfold, be based on reciprocity and not be constrained to GUUGLE-related topics. Such relationships may help overcome indifference or hostility towards GUUGLE (stage 1 of the self-selection process in Figure 9.2) or lower individual barriers to engaging in

GUUGLE events (stage 2). Moreover, they may prove invaluable as feedback channels through which the needs, wishes, expectations, ideas and concerns of professors may be elicited, who would otherwise not speak out. If the GUUGLE programme is about transforming collegial relationships, the GUUGLE team ought to lead by example since ‘the medium is the message’ (McLuhan 1964, 7. See section 2.4.1.).

9.5.3 Profiling and managing PLCs

Stage 3 of the self-selection process poses a more operational kind of challenge. In the past, PLC participants discontinued their engagement as a result of their dissatisfaction with the impact of the PLC. The problem seems to be a combination of unclear objectives, excessive expectations, PLC facilitation and the given circumstances. As the content analysis in section 7.4 has shown, 60% of the PLCs were focusing on reflection and learning, whilst 40% were stressing problem-solving. However, a focus on solving concrete problems is neither implied by the working definition of a PLC (see section 4.2.1) nor was it intended by the GUUGLE programme initiators. Consequently, the participants’ expectations could not be managed adequately beforehand; the facilitators tended to be unprepared for the direction which their PLCs were taking; and, with only one exception, no contact with committees had been made which could have helped to implement solutions proposed by a PLC.

Based on the insights gained through the realist evaluation, three modifications could be made to the way PLCs are managed in order to prevent dissatisfaction, improve word of mouth and increase repeat participation rates. First of all, to ensure impact receives sufficient attention, PLCs may need to become more results-oriented than in the past. Secondly, the main intended impact of a PLC, either learning and reflection or problem

solving, should be communicated in advance and always be kept in mind by the participants. Sharpening the profile of PLCs will help promote them and set expectations right. The third modification suggested here follows directly from the first two and concerns the facilitation of PLCs. Results orientation requires the facilitators to make sure that early on, end products are defined and agreed by the participants; that every session contributes to them; that meandering conversations are kept to a minimum; and that progress towards the agreed end products is assessed on a regular basis. The PLC profiles (learning and reflection versus problem-solving) also need to be clearly recognisable in the facilitation approach. In both cases, taking action and driving implementation rather than just conversing may be necessary to convince professors of PLCs' impact potential.

9.5.4 Experimenting with new intervention formats

For GUUGLE, stimulating CETL to overcome pedagogical solitude is not only a matter of ensuring that there is always something to talk about with colleagues; it is equally a matter of providing the necessary resources. These include occasions for CETL such as the PLCs and the GUUGLE Forum conference series. But since 2009 these occasions have basically remained the same. Apart from variations in the way the GUUGLE Forum conferences were organised, there has been little innovation in the programme's intervention formats.

Since novelty tends to attract attention, spur curiosity, revive interest and prevent routine from taking over, it may be a good idea to experiment with new types of OD interventions. Section 2.2.2 lists a number of dialogic OD approaches, some of which might fit very well to the existing programme. For example, an Open Space conference

or a World Café could be used to generate new topics for PLCs and other events, giving professors new motivating reasons to join GUUGLE and engage in CETL. And learning-centred PLCs could explicitly adopt dialogic action research as their core method for deeper learning and greater impact.

9.6 Transferring GUUGLE to other contexts

This brief section offers some thoughts about the GUUGLE programme's transferability to other HE contexts. These thoughts are not purely hypothetical since three partner universities in Poland, Spain and Uzbekistan have invited the GUUGLE team for presentations and workshops and are considering adopting the programme in parts or fully. Obviously, countless contextual aspects affect the success of OD programme implementations, including a committed team, funding and senior management support. This section, however, will limit itself to three aspects about which this study permits drawing some tentative conclusions: size of institution, type of institution and low/high context culture.

The close association between PLC participation choices and network proximity to the GUUGLE team suggests that informal social networks and communication channels are important to bring together enough participants in early phases of the programme to be able to announce 'quick wins', build momentum, establish a reputation and capture the attention of further potential participants. Snowball recruitment of PLC participants seems to have worked well in Bremerhaven and, in absence of extrinsic incentives to participate, may be the single most effective way to get the programme started. Smaller HE institutions like the *Hochschule Bremerhaven* may have an advantage over larger ones, because their academic staff's informal networks tend to reach further across the

institution and bridge structural holes between faculties and departments simply because the chances of sitting in the same meetings, attending the same events or bumping into each other on campus are greater when the campus and the number of staff are small. To transfer GUUGLE to a large HE institution, it may be necessary to pursue a ‘multi-seed’ strategy by treating institutional subunits as separate organisations, sowing departmental or faculty PLCs and then leveraging the different local social networks. Alternatively, GUUGLE can first be implemented in one subunit of a large HE institution, allowing the team to gain experience, build its own capacity and make mistakes only on a small scale, before scaling up the programme.

The second contextual aspect to be considered here is the type of institution. The *Hochschule Bremerhaven* is a *Fachhochschule* and as such, in various respects, similar to post-1992 universities in the UK. Like the new universities, the *Fachhochschulen* have a fairly young history, far less tradition to cling to, a greater professional focus and closer ties with the business world. Unlike the new universities, however, all professors at the *Fachhochschulen* have worked outside the HE system (see section 4.1.1) and received a rather untypical academic socialisation. Professorial habitus is far less common at the *Fachhochschulen* than it is at the traditional universities (Stegmann 2007). The accounts of the participants of the 2012 Dortmund Spring School (see the outset of chapter 1), none of whom was from a *Fachhochschule*, exemplify the level of resistance and marginalisation to be expected by those who attempt to establish a programme like GUUGLE at an old German university. With its informal nature, colourful appearance, emphasis on teaching and its behavioural and cultural change agenda, GUUGLE would not be applauded by the professorial establishment. GUUGLE at a traditional university would probably lose all senior professors and also a

significant part of their more junior colleagues as potential participants already in the first stage of the self-selection process in Figure 9.2. Therefore, it would probably make more sense to target primarily doctoral students and post-docs in order to influence their academic socialisation.

The third aspect to be taken into account when transferring GUUGLE to another context is national culture. One dimension on which national cultures have been ranked is Hall's (1976) low/high context continuum. The term high/low context characterises the cultural rules around information exchange and, in particular,

'the degree to which information in a culture is explicit, vested in words or precise and unambiguous meaning (low context) and the degree to which it is implicit, vested in shared experience and assumptions and conveyed through verbal and non-verbal codes (high context). [...] High-context communication requires far more time because trust, friends and family relationships, personal needs and situations will also be considered.' (Korac-Kakabadse et al. 2001, 6-7)

Higher-context cultures have a strong sense of tradition and history and remain relatively stable over time, rendering indirect communication through cues and references to elements of the shared cultural heritage possible. Of immediate relevance to pedagogical solitude and GUUGLE is Hall's (1976) observation that individuals in high context cultures are more likely to ask for assistance than those in lower-context cultures, who expect self-service support instead and tend to look for solutions independently. Support in high-context-cultures is provided by close-knit groups to their members, who are expected to reciprocate.

Hall proposed that all cultures can be ranked by their communication and information processing preferences. In this ranking, which is depicted in Figure 9.3, German features as one of the lowest-context cultures worldwide.

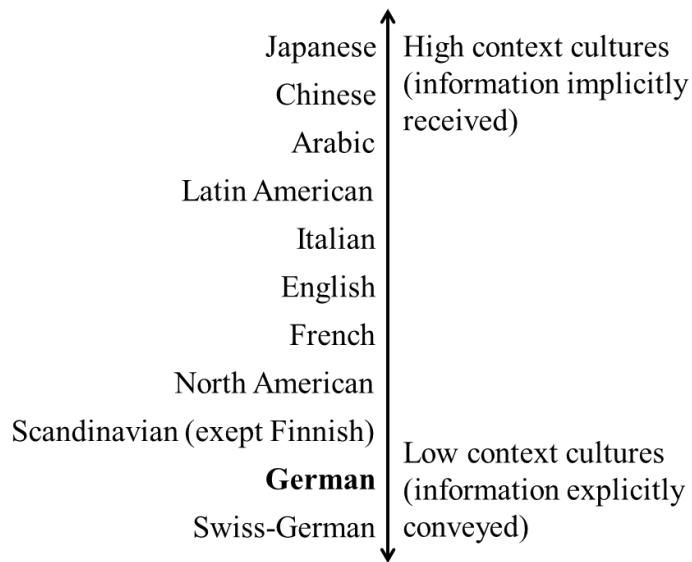


Figure 9.3. The high/low context continuum (adapted from Hall 1976, 91).

It might be more than just a coincidence that Germany, where GUUGLE is at home, and the United States (except the southernmost States), where most of the literature on pedagogical solitude has been published, are characterised by low-context cultures. Hall's findings suggest that pedagogical solitude may be less of an issue at HE institutions in higher-context cultural settings. Accordingly, implementing GUUGLE at an English or Italian university can be expected to be less difficult than in Germany, but also less necessary.

9.7 Summary

This chapter addressed the research question of this thesis: *How fit is GUUGLE for the purpose of overcoming pedagogical solitude and how might its fitness be enhanced?* In a first step, it operationalised the concept of fitness for purpose by decomposing it into theoretical, practical and prospective fitness for purpose and defining the criteria for each. Then it applied the criteria to GUUGLE, taking the results of the three realist

evaluation cycles into account. GUUGLE turns out to be fit for purpose in theory, but that its practical fitness is limited to professors who are intrinsically motivated to participate, as GUUGLE fails to provide new motivating reasons for CETL. The prospective fitness for purpose of the GUUGLE programme is given only for a fraction of the professoriate, since indifference, personal barriers to participation and dissatisfaction with the impact induce professors not to engage in CETL or to withdraw from it. Possible enhancements of the GUUGLE programme include providing extrinsic incentives to participate, networking and relationship building by the GUUGLE team, sharpening the profile and improving the management of PLCs as well as innovating with new dialogic OD interventions. Finally, if GUUGLE is to be transferred to other places, smaller and less traditional HE institutions in higher-context cultures are likely to simplify a successful implementation.

Chapter 10 Concluding Remarks

This thesis concludes with a recapitulation of the main findings and implications of the three evaluation cycles of chapters 6-8; a summary of its main contributions; and a reflection on some of the merits and shortcomings of Realistic Evaluation.

10.1 Evaluation findings and implications recapitulated

In the course of the three sequential evaluation cycles, the elements of the normative GUUGLE programme theory were empirically tested by means of different data sets and methods, and gradually refined so as to (better) explain the observable effects of the GUUGLE programme. The first cycle confirmed that professors at the *Hochschule Bremerhaven* regarded CETL as an important idea and practice, were engaging in GUUGLE-facilitated CETL and planning to continue doing so. However, the microsocial mechanism structure, posited by the normative programme theory and depicted by the whirlpool diagram, was found not to fit the data sufficiently well and had to be revised. Part of the microsocial structure is feedback loop 1, according to which professors engage in CETL because they find it an important practice, a view continuously reinforced by their own engagement in CETL. This crucial circular mechanism was found to be weaker and less reliable than expected. The unanticipated, significantly positive effect of professors' social network proximity to the GUUGLE team on their engagement in CETL was added to the model as a new mechanism.

The second evaluation cycle confirmed the refined model of the first cycle, with the exception of feedback loop 2, i.e. the cultural anchoring of CETL, for which insufficient evidence was found. Data analysis also suggested that the whirlpool model was still

incomplete. Firstly, professors saw CETL as an important practice not only for the opportunities it offers to learn from and with colleagues, as claimed by the normative programme theory, but also for the collective problem-solving it enables. Secondly, while the reasons for CETL, which GUUGLE had been communicating, appealed to many professors, they failed to do so to others and even triggered defensive reactions. And thirdly, some professors experienced CETL in PLCs as ineffective and refrained from joining further PLCs. All three effects were incorporated into the whirlpool model, either by describing mechanisms more precisely or by adding new ones.

The third evaluation cycle highlighted in particular two problems of the GUUGLE programme: its failure to provide motivating reasons for CETL; and its inability to convert the strong support for CETL as an important *idea* (see Table 6.5) into more CETL as a *practice*. Consequently, the whirlpool model was modified by replacing the claim that GUUGLE provides reasons for CETL by the more modest descriptive statement that GUUGLE raises issues in teaching and learning (which may constitute motivating reasons for some but not for others); and by introducing cognitive dissonance as a psychological mechanism explaining the gap between some professors' *belief* that CETL is an important idea, and their *practice* of refraining from CETL. Last but not least, feedback loop 2 again did not receive empirical support and remains the only purely normative component of the otherwise empirically founded, explanatory programme theory.

All three evaluation cycles pinpointed problems at the level of the social mechanisms on which GUUGLE's effectiveness and sustainability depend, explaining the programme's failure to work for certain groups of professors and to become self-sustaining.

10.2 Contributions of this thesis

The contributions of this thesis fall into two categories: practical and methodological. The practical contributions result from empirically evaluating GUUGLE. They benefit the programme's management and further development at the *Hochschule Bremerhaven* as well as HE institutions elsewhere intending to implement alike programmes.

In the spirit of Kurt Lewin's (1951) dictum that there is 'nothing so practical as a good theory' (169), the first practical contribution is the GUUGLE programme theory. The advantage of having it on paper and supported by empirical evidence is a great advantage over the previous situation, where each stakeholder had her or his own tacit mental model of the programme, with no guarantee that the models would resemble each other and guide action in the same direction. Moreover, in the face of the apparent pervasiveness of pedagogical solitude in German HE institutions, the programme theory may be of interest to a wider audience than just the GUUGLE team and professors at the *Hochschule Bremerhaven*.

The second contribution of practical relevance lies in the identification of flaws in the programme implementation, the specification of their causes and the development of suggestions for tackling them. Operational programme management hardly leaves time for detailed problem analyses, so chances are that these problems would otherwise have gone unnoticed until too late.

A third practical contribution reveals itself only at second glance. It consists in the framing of GUUGLE as a driver of organisational learning and in the clarification of its importance for the *Hochschule Bremerhaven*'s adaptive coping. While this may not sound very practical, it can help internal communication shift the professors'

perspective on the programme. Those who refuse to join GUUGLE because they see it as personally irrelevant or a threat to their self-concept may soften their attitude when they understand that GUUGLE is not ‘just’ about professional development but also about transforming the *Hochschule* to make it a more sociable place, more responsive and innovative, more competent in dealing with external challenges and ultimately more successful.

The *methodological* contributions of this thesis are innovations potentially benefitting academic and professional programme evaluators and organisation developers. For them, the modification of Realistic Evaluation and the critique of Pawson & Tilley’s approach are probably the most valuable bits of the thesis. I will extend the critique in the next section. The modification involves in particular the following three aspects.

Firstly, Coleman’s macro-micro-macro ‘bathtub’ scheme is proposed as an alternative to Pawson & Tilley’s ‘rugby ball’ diagram for representing context, mechanisms and outcomes of social interventions. This modification permits a more detailed modelling and analysis of programme mechanisms (i.e. it opens the black box even further), puts greater emphasis on programme subjects’ perceptions, decisions and actions and provides a more intuitive visual support for realist interviews.

Secondly, the ‘whirlpool’ scheme is introduced as an upgrade of Coleman’s bathtub. Feedback loops, which can be customised according to need, represent the homeostatic control mechanisms required to anchor any new behavioural, attitudinal or other patterns in an organisation. The circular mechanisms force whirlpool scheme users to consider the sustainability of intended organisational changes.

Thirdly, a more transparent and differentiated treatment of context is achieved by decomposing the programme subjects' context into programme context, which remains unaffected by a programme, and delta context, which comprises the changes introduced by the programme with the aim of influencing the subjects' decision-making and behaviour. The concept of delta context also helps reconcile the bathtub/whirlpool scheme with Pawson & Tilley's context-mechanism-outcome configuration or CMOC.

Last but not least, the sequential mixed-methods research design of this thesis may not be a methodological milestone, but its particular purpose might be worth stressing here. Given my triple position of GUUGLE programme initiator/director, programme evaluator/researcher and colleague of the researched, I needed to minimise the risk of my colleagues' strategic behaviour and my own biases polluting the data. So I developed a research design which largely builds on existing data that neither my colleagues nor I would be able to influence. This design approach to securing the validity of insider research might be considered a fourth methodological contribution.

10.3 Critical reflections on Realistic Evaluation

I chose Realistic Evaluation for this thesis because it promised to: avoid the black box problem by revealing a programme's inner workings; generate theory and thus contribute to knowledge; focus on causation and explanation, rather than just on outcomes; allow me to apply and mix research methods I wished to familiarise myself with (see my Personal Statement at the outset of the thesis); and because it is based on a philosophical paradigm I was curious about.

The last point could well have been the first on the list. Lectures by Roy Bhaskar at the Institute of Education had left me impressed and excited but also puzzled. I found it hard to imagine how critical realism would translate into concrete practice. Realistic Evaluation became the example I was looking for, and I tend to consent to Rogers's (1999) judgement that 'This is one of those rare books that has the potential to permanently change one's perspective on program evaluation' (381).

Realistic Evaluation held its promises, and due to the emphasis on explanation, theory generation, empirical testing and the use of standard social research methods, doing Realistic Evaluation felt very much like 'ordinary' research, which reassured me of its suitability for a doctoral thesis. In fact, in my opinion its degree of methodological sophistication makes Realistic Evaluation rather unsuitable for evaluators without substantial research training. This may explain why, in his sample of 52 published cases of Realistic Evaluation, Astbury (2011) finds that the approach was employed mostly by academic researchers who do not affiliate primarily with the field of evaluation.

It is not only methodological challenges that limit the usability of Realistic Evaluation for evaluation practitioners. I agree with Gill & Turbin (1999) that

'The requirements of data collection are far more specific using this methodology; notably, that each of the elements of context, mechanism and outcome require careful validation if they are to be proven. It is relatively easy to propose plausible CMO configurations but much harder to collect useful (or valid) data for all three, particularly where time and resources are limited.' (195)

I also have to agree with Patton (1999) who warns that the results of Realistic Evaluation may not get utilised in practice since stakeholders are de facto excluded from playing an active role due to the level of abstraction and lack of user-friendliness of the approach. With a touch of irony, Patton remarks that 'It's certainly refreshing to

see stakeholders referred to so honestly in *Realistic Evaluation* as the evaluation researcher's "subjects" (ibid, 387, original emphasis).

Despite my generally positive experiences with the approach, I have some further reservations about Realistic Evaluation. First of all, its basic tenet is that 'programmes are theories' (Pawson 2003, 472), meaning that programmes embody theories which, therefore, are *normative*. At the same time, Realistic Evaluation demands that programme theories be refined until they explain what the respective programmes actually do, i.e. programme theories are *explanatory*. This contradiction results from Pawson & Tilley's undifferentiated use of 'theory' and the failure to clarify the relationship between normative and explanatory programme theories.

My second reservation weighs heavier. It concerns gaps between normative programme theory and programme implementation, which affect how a programme works for whom and in what circumstances. As Figure 10.1 illustrates, evaluators may be led to wrongly reject or accept hypotheses on the basis of observations that reflect not the normative programme theory but its deficient implementation. For instance, its normative theory requires GUUGLE to provide reasons for CETL, which, as my research has shown, the actual programme has been failing to do. Clearly, this finding does not falsify the normative programme theory, it only points to an implementation problem. Realistic Evaluation, however, implicitly assumes that normative theories are put into practice accurately, basically forcing evaluators to adjust the explanatory programme theory to fit data resulting from faulty implementation.

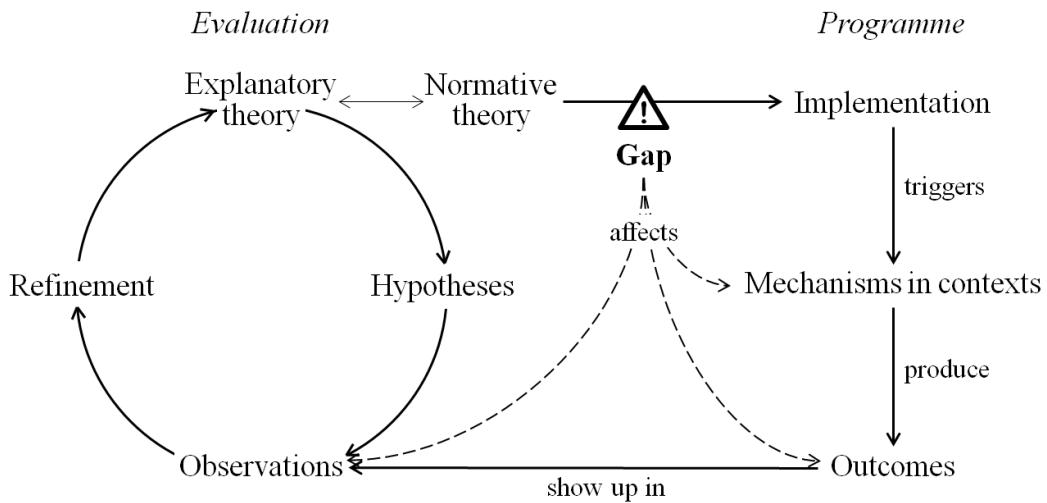


Figure 10.1. Influence of a theory-implementation gap on evaluation.

To avoid this problem, Realistic Evaluation could follow the example of the Theories of Change approach which combines programme theory with implementation theory (Weiss 1995). Programme theory specifies hypothesised causal links between mechanisms triggered by an intervention and their anticipated outcomes in a particular context, whereas implementation theory prescribes for that context what is required to trigger a certain mechanism. This way, implementation issues could more easily be identified and taken into account in the testing of programme theory.

My third reservation about Realistic Evaluation refers to a point which I already addressed in one of my earlier essays: the ‘many variables, few cases’ problem of the comparative method (Lijphart 1971), which is also closely related to the underdetermination thesis mentioned in section 5.5. The problem of not having enough observations available to sort out rival explanations for a phenomenon extends to retrodiction in open systems and therefore also to Realistic Evaluation. Pawson offers his own Realist Synthesis (Pawson 2002; Pawson et al. 2005) to cope with this problem:

increasing the number of observations by making many evaluations cumulate. This approach is promising in areas where large numbers of similar interventions take place in similar contexts, but it is ineffective when programme contexts are fairly unique or few Realistic Evaluations are conducted.

So it is no truism but a network effect that realising the potential of Realistic Evaluation requires more Realistic Evaluation. The potential, I believe, also merits further research and development to tackle the shortcomings of this innovative and, in a way, radical approach to programme evaluation.

Appendix

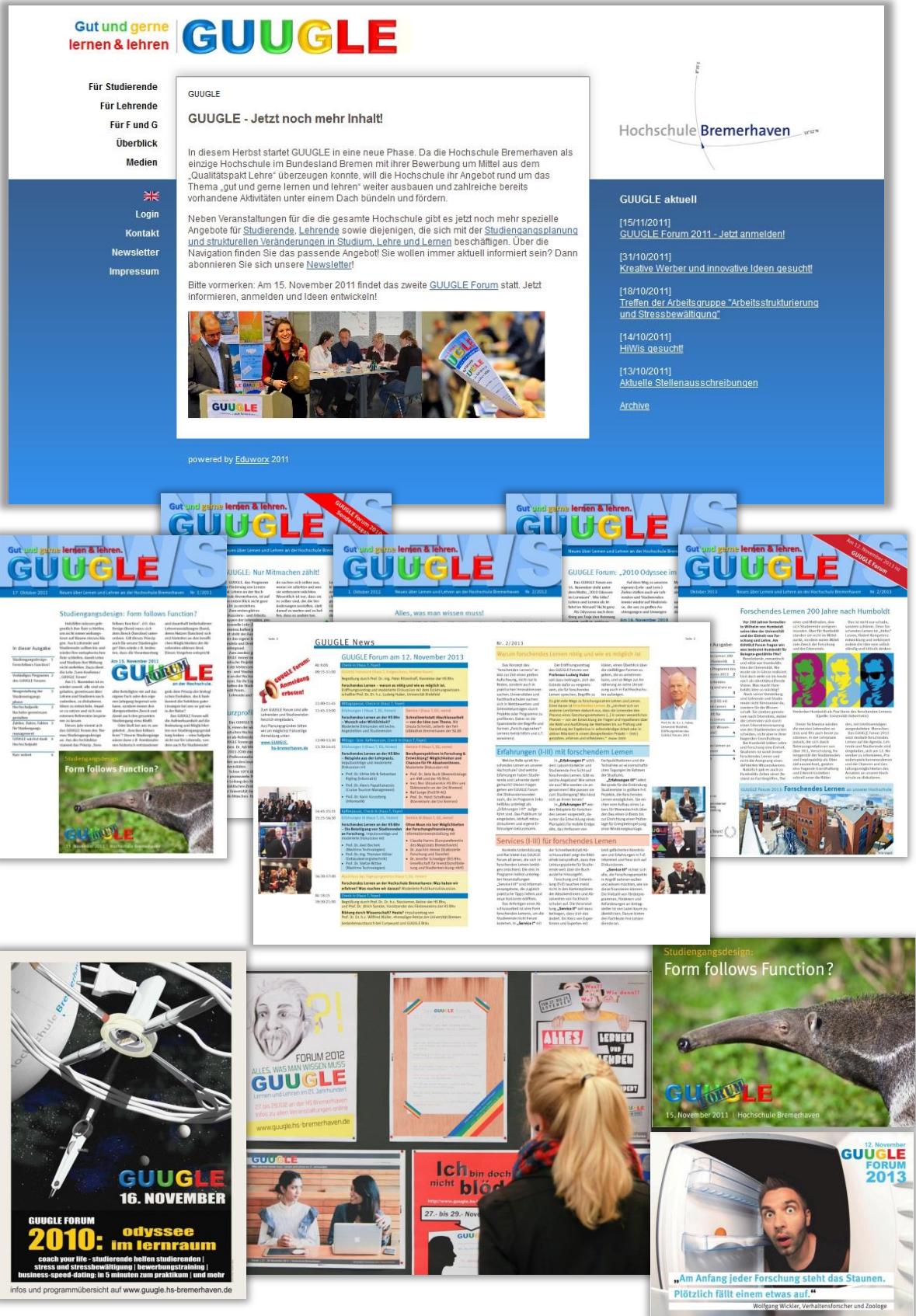
A1 Images



Figures A1 a-d. Impressions from GUUGLE PLC sessions and workshops.



Figures A2 a-d. Scenes from GUUGLE Forums.



Figures A3 a-c. GUUGLE website, newsletter, posters and postcards.

A2 Tables

Table A1. Path model fit summary.

Measure of model fit	Good fit	Source	Model 1	Model 2
Relative chi-square	$\chi^2/df < 2$	Ullman (2014)	5.342	1.186
Normed Fit Index	NFI > .95	Ullman (2014)	.767	.965
Bollen's Incremental Fit Index	IFI > .95	Hu & Bentler (1999)	.802	.994
Tucker Lewis Index	TLI > .95	Hu & Bentler (1999)	.584	.982
Comparative Fit Index	CFI > .95	Ullman (2014)	.792	.994
Standardised root mean squared residual	SRMR < 0.08	Hu & Bentler (1999)	.165	.046
Root Mean Square Error of Approximation	RMSEA < .07	Steiger (2007)	.409	.085
P of Close Fit	PCLOSE > .05	Byrne (2009)	.002	.329

Table A2. Correlations between PastGP and possible explanatory variables.

		PastGP
Self-concept (1=teacher; 5=expert)	Spearman's ρ p-value (2-tailed) N	-.053 .702 54
Responsibility (1=for learning; 5=for teaching)	Spearman's ρ p-value (2-tailed) N	.111 .427 53
Perception of students (1=as adults; 5=as kids)	Spearman's ρ p-value (2-tailed) N	-.110 .426 55
Energy impact of teaching (1=takes energy; 5=gives energy)	Spearman's ρ p-value (2-tailed) N	-.148 .296 52
Perception of student feedback (1=not important; 5=very important)	Spearman's ρ p-value (2-tailed) N	.038 .785 55
Typical size of classes taught (1=large, i.e. >40 students; 5=small)	Spearman's ρ p-value (2-tailed) N	.059 .670 55
Time available to cover course contents (1=not enough; 5=enough)	Spearman's ρ p-value (2-tailed) N	-.103 .456 55
Typical results of teaching evaluation by students (1=not so good; 5=very good)	Spearman's ρ p-value (2-tailed) N	-.132 .355 51
Discussing own teaching evaluations with colleagues (1=never; 5=always)	Spearman's ρ p-value (2-tailed) N	.017 .906 51

A3 Interview notes

Realist interview with respondent A.

The following pages are a translated and slightly polished exemplary excerpt from the notes I took in German during the realist interview with respondent A. The introduction to the GUUGLE programme theory (the interview's teacher-learner function) has just ended, and the interview is entering into its concept refinement phase, where A is asked '*to confirm, to falsify and, above all, to refine that theory*' (Pawson & Tilley 1997, 155, original emphasis).

General plausibility of the programme theory

- Interviewee A much prefers the diagrams to the 'usually wordy GUUGLE presentations and newsletters'.
- Finds the theory generally plausible, gets the point of what GUUGLE is trying to achieve and how.
- As an engineer, A is concerned about the fuzzy meaning of the arrows. They seem under-defined. E.g. how does the transmission between social and individual conditions work? Through conversation, observation, other forms of perception? Over what period? How often? With what intensity? Etc. [Me: *Arrows summarise all forms of influence. The transmission mode is not so important here.*]
- Another concern is about the feedback loops: is it reinforcing or dampening feedback? Homeostatic control needs both and in balance. [Me: *The loops*

represent reinforcing feedback since dampening forces are abundant at our Hochschule. And there is anyway a natural ceiling to the intensity of CETL].

- The model looks plausible but other models of GUUGLE may also be plausible.

So why this one? [Me: This model is based on the initial idea of how GUUGLE might work. Need to start somewhere. Aim is to find a plausible alternative that fits even better.]

Feedback on individual DMOCs

#	Delta context	+	Mechanism	=	Outcome
1	Communication of reasons for CETL		Reasons trigger learning processes in professors.	=	(a) Professors regard CETL as an important idea.

- Doesn't feel that GUUGLE communicates reasons for CETL. GUUGLE brings up topics and invites to events, but provides no reasons.
- Sees no compelling reason why CETL is important and why it should matter to her teaching what others think or say. Different subjects are taught differently by different people. CETL may be nice but means extra time pressure because less time is left for other things. Even if A learns from others, best way for her to improve teaching is to have time to prepare.
- Can't believe that many colleagues have changed their mind on CETL, even if they participate in PLCs. Imagines that GUUGLE meets pre-existing demand for CETL, which is a good thing.
- Rejects DMOC 1.

#	Delta context	+	Mechanism	=	Outcome
1a	Communication of reasons for CETL		Communicated reasons are ignored or considered a threat by professors.		Professors regard CETL as irrelevant or react with defensive routines.
<hr/>					
<ul style="list-style-type: none"> • Dislikes representation of non-participants by leakage symbol because pejorative connotation. Associations: waste, unproductive, annoying dripping tap. Non-participation is a legitimate choice. • Believes that GUUGLE doesn't threaten any professor. Threat implies importance, but GUUGLE is too unimportant for most. Among her engineering colleagues, it's seen as a playground for the FB2 [= economic faculty]. • Also believes that professors don't ignore reasons provided by GUUGLE. Reasons are personal, not universal. What GUUGLE communicates may be reasons for some but not for others. For A, GUUGLE communicates no reasons. • Reasons for CETL must be compelling and promise real benefits to her personally. Doesn't care much about benefits to the <i>Hochschule</i>. • Time spent talking with colleagues must be worth it, e.g. have useful, very concrete output. Can't see that PLCs have done this. • Doesn't think that defensive reactions by some colleagues result from perceived threat by GUUGLE. Rather, it's frustration that something as secondary as talk about teaching receives so much attention and money. • Rejects DMOC 1a. 					

#	Delta context	+	Mechanism	=	Outcome
2	Outcome (a)	+	Professors form expectations about how CETL may benefit them in practice.	=	(c) Professors regard CETL as a (potentially) important practice.

- Agrees with the mechanism because it describes her own reaction, but disagrees with outcome.
- Apologises for her sceptical attitude.
- Argues that adhering to a general idea does not imply that people act on it. Value conflicts, constraints of daily life, different long/short-term priorities etc. interfere with ideas. People are also inconsistent.
- Example: most people like the idea of harmony, but their behaviour is inconsiderate, self-centred and aggressive.
- *[Me: Outcome of DMOC 2 is not about action, only about linking the important idea of CETL with one's own situation. Triggering action is the concern of DMOC 3.]*
- Insists that the model confuses the abstract [*CETL as an important idea*] and the concrete [*CETL as an important practice*]. The mechanism only seemingly explains the outcome.
- Rejects DMOC 2.

[Excerpt covers about 20 minutes of the 60-minute interview.]

References

Alvesson, M. (2003). Methodology for close up studies - struggling with closeness and closure. *Higher Education*, 46(2), 167-193.

Alvesson, M. (2013). *Understanding Organizational Culture*. London: Sage.

Anderson, G. L., & Herr, K. (1999). The New Paradigm Wars: Is There Room for Rigorous Practitioner Knowledge in Schools and Universities? *Educational Researcher*, 28(5), 12-21+40.

Archer, M. (1998). Introduction: realism in the social sciences. In M. Archer, R. Bhaskar, A. Collier, T. Lawson & A. Norrie (Eds.), *Critical Realism: Essential Readings* (pp. 189-205). London: Routledge.

Argyris, C., & Schön, D. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading/MA: Addison Wesley.

Ashby, E. (1967). Ivory Towers in Tomorrow's World. *The Journal of Higher Education*, 38(8), 417-427.

Astbury, B. (2011). The practice of realistic evaluation: a conceptual and empirical review. Unpublished doctoral research study. Master, The University of Melbourne, Melbourne.

Astbury, B., & Leeuw, F. L. (2010). Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. *American Journal of Evaluation*, 31(3), 363-381.

Barnett, R. (2000). *Realizing the University in an Age of Supercomplexity*. Buckingham: SRHE and Open University Press.

Barnett, R. (1997). *Higher Education: A Critical Business*. Buckingham: SRHE and Open University Press.

Barr, R. B., & Tagg, J. (1995). From Teaching to Learning - A New Paradigm for Undergraduate Education. *Change*, 27(6), 12-25.

Becher, T., & Trowler, P. R. (2001). *Academic Tribes and Territories*. (2nd ed.). Buckingham: The Society for Research into Higher Education & Open University Press.

Beckhard, R. (2006). What Is Organization Development? In J. V. Gallos (Ed.), *Organization Development. A Jossey-Bass Reader* (pp. 3-12). San Francisco, CA: Jossey-Bass.

Bemelmans-Videc, M., Rist, R. C., & Vedung, E. (Eds.). (2010). *Carrots, Sticks, and Sermons: Policy Instruments and Their Evaluation*. New Brunswick, NJ: Transaction Publishers.

Bhaskar, R. (1986). *Scientific Realism and Human Emancipation*. London: Verso.

Bhaskar, R. (1989). *Reclaiming Reality: A Critical Introduction to Contemporary Philosophy*. London: Verso.

Bhaskar, R. (1998). *The Possibility of Naturalism* (3rd ed.). London: Routledge.

Bhaskar, R. (2008). *A Realist Theory of Science*. Oxon: Routledge.

Bhaskar, R., & Lawson, T. (1998). Introduction: Basic Texts and Developments. In M. Archer, R. Bhaskar, A. Collier, T. Lawson & A. Norrie (Eds.), *Critical Realism: Essential Readings* (pp. 3-15). London: Routledge.

Blamey, A., & Mackenzie, M. (2007). Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and Oranges? *Evaluation, 13*(4), 439-455.

Bohm, D. (1996). *On Dialogue*. London: Routledge.

Brookfield, S. D. (2005). *The Power of Critical Theory for Adult Learning and Teaching*. Maidenhead: Open University Press.

Brown, J., & Issacs, D. (2005). *The World café: Shaping our futures through conversations that matter*. San Francisco, CA: Berrett-Koehler.

Buchanan, D., Claydon, T., & Doyle, M. (1999). Organisation Development and Change: the Legacy of the Nineties. *Human Resource Management Journal, 9*(2), 20-37.

Bullock, R. J., & Batten, D. (1985). It's Just a Phase We're Going Through: A Review and Synthesis of OD Phase Analysis. *Group & Organization Management, 10*(4), 383-412.

Burnes, B., Wend, P., & By, R. T. (2013). The changing face of English universities: reinventing collegiality for the twenty-first century. *Studies in Higher Education, ,* 1-22.

Büro für Zukunftsfragen (Ed.). (2013). *Art of Hosting. Handbuch über die Kunst, Räume für gute Gespräche zu schaffen* (2nd ed.). Bregenz: Amt der Vorarlberger Landesregierung.

Bushe, G. R. (2013). Dialogic OD. A Theory of Practice. *OD Practitioner*, 45(1), 11-17.

Bushe, G. R., & Marshak, R. J. (2009). Revisioning Organization Development. *The Journal of Applied Behavioral Science*, 45(3), 348-368.

Byrne, B. M. (2009). *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming* (2nd ed.). New York: Routledge.

Carspecken, F. P. (2004). *Critical ethnography in educational research: A theoretical and practical guide*. New York: Routledge.

Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing Coping Strategies: A Theoretically Based Approach. *Journal of Personality and Social Psychology*, 56(2), 267-283.

Chiva-Gómez, R. (2003). The facilitating factors for organizational learning: bringing ideas from complex adaptive systems. *Knowledge and Process Management*, 10(2), 99-114.

Cilliers, P. (1998). *Complexity & Postmodernism: Understanding Complex Systems*. London: Routledge.

Clark, B. R. (1983). *The Higher Education System: Academic Organization in Cross-National Perspective*. Berkley/CA: University of California Press.

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A Garbage Can Model of Organizational Choice. *Administrative Science Quarterly*, 17(1), 1-25.

Cohen, S., Kimelfeld, B., & Koutrika, G. (2012). A Survey on Proximity Measures for Social Networks. In S. Ceri, & M. Brambilla (Eds.), *Search Computing III: Broadening Web Search* (pp. 191-206). Heidelberg: Springer.

Coleman, J. S. (1964). *Introduction to Mathematical Sociology*. New York: The Free Press.

Coleman, J. S. (1986). Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology*, 91(6), 1309-1335.

Coleman, J. S. (1987). Microfoundations of Macrosocial Behavior. In J. C. Alexander, B. Giesen, R. Münch & N. Smelser (Eds.), *The Micro-Macro Link* (pp. 153-173). Berkeley, CA: University of California Press.

Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard University Press.

Collier, A. (1994). *An Introduction to Roy Bhaskar's Philosophy*. London: Verso.

Connell, J. P., & Kubisch, A. C. (Eds.). (1998). *Applying a theories of change approach to the design and evaluation of comprehensive community initiatives: Progress, prospects, and problems*. Washington, DC: The Aspen Institute.

Connell, J. P., Kubisch, A. C., Schorr, L. B., & Weiss, C. H. (Eds.). (1995). *New Approaches to Evaluating Community Initiatives. Volume 1, Concepts, Methods, and Contexts*. Washington, D.C.: The Aspen Institute.

Cooperrider, D. J., Whitney, D., & Stavros, J. M. (2008). *Appreciative Inquiry Handbook* (2nd ed.). Brunswick, OH and San Francisco, CA: Crown Custom Publishing & Berrett-Koehler Publishers.

Cox, M. D., & Richlin, L. (Eds.). (2004). *Building Faculty Learning Communities*. San Francisco: Jossey-Bass.

Creswell, J. W., & Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage.

Crisp, R. (2006). *Reasons and the Good*. Oxford: Oxford University Press.

Crotty, M. (2003). *The Foundations of Social Research. Meaning and Perspective in the Research Process*. London: Sage Publications.

Cummings, T., & Worley, C. (2001). *Organization development and change* (7th ed.). Cincinnati, OH: Southwestern College.

Cummings, T., & Worley, C. (2014). *Organization development and change* (10th ed.). Stamford, CT: Cengage Learning.

Dalkin, S. M., Jones, D., Lhussier, M., & Cunningham, B. (2012). Understanding integrated care pathways in palliative care using realist evaluation: a mixed methods study protocol. *BMJ Open*, 2(4).

Danermark, B., Ekstrom, M., Jakobsen, L., & Karlsson, J. C. (2002). *Explaining Society: Critical Realism in the Social Sciences*. London: Routledge.

Davies, R. (2012). *Criteria for assessing the evaluability of Theories of Change*. Retrieved 26 January, 2014, from <http://mandenews.blogspot.co.uk/2012/04/criteria-for-assessing-evaluability-of.html>

de Souza, D. E. (2013). Elaborating the Context-Mechanism-Outcome configuration (CMOC) in realist evaluation: A critical realist perspective. *Evaluation*, 19(2), 141-154.

Destatis. (2013). *Studierende an Hochschulen - Wintersemester 2012/2013*. Fachserie 11 Reihe 4.1. Wiesbaden: Statistisches Bundesamt.

Downward, P., & Mearman, A. (2007). Retroduction as mixed-methods triangulation in economic research: reorienting economics into social science. *Cambridge Journal of Economics*, 31(1), 77-99.

Easterby-Smith, M., Burgoyne, J., & Araujo, L. (Eds.). (1999). *Organizational Learning and the Learning Organization: Developments in theory and practice*. London: Sage.

Edmondson, A., & Moingeon, B. (1998). From Organizational Learning to the Learning Organization. *Management Learning*, 29(1), 5-20.

Edwards, A. D., & Furlong, V. J. (1985). Ethnography and theory construction in educational research. In R. Burgess (Ed.), *Field methods in the study of education*. Lewes: Falmer Press.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Standford, CA: Stanford University Press.

Fetterman, D. (1994). Empowerment Evaluation. *American Journal of Evaluation*, 15(1), 1-15.

Finney, S. J., & DiStefano, C. (2006). Non-Normal and Categorical Data in Structural Equation Modeling. In G. R. Hancock, & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 269-314). Charlotte, NC: IAP.

French, W., & Bell, C. (1999). *Organization Development* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

French, W., & Bell, C. H. (1973). *Organization Development: Behavioral Science Interventions for Organizational Improvement*. Englewood Cliffs, NJ: Prentice-Hall.

Garvin, D. A. (1993). Building a Learning Organization. *Harvard Business Review*, 73(4), 78-91.

Gill, M., & Turbin, V. (1999). Evaluating Realistic Evaluation: Evidence From a Study of CCTV. In K. Painter, & N. Tilley (Eds.), *Surveillance of Public Space: CCTV, Street Lighting and Crime Prevention* (pp. 179-199). Monsey, NY: Criminal Justice Press.

Granovetter, M. S. (1973). The Strength of Weak Ties. *The American Journal of Sociology*, 78(6), 1360-1380.

Granovetter, M. S. (1978). Threshold Models of Collective Behavior. *American Journal of Sociology*, 83(6), 1420-1443.

Greenhalgh, T., Humphrey, C., Hughes, J., Macfarlane, F., Butler, C., & Pawson, R. (2009). How Do You Modernize a Health Service? A Realist Evaluation of Whole-Scale Transformation in London. *The Milbank Quarterly*, 87(2), 391-416.

Guba, E., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: Sage.

Guba, E., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging influences. In N. K. Denzin, & Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (3rd ed., pp. 191-215). Thousand Oaks, CA: Sage.

Habermas, J. (1984). *The Theory of Communicative Action. Vol. 1: Reason and the Rationalization of Society* (T. McCarthy Trans.). Boston/MA: Beacon Press.

Habermas, J. (1987). *The Theory of Communicative Action. Vol. 2: Lifeworld and System: a Critique of Functionalist Reason* (T. McCarthy Trans.). Boston/MA: Beacon Press.

Hair, J. F., Tatham, R. L., Anderson, R. E., & Black, W. (1998). *Multivariate Data Analysis* (5th ed.). New Jersey: Prentice-Hall.

Hall, E. T. (1976). *Beyond Culture*. Garden City, NY: Anchor Press/Doubleday.

Harvey, L., & Green, D. (1993). Defining Quality. *Assessment & Evaluation in Higher Education*, 18(1), 9-34.

Hawkins, P. (1994). Organizational Learning: Taking Stock and Facing the Challenge. *Management Learning*, 25(1), 71-82.

Hedberg, B. (1981). How organizations learn and unlearn. In P. C. Nystrom, & W. H. Starbuck (Eds.), *Handbook of Organizational Design* (pp. 8-27). Oxford: Oxford University Press.

Hedström, P., & Swedberg, R. (1998). Social Mechanisms: An Introductory Essay. In P. Hedström, & R. Swedberg (Eds.), *Social Mechanisms. An Analytical Approach to Social Theory* (pp. 1-31). Cambridge: Cambridge University Press.

Hedström, P., & Swedberg, R. (1996). Social Mechanisms. *Acta Sociologica*, 39(3), 281-308.

Henry, G. T., & Julnes, G. (1998). Values and realist evaluation. *New Directions for Evaluation*, 1998(78), 53-71.

Henry, G. T., & Rog, D. J. (1998). A realist theory and analysis of utilization. *New Directions for Evaluation*, 1998(78), 89-102.

Herr, K., & Anderson, G. L. (2005). *The Action Research Dissertation. A Guide for Students and Faculty*. Thousand Oaks/CA: Sage.

Hofstadter, R., & Metzger, W. P. (1955). *The development of academic freedom in the United States*. New York: Columbia University Press.

Holman, P. (2013). A Call to Engage. Realizing the Potential of Dialogic Organization Development. *OD Practitioner*, 45(1), 18-24.

Hochschulrahmengesetz (2007) [German Framework Act for Higher Education].

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.

Huber, G. P. (1991). Organizational Learning: The Contributing Processes and the Literature. *Organizational Science*, 2(1), 88-115.

Hutchings, P. (2001/2002). Reflections on the Scholarship of Teaching and Learning. *Essays on Teaching Excellence*, 13(2), 11 September 2011.

Ibarra, H., & Hunter, M. (2007). How Leaders Create and Use Networks. *Harvard Business Review*, 85(1), 40-47.

IOE (2013). *EdD (International) Student Handbook 2013/2014*. London: Institute of Education, University of London.

Jackson, S. F., & Kolla, G. (2012). A New Realistic Evaluation Analysis Method: Linked Coding of Context, Mechanism, and Outcome Relationships. *American Journal of Evaluation*, 33(3), 339-349.

Julnes, G., & Mark, M. M. (1998). Evaluation as sensemaking: Knowledge construction in a realist world. *New Directions for Evaluation*, 1998(78), 33-52.

Julnes, G., Mark, M. M., & Henry, G. T. (1998). Epilogue: Experience informing theory supporting practice. *New Directions for Evaluation*, 1998(78), 103-104.

Kaidesoja, T. (2007). Exploring the Concept of Causal Power in a Critical Realist Tradition. *Journal for the Theory of Social Behaviour*, 37(1), 63-87.

Karataş-Özkan, M., & Murphy, W. D. (2010). Critical Theorist, Postmodernist and Social Constructionist Paradigms in Organizational Analysis: A Paradigmatic Review of Organizational Learning Literature. *International Journal of Management Reviews*, 12(4), 453-465.

Kim, D. H. (1993). The link between individual and organizational learning. *Sloan Management Review*, (Fall), 37-50.

Klumpp, M., & Teichler, U. (2008). German Fachhochschulen: Towards the End of a Success Story? In J. S. Taylor, J. Brites Ferreira, M. d. L. Machado & R. Santiago (Eds.), *Non-University Higher Education in Europe* (pp. 99-122). Dordrecht: Springer.

Kolb, D. A. (1984). *Experiential Learning. Experience as the Source of Learning and Development*. Englewood Cliffs/New Jersey: Prentice-Hall.

Kolb, D. A., & Frohman, A. L. (1970). An organizational development approach to consulting. *Sloan Management Review*, 12(1), 51-65.

Korac-Kakabadse, N., Kouzmin, A., Korac-Kakabadse, A., & Savery, L. (2001). Low- and High-Context Communication Patterns: Towards Mapping Cross-Cultural Encounters. *Cross-Cultural Management*, 8(2), 3-24.

Kristiansen, M., & Bloch-Poulsen, J. (2004). Self-referentiality as a power mechanism: Towards dialogic action research. *Action Research*, 2(4), 371-388.

Lawson, T. (1997). *Economics and Reality*. London: Routledge.

Leone, L. (2008). Realistic Evaluation of an Illicit Drug Deterrence Programme: Analysis of a Case Study. *Evaluation*, 14(1), 9-28.

Levitt, B., & March, J. G. (1988). Organizational Learning. *Annual Review of Sociology*, 14, 319-340.

Lewin, K. (1951). In Cartwright D. (Ed.), *Field Theory in Social Science. Selected Theoretical Papers*. New York: Harper and Row.

Lijphart, A. (1971). Comparative Politics and the Comparative Method. *The American Political Science Review*, 65(3), 682-693.

Luhmann, N. (1996). *Social Systems* (J. Bednarz Trans.). Stanford, CA: Stanford University Press.

Lukas, W., Müller-Christ, G., Paskarbies, N., & Vogel, M. (2013). Organisationsentwicklung. In B. Jorzik (Ed.), *Charta guter Lehre* (pp. 68-75). Essen: Stifterverband für die Deutsche Wissenschaft.

Marchal, B., van Belle, S., van Olmen, J., Hoerée, T., & Kegels, G. (2012). Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation*, 18(2), 192-212.

Mark, M. M., & Henry, G. T. (1998). Social programming and policy-making: A realist perspective. *New Directions for Evaluation*, 1998(78), 73-88.

Mark, M. M., Henry, G. T., & Julnes, G. (1998). A realist theory of evaluation practice. *New Directions for Evaluation, 1998*(78), 3-32.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: an integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.

Marshak, R. J., & Bushe, G. R. (2013). An Introduction to Advances in Dialogic Organization Development. *OD Practitioner, 1*(45), 1-4.

Marsick, V. J., Bitterman, J., & Veen, R. v. d. (2000). *From the Learning Organization to Learning Communities toward a Learning Society*. Columbus, OH: Center on Education and Training for Employment, Ohio State University.

Maurer, M., & Githens, R. P. (2010). Toward a reframing of action research for human resource and organization development: Moving beyond problem solving and toward dialogue. *Action Research, 8*(3), 267-292.

McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. New York: McGraw Hill.

Mertens, D. M., & Wilson, A. T. (2012). *Program Evaluation Theory and Practice*. New York: The Guilford Press.

Merton, R. K. (1968). On Sociological Theories of the Middle Range. In R. K. Merton (Ed.), *Social Theory and Social Structure* (2nd ed., pp. 39-72)

Miller, D. (1996). A preliminary typology of organizational learning: Synthesizing the literature. *Journal of Management*, 22(3), 485-505.

Miller, K. D., & Tsang, E. W. K. (2011). Testing management theories: critical realist philosophy and research methods. *Strategic Management Journal*, 32(2), 139-158.

Mills, C. W. (2000). *The Sociological Imagination*. Oxford: Oxford University Press.

Mintzberg, H. (1980). Structure in 5's: A Synthesis of the Research on Organization Design. *Management Science*, 26(3), 322-341.

Mooi, E., & Sarstedt, M. (2011). *A Concise Guide to Market Research. The Process, Data, and Methods Using IBM SPSS Statistics*. Heidelberg: Springer.

Norman, G. R., & Streiner, D. L. (2003). *PDQ Statistics* (3rd ed.). Hamilton, Ontario: BC Decker.

Olsen, W. (2010). Editor's Introduction: Realist Methodology - A Review. In W. Olsen (Ed.), *Realist Methodology. Vol. I: Practical Realist Ontology* (pp. xix-xlvi). London: Sage.

Orton, J. D., & Weick, K. E. (1990). Loosely Coupled Systems: A Reconceptualization. *The Academy of Management Review*, 15(2), 203-223.

Owen, H. (2008). *Open Space Technology* (3rd ed.). San Francisco, CA: Berrett-Koehler.

Oxford Dictionaries. (2013). *Learning*. Retrieved 18 May, 2013, from <http://oxforddictionaries.com/definition/english/learning>

Palmer, P. J. (1993). Good Talk About Good Teaching. *Change*, 26(6), 8-13.

Palmer, P. J. (2007). *The Courage to Teach. Exploring the Inner Landscapes of a Teacher's Life*. San Francisco: Jossey-Bass.

Palmer, P. J., & Zajonc, A. (2010). *The Heart of Higher Education. A Call To Renewal*. San Francisco, CA: Jossey-Bass.

Parfit, D., & Broome, J. (1997). Reasons and Motivation. *Proceedings of the Aristotelian Society*, 71(1), 99-146.

Parsons, T. (1951). *The Social System*. Glencoe, IL: The Free Press.

Patton, M. Q. (1999). Book Review: Realistic Evaluation. *American Journal of Evaluation*, 20(2), 385-388.

Patton, M. Q. (2012). *Essentials of Utilization-Focused Evaluation*. Thousand Oaks, CA: Sage.

Pawson, R. (1995). Quality and Quantity, Agency and Structure, Mechanism and Context, Dons and Cons 1. *Bulletin De Méthodologie Sociologique*, 47(1), 5-48.

Pawson, R. (2000). Middle-Range Realism. *European Journal of Sociology*, 41(2), 283-325.

Pawson, R. (2002). Evidence-based Policy: The Promise of 'Realist Synthesis'. *Evaluation*, 8(3), 340-358.

Pawson, R. (2003). Nothing as Practical as a Good Theory. *Evaluation*, 9(4), 471-490.

Pawson, R. (2008). *Causality for Beginners*. Unpublished manuscript. Retrieved 23 March 2013, from http://eprints.ncrm.ac.uk/245/1/Causality_for_Beginners_Doc_07.doc

Pawson, R. (2013). *The Science of Evaluation. A Realist Manifesto*. London: Sage.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, 10(supplement 1), 21-34.

Pawson, R., & Manzano-Santaella, A. (2012). A realist diagnostic workshop. *Evaluation*, 18(2), 176-191.

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London: Sage.

Pawson, R., & Tilley, N. (2001). Realistic Evaluation Bloodlines. *American Journal of Evaluation*, 22(3), 317-324.

Pedersen, L. H., & Rieper, O. (2008). Is Realist Evaluation a Realistic Approach for Complex Reforms? *Evaluation*, 14(3), 271-293.

Ponterotto, J. G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology*, 52(2), 126-136.

Popper, M., & Lipshitz, R. (1998). Organizational Learning Mechanisms: A Structural and Cultural Approach to Organizational Learning. *The Journal of Applied Behavioral Science*, 34(2), 161-179.

Popper, M., & Lipshitz, R. (2000). Organizational Learning: Mechanisms, Culture, and Feasibility. *Management Learning, 31*(2), 181-196.

Preskill, H., & Torres, R. T. (1999). Building Capacity for Organizational Learning Through Evaluative Inquiry. *Evaluation, 5*(1), 42-60.

Quine, W. V. (1975). On Emprically Equivalent Systems of the World. *Erkenntnis, 9*, 313-328.

Ranmuthugala, G., Cunningham, F., Plumb, J., Long, J., Georgiou, A., Westbrook, J., et al. (2011). A realist evaluation of the role of communities of practice in changing healthcare practice. *Implementation Science, 6*(1), 49.

Rashman, L., Withers, E., & Hartley, J. (2009). Organizational learning and knowledge in public service organizations: A systematic review of the literature. *International Journal of Management Reviews, 11*(4), 463-494.

Reed, M. I. (2002). New Managerialism, Professional Power and Organisational Governance in UK Universities: A Review and Assessment. In A. Amaral, G. A. Jones & B. Karseth (Eds.), *Governing Higher Education: National Perspectives on Institutional Governance* (pp. 163-185). Dordrecht: Kluver Academic Publishers.

Richlin, L., & Essington, A. (2004). Overview of Faculty Learning Communities. In M. D. Cox, & L. Richlin (Eds.), *Building Faculty Learning Communities* (pp. 25-39). San Francisco: Jossey-Bass.

Ritzenhoff, P., & Vogel, M. (2012). Schweigen ist Silber, Reden ist Gold: kollegialer Austausch über die Lehre als Treiber von Wandel. In W. Benz, J. Kohler & K.

Landfried (Eds.), *Handbuch Qualität in Studium und Lehre* (pp. 1-20). Section E/2/8. Berlin: Raabe.

Robson, C. (2002). *Real World Research* (2nd ed.). Malden/MA: Blackwell Publishing.

Rogers, P. J. (1999). Book Review: Realistic Evaluation. *American Journal of Evaluation*, 20(2), 381-383.

Rothwell, R. (1994). Industrial Innovation: Success, Strategy, Trends. In M. Dodgson, & R. Rothwell (Eds.), *The Handbook of Industrial Innovation* (pp. 33-53). Brookfield, VT: Edward Elgar.

Samuelowicz, K., & Bain, J. D. (2001). Revisiting Academics' Beliefs about Teaching and Learning. *Higher Education*, 41(3), 299-325.

Sayer, A. (1992). *Method in Social Science: A Realist Approach*. London: Routledge.

Sayer, A. (2000). *Realism and Social Science*. London: Sage.

Schein, E. (1965). *Organizational Psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Schein, E. (1993). On Dialogue, Culture, and Organizational Learning. *Organizational Dynamics*, 22(2), 40-51.

Schelling, T. (1978). *Micromotives and Macrobbehavior*. New York: W.W. Norton.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago/IL: Rand McNally.

Scriven, M. (1991a). Beyond formative and summative evaluation. In M. W. McLaughlin, & D. C. Phillips (Eds.), *Evaluation and Education: At Quarter Century* (pp. 18-64). Chicago, IL: The University of Chicago Press.

Scriven, M. (1991b). *Evaluation Thesaurus* (4th ed.). Newbury Park, CA: Sage.

Senge, P. M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Currency Doubleday.

Shotter, J. (2010). Situated Dialogic Action Research: Disclosing “Beginnings” for Innovative Change in Organizations. *Organizational Research Methods*, 13(2), 268-285.

Shrivastava, P. (1983). A Typology of Organizational Learning Systems. *Journal of Management Studies*, 20(1), 7-28.

Shulman, L. S. (1993). Teaching as Community Property. *Change*, 25(6), 6-7.

Simon, H. A. (1999). Can there be a science of complex systems? In Y. Bar-Yam (Ed.), *Unifying themes in complex systems: Proceedings from the International Conference on Complex Systems 1997* (pp. 4-14). Cambridge, MA: Perseus Press.

Simon, H. A. (1991). Bounded Rationality and Organizational Learning. *Organization Science*, 2(1, Special Issue: Organizational Learning: Papers in Honor of (and by) James G. March), 125-134.

Skinner, Q. (Ed.). (1985). *The Return of Grand Theory in the Human Sciences*. Cambridge, UK: Cambridge University Press.

Sklar, L. (1981). Do Unborn Hypotheses Have Rights? *Pacific Philosophical Quarterly*, 62, 17-29.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

Stegmann, S. (2007). Herrenzimmer mit Sofa. Professoraler Habitus in universitären Alltagskulturen. *Freiburger Zeitschrift Für Geschlechterstudien*, 13(21)

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5), 893-898.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional Learning Communities: A Review of the Literature. *Journal of Educational Change*, 7(4), 221–258.

Stoll, L., & Louis, K. S. (2007). Professional learning communities: elaborating new approaches. In L. Stoll, & K. S. Louis (Eds.), *Professional Learning Communities: Divergence, Depth And Dilemmas* (pp. 1-13). Maidenhead: Open University Press.

Trowler, P. (2010). UK Higher Education: Captured by New Managerial Ideology? In V. L. Meek, L. Goedegebuure, R. Santiago & T. Carvalho (Eds.), *The Changing Dynamics of Higher Education Middle Management* (pp. 197-2012). Dordrecht: Springer.

Ullman, J. B. (2014). Structural Equation Modeling. In B. Tabachnick, & L. S. Fidell (Eds.), *Using Multivariate Statistics* (6th ed., pp. 731-836). Harlow: Pearson Education.

van der Knaap, L. M., Leeuw, F. L., Bogaerts, S., & Nijssen, L. T. J. (2008). Combining Campbell Standards and the Realist Evaluation Approach: The Best of Two Worlds? *American Journal of Evaluation*, 29(1), 48-57.

Vogel, M. (2009a). Exploring the conditions for academic teachers' informal collegial learning about teaching. A social network approach. *Educate~, 9*(2), 18-36.

Vogel, M. (2009b). The Professionalism of Professors at German Fachhochschulen. *Studies in Higher Education*, 34(8), 873-888.

Vogel, M. (2009c). Understanding Emergent Social Phenomena Comparatively: The Need for Computational Simulation. *European Journal of Social Sciences*, 7(4), 84-92.

Vogel, M. (2011). Pedagogical Solitude at a German Higher Education Institution. A Critical Discourse Analysis. Unpublished doctoral research study. EdD, Institute of Education, London.

Walker, M. (2001). *Reconstructing Professionalism in University Teaching*. Buckingham: The Society for Research into Higher Education & Open University Press.

Weaver, K., & Olson, J. K. (2006). Understanding paradigms used for nursing research. *Journal of Advanced Nursing*, 53(4), 459-469.

Weick, K. E. (1976). Educational Organizations as Loosely Coupled Systems. *Administrative Science Quarterly*, 21(1), 1-19.

Weick, K. E. (1982). Management of organizational change among loosely coupled elements. In P. S. Goodman & Associates (Ed.), *Change in organizations* (pp. 375-408). San Francisco, CA: Jossey-Bass.

Weick, K. E. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: Sage.

Weiss, C. H. (1995). Nothing as Practical as Good Theory: Exploring Theory-based Evaluation for Comprehensive Community Initiatives for Children and Families. In J. P. Connell, A. C. Kubisch, L. B. Schorr & C. H. Weiss (Eds.), *New Approaches to Evaluating Community Initiatives. Volume 1, Concepts, Methods, and Contexts* (pp. 65-92). Washington D.C.: Aspen Institute.

Wirtenberg, J., Abrams, L., & Ott, C. (2004). Assessing the Field of Organization Development. *The Journal of Applied Behavioral Science*, 40(4), 465-479.

Wittek, L., & Kvernbekk, T. (2011). On the Problems of Asking for a Definition of Quality in Education. *Scandinavian Journal of Educational Research*, 55(6), 671-684.

Woodhouse, D. (2003). Quality Improvement through Quality Audit. *Quality in Higher Education*, 9(2), 133-139.