Investigating the Evaluation of Higher Education in Germany: A Case Study of Educational Science (*Erziehungswissenschaft*) in Baden-Württemberg

Susan Harris-Huemmert

Thesis submitted to the University of Oxford
for the degree of Doctor of Philosophy

New College, Oxford Trinity Term 2009

ABSTRACT

This thesis is an empirical qualitative case study of the work of an evaluation commission which was established by the Ministry of Culture, Research and Art in the German state of Baden-Württemberg in 2003 to undertake a cross-sectional examination of the quality of teaching and research in universities and teacher training colleges offering courses in Educational Science (Erziehungswissenschaft). Although much literature on evaluation methods and models exists, little addresses the minutiae of evaluation work through the eyes of those doing it. This study therefore aims to augment this knowledge by examining the attitudes, experiences and difficulties which evaluators face. It contributes to the discourse on evaluation theory and practice by providing suggestions for what might help to improve decision-making processes in this particular environment. Informed by phenomenological theory, this exploratory study also uses decision-making theories as a means of providing knowledge on both individual and organisational issues which can augment our understanding of how expert commissions work: the social, political and cultural mechanisms that are involved, and the techniques that evaluators employ in order to provide institutions with an objective, fair, trustworthy and reliable evaluation report.

As external review has become an accepted means of quality assurance, this research constitutes a contribution to the discourse on ways of improving quality assurance in higher education on a broader scale, not just within the context of German higher education. Furthermore, it also offers insight into the discipline of Educational Science itself and the notion of competing research paradigms, which have an impact on the way the discipline perceives itself and is perceived by others.

The study is broadly divided into three main sections. The first contextualises the history of higher education evaluation, specifically within the German context. It looks at how the idea of what a university is has changed, especially during the last few decades of the 20th century, and notions of quality within this particular environment. The evaluation is also briefly introduced within the context of Baden-Württemberg. The second section explores the evaluation by examining the documents available in the *evalag* agency archive in Mannheim, thereby facilitating an understanding of the background to the commission and the processes which the evaluators underwent as work progressed. The third provides insight into what the evaluators themselves thought of the process and is based on interview analysis. The thesis concludes with a brief survey of more recent developments in quality assurance in Germany.

ACKNOWLEDGEMENTS

Writing a thesis is perhaps somewhat akin to driving a car. At times you are in the fast lane on a straight road making tremendous progress. Just as quickly you get diverted, run into thick fog, or have a puncture. The following have all been there to help me avoid major collisions.

● Firstly, I would like to thank my supervisor, Professor David Phillips, who provided the ignition for this project, and gave me valuable support, advice and encouragement throughout. His ability to guide my doctoral vehicle back on course when it was threatening to steer off route is exemplary. My thanks also go to his wife, Val, for her culinary support when I was working in their private library.

● My thanks go in particular to Dr Heribert Knorr of the Ministry of Science, Research and Art in Baden-Württemberg, Dr Gerhard Schreier, Bertelsmann Foundation, and the employees of the *evalag* agency in Mannheim, for having the vision to allow this research to take place.

● My interview partners, including many professors of Education, ministerial members and various heads of institutions were extremely obliging to give up of their time. They are listed in this thesis. I hope that they will enjoy reading the results.

● Within Oxford I would like to thank not only my college advisor, Mr David Palfreyman, for his support and insights into higher educational legislation and the Bologna Process, but also Professor Lynn McAlpine and Dr Nick Hopwood of The Learning Institute, for their ongoing encouragement. Barbara Whitfield, Luba Mandzy, Paul Tosio and Naomi Weiss of the cast of *An Ideal Husband*, and *The Importance of Being Earnes*t also ensured *mens sana* during occasionally difficult times.

● I would also like to thank Natalie Lundsteen, Helen Carasso, Dorota Campfield, Cindy Naumann, Rachel Mazyck, James Hall and Alaster Douglas, my fellow DPhil students. Cindy, James, Rachel and Natalie all helped out with occasional childcare and accommodation logistics, without which I would have had a much harder time completing.

● My thanks also go to Drs Evelyn Reed-Sanderson, Anna Zimdars and Alexander Foss for being so supportive and such good friends. My fellow Comparative and International Study Room colleagues Wen Wen, Jeremy Rappleye and Julia Paulson were great to work with. I enjoyed sharing parts of my journey with them.

● I would also like to thank Professor Rudolf Tippelt and Dr David Mills for taking the time to examine my thesis.

● My father died shortly before I took up my doctorate, but I know that he would have been proud of my personal Le Mans. My mother has been very supportive of my project throughout, for which I thank her so much.

Lastly, this journey would have been inconceivable without the emotional support, financial backing and continuing encouragement of my husband Martin, who remained in Germany for four years while I undertook my research. We covered thousands of miles in the process. His patience and love are beyond measure. My two children came to the UK with me and have been so much a part of the experience. They had unerring faith in me, for which I am so grateful.

This thesis is therefore dedicated, with much love, to Martin, Felix and Sophie.

GLOSSARY

The following is a list of acronyms which appear in this thesis.

DeGEval	Deutsche Gesellschaft für Evaluation
DFG	Deutsche Forschungsgemeinschaft
EFQM	European Foundation for Quality Management
ESRC	English Social Sciences Research Council
HE	Higher education
HEI	Higher Education Institution
HEFCE	Higher Education Funding Council for England
HRG	Hochschulrahmengesetz
HRK	Hochschulrektorenkonferenz
KMK	Kultusministerkonferenz
OECD	Organisation for Economic Co-operation and Development
PH	Pädagogische Hochschule
PISA	Programme for International Student Assessment
QA	Quality assurance
QAA	Quality Assurance Agency
RAE	Research Assessment Exercise
VSNU	Association of Universities in The Netherlands
WR	Wissenschaftsrat

TABLE OF CONTENTS

	Abstract	i
	Acknowledgements	ii
	Glossary	iii
	Table of contents	iv
	List of illustrations	viii
	List of tables	ix
Chapter 1	**Introduction**	1
1.1	Context of study	2
1.2	Thesis rationale	2
1.3	Research questions	3
1.4	Methodology and theoretical background	6
1.5	Thesis outline	7
1.6	Contribution of study	8
SECTION 1	**LITERATURE REVIEW; THEORETICAL FRAMEWORK METHODOLOGY**	
Chapter 2	**General contextualisation with definitions of the university, quality, evaluation and an introduction to higher education evaluation culture in Germany and the evaluation of *Erziehungswissenschaft***	10
2.0	The German university and Educational Science contextualised	12
2.1	Concepts of the university in brief	12
2.2	Educational Science in German higher education	15
2.3	Notions of 'quality' in higher education	17
2.4	Stakeholders in higher education and attitudes towards accountability	22
2.5	Definitions of evaluation	27
2.6	The habitus of the evaluator	31
2.7	Professors within the dichotomy of egalitarianism and elitism	33
2.8	Influence of Dutch evaluation methodology in German higher education	36
2.9	Analysis of German HE evaluation landscape	37
2.10	Political factors in association with German higher education evaluation	39
2.11	Influence of PISA and the Bologna Process on German higher education	41
2.12	Evaluation landscape in Germany	42
2.13	Evaluation agencies and significance of launch date	44
2.14	History of the *evalag* agency in Baden-Württemberg	45
2.15	*Evalag's* evaluation methodology	46
2.16	Impact of higher education evaluation in Germany 2000-2004	48
2.17	Background to evaluation of *Erziehungswissenschaft* 2003/2004	49
2.18	Summary	51
Chapter 3	**Theoretical Framework and Methodology**	52
3.1	Theoretical framework	52
3.1.1	Phenomenology and decision-making theory	52
3.2	Using qualitative methods	57
3.2.1	Case study method	58
3.2.2	Emic and etic perspectives	59
3.3	Reason for not conducting pilot study	59

3.4	Gaining overarching access	60
3.5	Archival research	61
3.5.1	Method of archival research	62
3.5.2	Limitations of archival research	64
3.6	Reason for conducting retrospective analysis	65
3.7	Method of interviewing agency employees	66
3.7.1	Method of interviewing evaluators	67
3.7.2	Duration; location and number of interviews	68
3.7.3	Choice of interview partner / sampling	69
3.7.4	Interviewing elites: culture and the researcher	70
3.7.5	Limitations of participant recall	73
3.7.6	Ethical considerations regarding interviews	74
3.7.7	Method of interview data analysis	76
3.8	Summary	79

SECTION 2 DOCUMENTARY ANALYSIS

Chapter 4	**Chronology of evaluation as revealed through documentary analysis of *evalag* archives**	81
4.1	First steps into documentary research	81
4.2	Historical background preceding evaluation 2003/2004	83
4.3	Organisational steps towards evaluation of *Erziehungswissenschaft*; composition of expert committee	91
4.4	Ministry's criteria for selection of experts	97
4.5	Context preceding constitutional meeting 16 April 2003	99
4.6	Constitutional meeting 16 April 2003	101
4.7	Compilation of questionnaires A & B for self report	107
4.7.1	Professor Phillips's comments on draft questionnaire of 10 April 2003	108
4.8	Interim organisational details	114
4.9	Preparations in advance of plenary meeting 25 September 2003	116
4.9.1	Plenary meeting 25 September 2003	121
4.10	Towards determining evaluation criteria for *Erziehungswissenschaft*	125
4.11	Interim correspondence and proceedings prior to onsite visits	127
4.12	The Ulm Declaration 2001	129
4.13	Allocation of experts to HEIs for onsite visits	131
4.14	Developing evaluation criteria for judging quality of research	133
4.15	Typical timetable for HEI visits; onsite reflections	137
4.16	Post onsite plenary meeting 10-11 December 2003	140
4.17	Meeting of sub-committee 16 January 2004	144
4.18	Plenary meeting 27 February 2004	147
4.19	The Achtenhagen/Schelten statement 23 March 2004	151
4.20	Final meeting with HEI representatives, 1-2 April 2004	152
4.20.1	Discussions with representatives from PH Weingarten	154
4.21	Editorial meeting, 29 April 2004	159
4.22	Agreed general criteria regarding resources for research	161
4.22.1	Scientific quality of research	162
4.22.2	Productivity and practical relevance of research	162
4.22.3	Research funding	163
4.23	Early career academic support	163
4.24	Cooperation and international networking	164
4.25	Criteria for judging the structures of teaching and learning provision	164
4.26	Summary	165

SECTION 3 INTERVIEW ANALYSIS

Chapter 5	**Context of evaluation: roles of individuals and their work ethics**	170
5.1	Special areas of knowledge, age of experts, and other experience	171
5.2	Role of non-specialist expert	173
5.3	Prior knowledge of *evalag* and recruitment to evaluation	174
5.4	Familiarity with other colleagues / previous networks	175
5.5	How experts viewed their selection	179
5.6	Reasons for participating in evaluation	181
5.6.1	Knowledge of institutions	181
5.6.2	Curiosity because of peers	182
5.6.3	Accommodation of evaluation into existing schedules	183
5.6.4	Experience of being evaluated	184
5.7	Significance of 1st plenary meeting	186
5.8	Extent of ministerial involvement	188
5.9	Experts' opinions on how HEIs perceived evaluation	193
5.10	Specific roles within evaluation commission	196
5.10.1	Role of foreign experts	196
5.10.2	Role of chairperson	199
5.10.3	Role of sub-committee	201
5.10.4	Role of Baumert	202
5.11	Summary	203
Chapter 6	**Establishing evaluation criteria: reaching consensus on the notion of quality in German Educational Science**	205
6.1	Working towards evaluation criteria	205
6.1.1	Frameworks for quality	208
6.1.2	Considerations for 'first-time' evaluations	212
6.1.3	Camps and hierarchies: consequences for establishing evaluation criteria	213
6.1.4	Influences from The Netherlands	216
6.1.5	Publications and impact factor	220
6.1.6	Research funding and notions of quality	223
6.2	Summary	226
Chapter 7	**Organisational issues and problems; self-concept of evaluators and learning processes**	228
7.1	Techniques for working with large amounts of information: the self reports	228
7.1.1	Self reports and the art of seeing through façades	230
7.2	Attitudes to gaps in statistics	233
7.3	Numbers of students at onsite meetings	236
7.4	Compiling reports	238
7.5	Discrepancies between discussions and final report	240
7.6	How the experts thought they were perceived by HEIs	243
7.6.1	Self-concept of evaluators	245
7.6.1.1	Ability to reach compromise	245
7.6.2	Knowledge of field and publishing experience	246
7.6.3	Empathy with academic staff	248
7.7	Gender issues	249
7.8	Difficulties with HEIs	251
7.9	What the experts learnt from the evaluation	255
7.10	Summary	259

CONCLUSION

Chapter 8	**Review of the evaluation, critical discussion and conclusions**	260
8.1	Ministerial contract, impact on evaluation and appointed experts	261
8.2	Connectivity of experts and differing paradigms	263
8.3	Management of literature	264
8.4	Statistical problems	265
8.5	Experiences gained onsite and strategies for dealing with specific problems	266
8.6	Exploiting particular expertise	267
8.7	Problems of size	267
8.8	Attitudes to selection, evaluation experience and status	269
8.9	Choice of expert and impact on evaluation	271
8.9.1	Evaluator types	272
8.10	Evaluatory impact and perception of experts	273
8.11	Conclusions	274
8.12	Limitations of study	276
8.13	Implications for policy and future research	276

POSTSCRIPT

Chapter 9	**Developments in German HE quality assurance since 2004**	278
9.1	Introduction of evaluation standards and the 'Exzellenzinitiative'	278
9.2	Organising quality assurance in HE	280
9.3	Evaluations in *Erziehungswissenschaft*	282
9.4	Changing role of *evalag*	282
9.5	A vision for the future?	283

References 285

Appendices 301

A:	Evaluations of *Erziehungswissenschaft* in Germany until 2004	301
B:	Map of 15 HEIs in evaluation of *Erziehungswissenschaft*	302
C:	Informed consent for participants in a research project	303
D:	Questions for evaluators	304
E:	Interview participants, place and date of interview	307
F:	Questionnaires A & B (final version)	308
G:	Specimen interview	315
H:	Ulm Declaration	340
I:	Achtenhagen / Schelten statement	344

LIST OF ILLUSTRATIONS

Figure 1:	Adaptation of Phillips & Filmer-Sankey (2003) research model	6
Figure 2:	Holistic view of evaluator in phenomenological context	55
Figure 3:	Matrix of evaluator contacts	177

LIST OF TABLES

Table 1: Organisation of evaluations throughout Germany 43
Table 2: Participants in plenary meeting 16 April 2003 102
Table 3: Evaluation schedule 6 May 2003 106
Table 4: Participants in plenary meeting 25 September 2003 121
Table 5: Allocation of experts for first onsite week 20-24 October 2003 131
Table 6: Allocation of experts for second onsite week 3-7 November 2003 131
Table 7: Allocation of experts for third onsite week 17-21 November 2003 132
Table 8: Allocation of experts in HEI meeting, morning, 1 April 2004 153
Table 9: Allocation of experts in HEI meeting, afternoon, 1 April 2004 153
Table 10: Allocation of experts in HEI meeting, morning, 2 April 2004 153
Table 11: Allocation of experts in HEI meeting, noon, 2 April 2004 154
Table 12: Overview of evaluation 167
Table 13: Area of expertise/specialisation 171
Table 14: Positions of responsibility held be evaluators 172
Table 15: Skills of evaluators 275

Chapter 1

Introduction

1.1 Context of study

The OECD conducted its first Programme for International Student Assessment (hereafter PISA) study in 2000, which examined the reading literacy among 15 year old children in 47 countries (see http://www.pisa.oecd.org). Shortly before this, in June 1999, 29 European countries signed the Bologna Declaration in which they documented their intention to create convergence in their higher education structures (see http://www.bologna-berlin2003.de/pdf/bologna_declaration.pdf). Both of these events have had a momentous effect on German higher education. An avalanche of debate in relation to quality and standards in both secondary and tertiary education in Germany, in particular with regard to teacher training, was initiated by these two events. Part of this discussion surrounded the ongoing discontent and feeling that teacher training in Germany overall was 'off the mark' and that children were not subject to the best teaching practice, which was perceived as one of the reasons for Germany's poorer PISA results than those in other industrialised countries. Assessment practice in association with quality assurance therefore became increasingly important within both schools and universities, mirroring increasing public concern with accountability. From a position of employment in German secondary education I became interested in how institutions of higher educations (HEIs) assess the quality of what they do. Germany had not addressed the quality of its HEIs very directly before this point, although there was a general understanding that its higher education system would benefit from reform.

External peer review, an approach to quality assurance which had already been introduced in various universities around the world (e.g. Canada and The Netherlands to name just two examples) was seen as one main way of assisting German HEIs to learn more about their practice and help to improve the system in general. During the late 1990s and early part of the 21st century agencies were established to assist universities with the

organisation of evaluations, a process which was seen as informative and supportive, with none of the possible financial retribution commonly associated with quality assurance (hereafter QA) practice in other countries, notably Great Britain, where HEIs can suffer financial losses as a consequence of poor assessment results. Some of these agencies were local, working for one federal state only, whereas others spanned separate states, but organised concurrent evaluations of disciplines across state borders. Others again organised individual university groups (see Chapter 2). The practice of these agencies was also diverse, with some organising one-off evaluations, whereas others organised waves of evaluation, with repeats after a few years. One characteristic which was common to all, however, was that they employed external experts. I started asking questions such as: Who are these experts? What is their training or their background? How are they chosen? What is their disciplinary understanding, or does this vary? What does it feel like to be an external reviewer assessing your own peers? These questions persisted and I became preoccupied with definitions of evaluation, quality, peer review processes and so on. One of my underlying assumptions was that something could be done to improve the system. Evaluation itself might even be done better, but before I could determine that, I needed to see how evaluation functioned.

1.2 Thesis rationale

The literature deals substantially with evaluation theory and practice and has already gained a clear notion of the kinds of issues evaluators have to deal with. However, hardly any literature offers insight into the processes of actual evaluations, which is acknowledged as an area that needs more research (Mittag, 2006). Following his own in-depth analysis of present evaluation literature and forming a similar conclusion with regard to a general lack of research on evaluation processes and those involved, Balzer (2005) remarks:

> A consistent attention to the results of such empirical research approaches should be of theoretical and practical relevance and as a consequence be of no little influence to theoretical evaluation approaches and the form of evaluation projects.

I decided then that there was scope for a study which would look at the minutiae of evaluation work, as this might give those involved in HE quality assurance insight into some of the aspects of evaluation practice which they might need to consider more carefully when asking external experts to assess their HEI. The purpose of this study then is to explore what happens when evaluators of higher education go about their work and examine the decision-making involved in their selection prior to the evaluation, but also the decision-making that is involved throughout the evaluation. As Van Vught (1997, p.82) states, the 'definition, selection and training of peers to be employed in a review procedure merit a good deal of attention.'

My study expands on work already undertaken by researchers such as Alkin, Daillak and White (1979), Balzer (2005), Volkov and Volkova (2005), and Mittag (2006) who have examined the work of evaluators. There remains nonetheless an acknowledged lack of research in this field, most of which has been gained within either British or American contexts. My study aims to describe the mechanisms of evaluation work through the eyes of those involved, in particular the evaluators themselves. This study is also unique as to date it is the only one which describes in detail the work of an evaluation organised by the *evalag* agency, which operates in one of Germany's largest federal states, Baden-Württemberg. My study therefore provides an original contribution to our understanding of how evaluation commissions work, as not only is this area under-researched in general, but also there is specific lack of knowledge of it within the German academic field.

1.3 Research questions

Having determined the locus of my study, I grappled for a long time with the precise questions I wished to ask. I wanted to explore the experience of evaluation in all of its facets and with all of the 'messiness' that I imagined it involved, and so I wished my research questions to reflect this. When I began to conceptualise them I realised that they would have to be rather general to allow for the breadth of what I believed I would be exploring. By being too specific I would be limiting what I could explore. Although I felt

instinctively that I knew what I was searching for, it was hard to encapsulate this in a manner which would summarise all of what I was trying to find. Fortunately I was pointed in the direction of a research model which seemed a useful starting point for my own work. Although it was constructed for an entirely different context (schools), I adapted it and constructed a model that was appropriate to my own research field (see Fig.1 overleaf**)**:

> What kind of *experience* do evaluators have when they engage in large-scale evaluations of higher education?
>
> What *attitudes* towards evaluation and higher education quality analysis do evaluators have and do these change as work progresses?
>
> What kinds of *difficulties* do evaluators encounter and what strategies do they develop to deal with these?

By focussing on the *experiences* and *attitudes* of evaluators during this process and the *difficulties* they encountered, I hope to include all of the aspects which this particular evaluation covered: the emotive aspects which this work entailed, the backgrounds and values of those concerned, the influences from various disciplinary areas, networks, the problems they encountered when they went onsite, their decision-making strategies and the resolution of their assessment in the writing-up of their final report. It is based on the overall assumption that evaluations are now standard practice in HEIs as a means of assessing their quality.

Although the above research questions appear to be discrete, there is considerable overlap in the process of evaluation. Experience is what is felt, suffered or learnt and this is inextricably linked to the individual's attitude or way of feeling about something (*Eigenwelt*) (Van den Berg, 1972; Giorgi 1985, 1992, 1994). As Jung (1921/1971) reminds us, attitudes are what we like or dislike about something. Attitude is a 'readiness of the psyche to act or react in a certain way', and this readiness can be influenced at any point. The 'culture' of each evaluator determines which aspects in an evaluation will be attended to or not (cf. Sproull & Larkey, 1979, p.99).

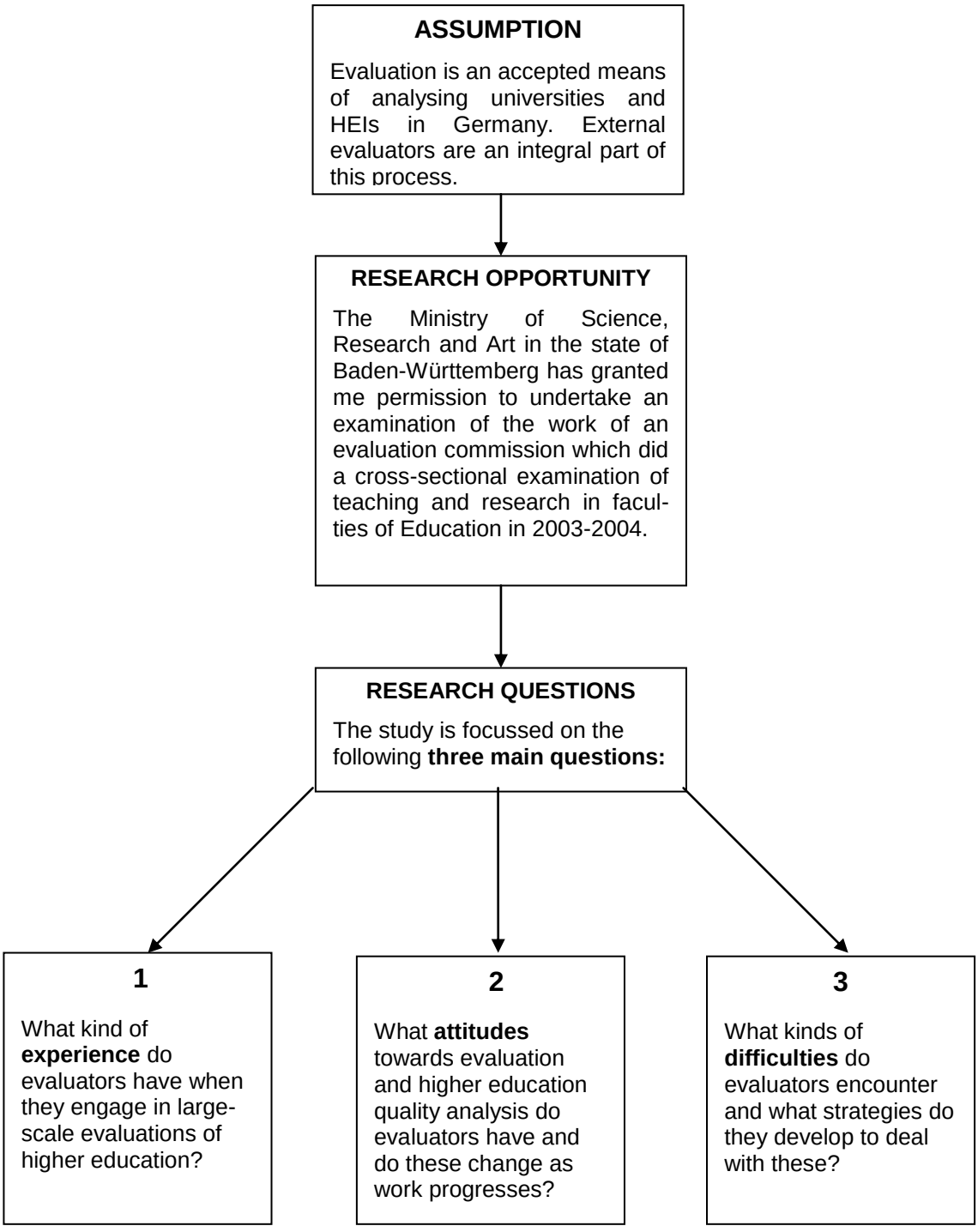


Figure 1: Adaptation of Phillips & Filmer-Sankey Research Model (1993)

Each evaluator comes to the task with a personal history, lesser or greater experience of the educational environment (*Lebenswelt*), greater or lesser reputation and knowledge of the field. Furthermore, the experience of evaluation touches on the structures of the overarching *Überwelt* of the university (Spinelli, 1989; Van der Mescht, 2004). Together, the evaluators try to uncover the reality of each HEI and engage in discourse with each

other (*Mitwelt*) to determine the criteria according to which they will assess these. While attempting to reach consensus therefore, a host of influences play out, some of which are subsumed, whereas others contribute to the final decisions that are taken.

1.4 Methodological and theoretical background

The research has been mainly informed by phenomenological, but also decision-making theory, which I describe in greater detail in Chapter 3. It employs qualitative interviewing techniques and documentary analysis to examine the evaluators and gain understanding of what they do, not only from their perspective, but also from the perspectives of those involved: the commissioning body (the Ministry of Culture, Research and Art in Baden-Württemberg), the institutions being examined, and members of the evaluation agency responsible for its organisation. I use a number of lenses to focus on various aspects of what takes place, adding myself as a further lens. A period of employment at Regensburg University also gave me insight into German university structures and systems, although I appreciate that much has changed since my term of employment in the 1980s. Nonetheless, being both an insider and an outsider coming to this research from a British perspective adds an additional perspective. Apart from being a witness to ongoing evaluative work, which I attempted to realise, but which was prohibited for a number of reasons, I believe that this is as near as I can get to the people concerned. Documentary analysis of institutional self-reports, minutes of meetings, diary entries, memos etc. helps to triangulate data.

Although pilot studies can help to establish the most appropriate questions in interviews, this is a channel of opportunity I was unable to pursue as there were 18 evaluators and 15 institutions in this evaluation. However, to prepare myself for interviewing, I held informal discussions with professors of Educational Science from other countries (Austria, Canada, Germany, and Mexico). This enabled me to develop an idea of the kinds of questions and issues that should be addressed.

The identities of those involved are in the public domain, as the evaluation report has been published. Documents on the evaluation are also available, albeit with the permission of the Ministry involved. At least three of the evaluators in the commission occupied specific positions which are identified in the public report. They have, however, granted permission for quotations from interviews to be used in this thesis. In my interview analysis quotations have been anonymised unless the discussion turned to the function of a specific individual (e.g. chairperson). Inevitably, it is the researcher's own decision as to which quotations to use. On some occasions where the subject matter might have implicated an individual or an institution I have refrained from using a quotation, even if it was especially pertinent.

1.5 Thesis outline

The first chapter provides the introduction to the present study and discusses briefly why the evaluation of Educational Science was commissioned in 2003/2004. It provides the research rationale and presents the overarching research questions to the study.

Chapters 2 and 3 fall into the first section of my thesis (literature review, theoretical framework and methodology). Chapter 2 explores definitions of the university and in Germany in particular and situates Educational Science within the German HE context, as this was the subject under examination. It reviews definitions of quality and evaluation, presents descriptions of evaluation theory and practice, and looks at the rise of quality assurance within academic contexts and in Germany. This chapter concludes by examining the German evaluation landscape at the time of the evaluation of Educational Science 2003/2004 and situates this case study within the political framework of Baden-Württemberg.

Chapter 3 describes the methodology of my work, examining the use of documentary and interview analysis within the theoretical framework of phenomenology and decision-making.

Chapter 4 (the second section) presents a detailed chronology of the evaluation based on documentary evidence from the archives. As I was not witness to the evaluation, I needed to reconstruct as much of what happened in order to begin to understand what took place while it was being conducted. This was possible by examining the archives of the *evalag* agency and piecing together historical evidence in the form of minutes, emails, correspondence, self reports, ministerial documents, diary extracts, drafts of the final report, and so forth.

The third section of the thesis is devoted to interview analysis. Chapter 5 examines the context of the evaluation through the eyes of the individual experts from information gained in interviews (oral history). It explores individual roles, their backgrounds and special areas of knowledge and specifically their attitude to the ministerial commissioning body. Chapter 6 reveals the work entailed in reaching consensus on the criteria which would be used for assessing quality in this evaluation. Chapter 7 investigates the evaluation in terms of its organisation and the problems that arose as work progressed. It explores the self-concept of evaluators and what they learnt during the process.

Forming the conclusion, Chapter 8 provides a final analysis of the evaluation, engages in a critical discussion of issues arising from this study, and offers some conclusions as to directions for future evaluations.

Chapter 9 forms a postscript to the study, as it looks briefly at the developments that have taken place in the field since this evaluation was concluded and looks at the role of *evalag* in the changing landscape of university evaluation.

1.6 Contribution of study

This research offers an ideal opportunity to expand our knowledge of how evaluators of higher education in Germany go about their work, the problems they encounter and the

strategies they develop to overcome them. It is appropriate in the context of improving quality control in higher education, which is part of an ongoing international debate.

This study investigates a single evaluation commission which was concerned with one academic discipline only. Experiences from an evaluation of a different discipline might reveal a different picture. However, I believe that the findings suggest similar problems might be an issue in other evaluations. Perhaps this line of investigation might open after this study has been completed. I am extremely grateful not only to the Ministry of Science, Research and Art in Baden-Württemberg for allowing this research to proceed, but also for the welcome that members of the *evalag* agency have extended to me and the cooperation of my interview partners. I hope that this research may be of some interest to them.

Section1: LITERATURE REVIEW, THEORETICAL FRAMEWORK, METHODOLOGY

Chapter 2

General contextualisation with definitions of the university, quality, evaluation and an introduction to higher education evaluation culture in Germany and the evaluation of *Erziehungswissenschaft*

'The purpose of evaluation is not to prove, but to improve' (Egon Guba)

Guba's simple premise, which appears on *evalag's* website, lies behind the work of the evaluation at the heart of this study, namely to identify what there is, and make suggestions for improvements if necessary. This study focuses on the people who examine the status quo in German HE with an aim to suggesting ways of improving it. To do this, a number of areas need to be considered and defined. The following chapter therefore establishes the context within which this study is to be understood and helps to avoid misunderstandings of interpretation.

In the first part I examine how the concept of the *university* in Germany has developed and where it now stands in the early 21st century, as this is the over-arching backdrop to the work of the evaluators in this study. I therefore also discuss German Educational Science, as it is this discipline from which the experts in this evaluation commission originated.

Secondly, and following on from an overall contextualisation, I explore notions of *quality* within higher education, and in Germany in particular. Quality is possibly one of the most difficult terms to define accurately as it encompasses a myriad of notions. However, we

need to engage with the different concepts of quality as these impact on the way in which evaluators work and their understanding of their evaluative task. This section therefore includes an examination of the various stakeholders involved in quality assurance in higher education and their different interests within the HE arena.

I then move into an examination of what evaluation is, and of how evaluations of higher education have changed over time. This also looks at the influences which different countries have had in how evaluations are conducted in Germany today. It is therefore necessary to discuss the term *evaluation* and examine how evaluation theory and practice have developed during the last few decades. As we also need to consider the political influences that existed at the time in which evaluations are conducted, I examine two main factors that were pertinent to the evaluation in this case study: the Bologna Process and the PISA studies. As the main emphasis of my research is a discussion of the experiences and attitudes of the evaluators, I conclude this section with an examination of the *habitus* of the evaluator situated within German Educational Science.

Lastly, I turn to the different evaluation agencies that have been established in German higher education and look at the methodology of the *evalag* agency in particular and the reasons behind the evaluation of Educational Science, which forms the focus of this present study.

All of the above forms a kind of *bricolage*, a term which Denzin and Lincoln (2000) used in reference to the range of methods which can be employed in educational research. I see myself also as a *bricoleur* bringing together a number of different areas within this literature review, challenging me to create a narrative which combines all of the above elements into a coherent story. I hope that by addressing the above aspects we are able to form a mutual understanding of the contexts in which the evaluation in this study took place and are suitably equipped to enter into the detail of what the experts experienced as the evaluation progressed.

2.0 The German university and Educational Science contextualised

2.1 Concepts of the university in brief

There have been numerous attempts to conceptualise what a university is and what it should aim to accomplish (see Newman, 1899 (reprinted 1996); Humboldt, 1879 for two main schools of thought which have been of major influence throughout the world). No one would argue today that a university's role should be to provide a place for teaching universal knowledge in order to mandate the presence of theology as a science of sciences as was one of Newman's underlying ideas. He believed that universities should develop the mind, expand outlook, and increase an ability to interact in society. To use his terms: they should develop that which was 'gentlemanly', which included notions of uprightness of character, discretion and judgment. Those being taught were expected to develop a sense of responsibility to society, a view later taken up by Northrop Fry when he said that 'it is the university's task to define the vision of society' (O'Grady & French, 2000, p.420). Indeed, Newman's notion of a university education appears incompatible with the present notion of vocational or professional training which some might put forward as being one of the university's main roles today.

Humboldt's notion of the university combined the idea of culture with the development of the individual; teaching and research were synthesised in this concept (Readings, 1996). It focused on three 'unities': educators and learners were equal, as they were both engaged in a mutual pursuit of knowledge; research and teaching went hand in hand; lastly, knowledge as a whole was united, with philosophy as the supreme discipline which brought together the empirical and the speculative (Schleiermacher, 1808/1956).

The German university allowed students to develop their own interests in any number of subjects and it served the purpose of broadening their horizons as they saw fit in the manner of self-cultivation. It should protect and develop the national culture. However,

referring to the role of the state in higher education Humboldt was specific by saying that it also had particular duties:

> […] there [should be] wealth (strength and diversity) of intellectual power through the selection of those men to be gathered together and for freedom in their effectiveness. However, freedom is threatened not only by the danger inherent in this, but also from the institutions themselves which, as they begin to foster a particular spirit, can suffocate the emergence of another. The [state] should take measures to prevent the disadvantages which might arise from this. (Humboldt, 1810, p.380) [1]

A university education was for a tiny percentage of students who came from the *Gymnasium*[2], having gained the so-called 'maturity' (*Reife*). They were thus already the elite of society in terms of their intellectual status, and so the notion of excellence was implicit. If we examine the American context during the latter part of the 19th century, we see that colleges such as Princeton and Yale were operating in a manner similar to the *Gymnasium* where courses were rigid and prescribed, whereas Harvard was beginning to move towards the Humboldt idea where students were free to select their own courses (Chamberlain, 1885). It was not until after the Second World War when increasing numbers of students started enrolling that the notion of prescribed study began to take root and more restrictions were put in place for access to certain disciplines such as medicine.

Before the First World War the core of the German university was regarded as 'healthy' (Becker, 1996, p.7). However, Simon (1996, p.7) rephrased this expression following vast increases in student numbers leading to overflowing lecture theatres and poor morale among teaching staff: 'The university's core is rotten'.[3] Glotz (1996) portrayed the German university in the last decade of the twentieth century as a very different place from its predecessor in the first half of that century as the students themselves now viewed their studies in a different light, more in terms of leading directly to a chosen career rather than as a time when they could experiment in any number of different areas (ibid., p.55).

[1] […] Reichtum (Stärke und Mannigfaltigkeit) an geistiger Kraft durch die Wahl der zu versammelnden Männer und für Freiheit in ihrer Wirksamkeit. Der Freiheit droht aber nicht bloß Gefahr von ihm, sondern auch von den Anstalten selbst, die, wie sie beginnen, einen gewissen Geist annehmen und gern das Aufkommen eines anderen ersticken. Auch den hieraus möglicherweise entstammenden Nachteilen muss er vorbeugen.

[2] The *Gymnasium* is the most academic school type in Germany which only takes in students thought capable of continuing their education to university level.

[3] Die Universitäten sind im Kern verrottet.

Equally, he suggested, the management structures of the German university were unable to cope with the increasing numbers of students entering higher education. One of his proposals for improving the situation was to increase the professionalism of university leadership by providing training which would help those concerned to come to terms with the new tasks they would be facing (ibid., p.110).

Like Glotz, Readings (1996) discusses his understanding of the university at the end of the 20th century and he has provided us with an interesting notion of what the university today is when he describes the environment as one in which society has moved away from the notion of the 'modern' university, by which he means the German Humboldtian idea, to one that has superseded itself and therefore become 'posthistorical'. He argues that the university as such has lost its sense of culture as the object, as both origin and goal, of the human sciences. Although this appears pessimistic and one might argue with this notion, it is against this backdrop of a changing status and rapid developments in higher education that the nascence of quality assurance should be regarded.

Barnett (2000) was the first to state that the age in which universities today find themselves is 'supercomplex' as society has become more consumer-oriented and is increasingly losing its faith in professionals and standards overall, in spite of the audit culture we find ourselves in. The present public disillusionment among the British public with its politicians and their code of ethics regarding tax claims is a case in point. The university's role is inevitably changing as governments and society in general place new expectations on what they believe a university should be achieving. As a result, everything and everyone within the higher education arena is coming under close scrutiny. Brint (1994) has commented on the resulting erosion of autonomy as external (usually market) forces increasingly dictate what can and cannot be done within the university.

Whatever our understanding of what a university is, or what it should be doing, it remains a place in which individuals operate autonomously *and* in cooperation with others in the

quest to acquire knowledge and develop what is known further. Its leaders are academic experts in their own rights who take on additional management tasks, often without any direct management training. This will have usually been acquired in 'lesser' positions as directors of departments, or other similar roles. Academia encompasses a wide range of disciplines, outlooks and even abilities, which it is the task of this leadership to identify, nurture and encourage to maintain standards and improve when necessary. Expanding on Ouichi's (1979) conceptual framework for analyzing how organisations 'can be managed to move towards their objectives,' Frackmann (1992) concludes that universities consist of clan-like structures which are becoming increasingly monitored by external bureaucracy, a risk factor as external control mechanisms may not understand the innate cultures of what they are examining.

Before we move to a discussion of definitions of quality and examine how quality assurance has impacted on German higher education, let us look at the specific discipline of Educational Science in which the evaluation at the heart of this study is embedded.

2.2 Educational Science in German higher education

Educational Science in Germany is historically located as originating within the fields of philosophy and psychology, two disciplines which remain of fundamental importance. If we look at the total numbers of students engaged in Germany who take either teacher training courses or take Education as a subject in its own right, Education is the third largest subject, and huge numbers of students are enrolled on teaching courses (Otto et al., 2000, p.15). They may take the state examination (*Staatsexamen*) to qualify as a teacher in any one of the main different school types (e.g. junior school, special needs, *Haupt-* or *Realschule, Gymnasium)*[4]. Once the first state examination is completed, the *Lehramt* (teaching) candidates spend two more years as trainee teachers (*Referendare)* in teaching

[4] Students have to decide from the outset of their studies which school type they wish later to be employed in. At age 11 a child is sent to either the *Hauptschule* for vocational training, the Realschule for a mixture of academic and vocational training, or the *Gymnasium*, for those aiming at tertiary education. So future teachers have to decide from the outset which subjects within which school type they wish to teach. For example, a *Gymnasium* teacher will be qualified to teach two full academic subjects, whereas a teacher for the *Hauptschule* will be trained in a number of less specialised subjects.

seminars, most of which is spent in schools on teaching practice. This period culminates in the second state examination which entitles candidates to apply for civil servant status on appointment, as teachers are state employees (Pritchard, 2004).

Depending on the type of *Staatsexamen*, e.g. if the student is preparing herself to teach at the *Haupt-* or *Realschule*, or *Gymnasium*, the extent of her specific knowledge in pedagogy will vary considerably. Those training to work in the *Gymnasium* will have a far deeper knowledge of two academic subjects and will concentrate less on pedagogy, whereas for those training for the *Hauptschule* the situation will be reversed with a far broader spread of subjects. Those studying for work in other areas such as primary education and special needs will have even more differentiated subject knowledge. Special needs, for example, is a vast area in itself, covering those educating the blind, deaf, dumb and those with various syndromes. It also includes the training of people rehabilitating from accidents or illnesses who have to re-learn various taken-for-granted abilities such as speech or movement, etc. Educational Science throughout Germany is now taught in universities, with the exception of Baden-Württemberg, where the training of non-*Gymnasium* teachers is still handled by teacher training colleges (*Pädagogische Hochschulen*). Areas of specialisation are diverse, and so a professor working in one particular area may have no knowledge of the research, literature, or international trends in another.

Those not wishing to become teachers have, until recently, taken either Diplomas or Masters examinations which were usually highly theoretical. Bachelors and Masters degrees have now been introduced to replace the former system. The discipline therefore encompasses not only vast numbers of students, but a whole array of further subjects and different specialised areas. It is rare for students to enter Educational Science at postgraduate level from another discipline, a situation which is diametrically opposed to that in the UK where large numbers enter postgraduate studies in Educational Science from any number of other disciplines.

The discipline has been categorised by Kolb (1981) as 'concrete-active' (soft applied), by which he meant that research is predominantly based on interpretation rather than 'hard' empirical evidence. Although we might argue with this interpretation today German Educational Science at the start of the 20th century was based on philosophy, a discipline which occupied high international prestige and hermeneutic in its outlook. Educational Science has undergone an interesting evolution in terms of its development which has seen numerous swings from either more empirically-based (hard) research, which today is regarded as high in status, to that which is hermeneutic (soft) and therefore apparently lower in status (McCulloch & Richardson, 2000). The debate among educationalists continues to this day and we shall see it echoed in the work of the experts within this case study. Overall it seems fair to say that the status of Educational Science is lower than subjects in hard pure areas such as mathematics or natural sciences. To increase status, therefore, it is seen as desirable to engage in empirical work of a quantitative nature. Some have argued that this is also open to criticism as 'science, as we all know by now, is not nearly as clean, simple, and procedural as scientists would have us believe' (Kincheloe & Berry, 2004, p.1).

Let us now turn from the discipline of Educational Science to see how 'quality' within higher education, both overall and specifically within Germany, has come to be seen and has developed over time.

2.3 Notions of 'quality' in higher education

One could argue that Western society in the first decade of the 21st century has turned into an audit society where everyone and everything is being measured and assessed. Power (1994) was one of the first to address this in detail and asks why there should be this necessity for audit to occur, as society 'managed' before with less auditory control. One argument for this could be that we have seen an erosion of trust in society and in those working within higher education where public money is spent in such large quantities

(Norris, 1999). Society has also become aware that it *can* have a say in how its money is spent. The student 'consumer' wishes to have a learning experience with which he/she is satisfied and has become increasingly vocal about what he expects to experience and which knowledge he thinks he should have acquired by the end of his studies, especially if he is paying for his education in total or in part. The qualification at the end (outcome) should also be trusted and accepted by future employers. Control mechanisms that determine the quality of courses and the worth of departments/institutions have therefore become an integral part of the academic landscape. However, too much control can also be a negative factor and might inhibit the free expansion of intellectual creativity. Some forms of research, for example, do not necessarily have an outcome that will help the public good, so to speak, and could be perceived as being less worthy of funding. Areas of research might not be given the necessary financial support. Indeed, this could even be perceived as being anti-democratic (Becher & Trowler, 2001). Nonetheless, the examination of the quality of teaching and research has become an accepted norm.

Although implicit notions of quality have existed in higher education for centuries, these have developed more specifically throughout much of the 20th century and have moved with even greater pace into the 21st as higher education has, post-Bologna Process in particular, come to be perceived as an international marketable commodity (Green, 1994). However, quality is 'notoriously elusive of prescription, and no easier to describe and discuss than deliver in practice' (Gibson, 1986, p.128). Quality is a 'contested' notion (Barnett, 1992). Indeed, its contradictory nature has been addressed in a general sense: 'Quality… you know what it is, yet you don't know what it is' (Pirsig, 1974, p.163-164). Expanding on Vroeijnstijn's (1992) comparison of quality to love, Müller-Böling (1997, p. 90) tells us:

> Quality in the field of higher education is similar to love: you cannot grasp it, but it is nonetheless present; you can experience it, but it isn't quantifiable; it is always transient, you have to constantly and permanently engage with it.[5]

[5] Danach ist die Qualität der Hochschule der Liebe ähnlich:
* nicht fassbar, aber doch vorhanden,
* erlebbar, aber nicht quantifizierbar,

Furthermore, Pellert (2002, p.24) has indicated, in terms of how one can deal with quality, that it is not possible to manage it in a linear way, rather it is a 'multi-dimensional construct'. This is supported by the fact that excellent quality can exist in one part of an institution, but there can be poor quality elsewhere (Van Vught, 1997, p.51).

Definitions of quality have also changed over time. Schöne (1966) said that in terms of the locus of quality per se it 'can only be created in an intellectual field where there is *selection* (his italics). However, selection presupposes clear conditions and a selection by experienced and appointed judges'[6] (my translation). This is problematic as selection itself can be based on widely differing criteria and although quality can be said to exist, the quality of the judges will not necessarily be the same. Van Vught (1997) supports the view that 'quality' can have a different meaning depending on the country where the quality debate is taking place. He also reflects on two different traditions of universities within Europe: the British and the Continental. In the former, there has been less state control (now shifting towards more), whereas in the latter there was more (now shifting towards less).

Garvin (1984) reviewed definitions of quality as originating from within the disciplines of philosophy, economics, marketing and operations management. He presented five definitions. The *transcendent* definition was one that could not be clearly defined, but referred to 'innate excellence' (ibid., p.25). The *product-based* definition was by contrast specific. Here quality was clear and measurable depending on 'differences in the quantity of some ingredient or attribute' (ibid., p.26). The third definition was *user-based*, where quality depended on the person's needs regarding the product. This was inevitably a subjective view, as we might ask if personal satisfaction is synonymous with a superior product. Arguably it is not. The fourth definition of quality was *manufacturing-based*, in other words the extent to which a product met the requirements of conforming to a set

* stets flüchtig, so dass man sich immer erneut und dauerhaft um sie bemühen muss.

[6] Qualität bildet sich im geistigen Bereich nur dort, wo es eine Auslese gibt. Eine Auslese aber setzt überschaubare Verhältnisse und ein Auswählen durch erfahrene und dazu berufene Beurteiler voraus.

standard. Finally, Garvin presented a *value-based* approach where quality was defined in monetary terms, e.g. 'in terms of costs and prices'. Here, 'quality, which is a measure of excellence, is being equated with value, which is a measure of worth' (ibid., p.28), a notion of 'best' for the amount of money available, but not necessarily *the best*, as this was beyond what the pocket could cover. Garvin argued that depending on the viewpoint there might therefore be competing views of quality, but that this should be regarded as a positive rather than detrimental aspect.

Ball (1985) then introduced the notion of 'fitness for purpose', an idea which has certainly taken root in academia in terms of the appropriateness of courses, those selected to transmit knowledge and those assessing it (Harris-Huemmert, 2008). McClain et al. (1989) provided the idea of added value, a notion which has been adopted by schools in the UK in particular. They present what they can offer students over and above the usual curriculum. Harvey and Green (1993) expanded these ideas even further by creating notions of quality that included 'fitness for purpose' *and* 'transformation' (in addition to exceptional, perfect, and value for money). In other words, quality could only be achieved if the person experiencing it underwent an enduring change (acquisition of knowledge and understanding).

More recently Van Kemenade et al. (2008, p.175) have argued that these notions of what we think quality is 'are not sufficient to explain what is happening'. They have concluded that recent changes in values should lead us to rethink our understanding as 'values [being] the motor of our behaviour' (ibid., p.177). Notions of quality are in a constant state of flux and it has even been debated that trying to find a single definition of quality is a 'waste of time' (Vroeijnstijn, 1992, p.113), as what might be perceived to be the cutting edge one moment will be discarded the next according to the changing values of those concerned.

Van Kemenade et al. (ibid.) have simplified the notions of four value systems originally proposed by Beck and Cowan (1996) into value systems on quality and quality management: process control, continuous improvement, commitment, and breakthrough. Process control, they tell us, 'will not result in quality improvement, but in quality standardisation' (ibid., p.178). However, while this may be useful for products to ensure that they are consistently built, for example, there is a danger in this with regard to academia, for standardisation disallows for creativity and the opportunity to try out something new, especially where quality is not immediately assured. In the notion of continuous improvement their definition of quality 'is the extent to which the object exceeds the expectations of the customer' (ibid., p.179). We could therefore see the department of a low-ranking HEI being perceived by its students as 'quality' if it exceeds their expectations. In the third value system 'quality is the extent to which the goals of all stakeholders are fulfilled, taking into account [the] here and now and the future' (ibid., p.180). Their final and fourth value system, which they call 'breakthrough', pays tribute to an ever-changing environment, and they define it as 'innovation' (ibid., p.180). This is the extent to which goals will be fulfilled in the future which links in with Barnett's suggestion (1992, pp.47-51) that the best possible approach to quality assurance is a combination of both relativist and developmental concepts. The former includes the idea of fitness for purpose as a test for performance using both quantitative and qualitative methods; the latter can possibly be regarded as a bottom-up autonomous approach where staff members within institutions take responsibility for the quality of what they do.

I would like to contribute to the discourse by adding my own understanding of what I understand quality to involve.

1. Action: quality is concerned with what people do, the extent to which they do it and whether or not this doing takes in all of the facets of what it is that the doing involves. Accordingly, this *doing* needs defining, something which is notoriously difficult to pinpoint accurately in terms of teaching quality, for example.

2. Motives and values: quality involves the attitudes and values people have to what they do, the intrinsic *why* of what they are doing. A high degree of self-analysis and cooperation with others would seem the way forward here to ensure that everyone operates from a shared base of understanding.
3. Management: quality is about how people organise what they do.
4. Investment: this involves what people are prepared to invest in terms of training towards improving what they do, and also maintaining this level once they have achieved it. Arguably, this process can never be concluded.

Let us consider for a moment the task of anyone who has been asked to evaluate quality in higher education. The above gives us some idea of the problematic nature of this task. Before any evaluative work can be undertaken, evaluators need to reach agreement therefore on their own understanding of the term 'quality', not only in general, but particularly within the context of the field or discipline they are being asked to examine, and how they will be able to identify it and then rate it according to the context in which they are operating. They need to have a clear idea about their own values and their attitudes to different subjects within the discipline. They will also need to be aware of the changing status or paradigms of their discipline, as these are non-static. They will need to know the politics of the state in which they are operating, and the politics of the discipline within the university. The interview analysis of this study in Section Three of this thesis will show that there were different levels of understanding among the evaluators on all of these subjects.

2.4 Stakeholders in HE and attitudes towards accountability

Quality assurance in higher education involves a number of stakeholders, each with their own vested interests and values: from students to university staff, administrators, the institutions as a whole, governments and funding bodies (Barnett, 1992). Each has its own ideas of what it understands quality to be and these can differ from one institution to another and even be slightly different among sub-disciplines. The university environment

which has enjoyed and to a certain extent still enjoys levels of autonomy not found elsewhere has been increasingly influenced during the last decades by managerial concepts lent from the world of business (Kieser, 1999a; Nickel, 2005), to which some of the stakeholders within institutions have been rather averse to accommodate. One common theme, however, which can be identified throughout the world is that although standards may vary, institutions of higher education should at least be recording what they do so that audits can identify where money is being spent and where money might be better spent in the future.

Developing the idea of higher education in the late twentieth century, Barnett (1992) conceptualised two main groups within it, which he categorises as external and internal. Group 1 (external) is concerned with higher education as the production of qualified manpower; the starting point for a research career; the efficient management of teaching provision; and lastly, a field for extending life chances. Implicitly concerned with Group 1's understanding of higher education are policymakers, funding bodies and other interested parties. Group 2 (internal) focuses on the student. Here the role of higher education is to develop the student's personal qualities, his intellect, knowledge, autonomy, critical skills, character and ability to fit into society (echoing back to both Newman's and Humboldt's idea of developing culture). We can see that although these two groups are compatible in part, they do not always share the same aims in terms of what their understanding of the 'output' is. If this understanding is divergent and if we recall the number of different stakeholders involved, then any focus on quality and quality assurance will lead to a number of different interpretations and outcomes where it becomes impossible to create one notion of quality, or quality assurance.

Kogan (1986) has described three forms of accountability-reporting: state-controlled, which is reflected by the democratic commission assigned through the electorate; professionally-controlled, legitimised through the educators in higher education themselves; and lastly consumer-controlled, through those who pass through the HE

institutions. Altrichter et al. (1997, p.13) believe that the second type will remain the most successful model inasmuch as those directly involved with the transmission of the product (knowledge) should also be those most directly involved with the assessment of the success whereby this product is transmitted, therefore creating self-critical models. They also refer to the fact that personal motivation and autonomy are driving forces towards what individuals can achieve. Lonsdale (1990, p.101) has referred to performance management as:

> […] a process which integrates role clarification, goal setting, development, review and leadership. It is not a process of managing the performance of staff... [but] the process of managing so that job satisfaction, motivation and performance are enhanced… so as to create the conditions under which others can best work.

This may be a realistic notion, but it assumes that there will be those leaders within institutions who can create these conditions, which is not always the case, especially in those examples where differences of academic outlook form barriers and can cause bad will among the staff. However, this apart, whichever system is employed, quality assurance 'is slowly but steadily becoming an integrated part of higher education' Stensaker (2008, p.3). Although initially restricted to within the institutions themselves, the debate on quality has become public since what higher education has to offer is answerable not only to the student stakeholder in particular (in market language: the consumer), but also the taxpayer who is largely paying for it.

In the past academics were almost entirely autonomous to decide what they wished to teach or research. Problems with this system arose when academics either rested on their laurels (e.g. no production of new research or poor standards of teaching), or did not engage with any ongoing processes of self-control to examine what they were doing in their lectures/seminars that worked well, and see what might need improving. A European initiative attempted to introduce standards for how quality in academia should be managed. This was called the 'Excellence Model' of the European Foundation for Quality Management (EFQM). However, it came as no surprise that it was rejected by academics as being too control-oriented.

A specific system of quality control from the business world was adopted into academia. Van Kemenade et al. (2008, p.10) have argued that the growing popularity of the ISO9000:2000 model[7] is possibly due to the fact that it 'incorporates the customer and has moved from control in the direction of continuous improvement'. Numerous countries have included this as a means of raising standards in particular areas of quality assurance.[8] This conforms with Altrichter's above-mentioned idea of professional control in universities being the key to successful methods of quality assurance. However, as Stock (2004, p.36) says: 'In research and teaching there are simply no guarantees of success, no calculable rules of transformation of an 'input' into an 'output'. What has worked well in the lecture theatre for a few years might still fail with a particular group of students. Nickel (2005), too, has explored the ways in which management structures from outside the field of higher education can be incorporated into this very different environment. She comes to the conclusion that we should be very wary of imposing such rigid management concepts on higher education, as this field warrants its own special structures.

It is not very straightforward to superimpose one system of quality control from one environment on to another, especially in one where success or quality cannot always be measured accurately and where personal autonomy is largely regarded as a given. The quality of a department cannot be found merely in the number of graduates passing their final examinations (to quote one frequently cited example), nor can quality be measured only in terms of publication numbers or references in academic journals. Ory (2000, p.17) also states that 'Assessment is more than counting, measuring, recording, or accounting. […] It touches on self-reflection, dialogue, and discussion.' Gregory (1996) refers to the list of criteria developed by Elton & Partington (1991) that should be employed in the assessment of teaching in higher education, for example, which includes aspects such as

[7] ISO 9000 refers to a set of quality management standards. The ISO (International Standards Organisation) was developed in Switzerland to provide international standards that could be applied in business across the globe. Many of these standards are applicable within the context of higher education and have been adapted by some institutions. See Stockmann (2006b, p.39) for a comparison of the ISO model and TQM (Total Quality Management). See also Vouzas & Gotzmani (2005) for information on best practice in ISO9000.

[8] see Magd & Curry (2003) for an example from Egypt, or the website below for the Taiwanese example: http://www.icdf.org.tw/web_pub/20080619084537%E7%AC%AC%E5%9B%9B%E7%AB%A02.pdf

innovation and evaluation of one's own teaching. Johnes & Rothschild (1975, p.17) were early proponents who argued against a metrics-based system in the quality debate, stating that 'we ought to get away from the idea of assessment in the sense of ranking or grading. We should aim to get to the point where colleagues look at it as the normal thing to assist one another.' To this day, however, staff members dislike the notion of letting others come into their lectures with a view to discussing the merit of their work. As Barnett (1988) reminds us, internal rivalries can work against any quality improvement, a situation which the experts in the case study evaluation were witness to. It would seem then that quality incorporates far more factors such as relationships as they exist between individual faculty members, student relationships, their attitudes, expectations, and life circumstances, ratios of staff to student, but also the manner in which networks lead to an exchange of existing knowledge, thereby creating new channels of thought.

Some countries have organisations which monitor performance in terms of research (e.g. the UK's RAE – the Research Assessment Exercise) in which very clear guidelines on what is expected of various levels of 'quality' are provided: 'The RAE is the principal means by which institutions assure themselves of the quality of the research undertaken in the HE sector'[9]. However, even quality-determining exercises such as the RAE are subject to modifications and improvements. Since the last RAE in 2001, a number of changes have been implemented[10], one of which has been a reduction in punitive measures which inevitably caused serious funding issues for those institutions ranked in the lower half of those measured. The RAE explored the idea of replacing the peer review process by metric assessment, but this has recently been overthrown as a pilot study revealed that a predominantly metrics-based system did not work.

Comparisons across the world reveal that university performance/quality measuring mechanisms are now as diverse as the university landscape itself, and to some extent this is influenced by the political structure and culture of the country in question. Germany is a

[9] See http://www.rae.ac.uk/aboutus/history.asp
[10] See http://www.rae.ac.uk/aboutus/changes.asp

federal country, which immediately suggests that we might find a diverse landscape in terms of quality assurance in higher education and indeed, this is the case (see Harris-Huemmert, 2006; Orr & Pätzold, 2006). Before we examine how quality assurance is organised, let us look at the development of evaluation in the higher education arena in its entirety.

2.5 Definitions of evaluation

Evaluation has existed for a long time and has been defined in many ways. Reflecting on Scriven's early work on the formative and summative aspects associated with the methodology of evaluation, Stake (1975) defined evaluation as 'an observed value compared to some standard.' He raised the valid point that any evaluation will leave 'some things de-emphasized', and so no evaluation process will be able to do justice to everything being observed. There can never be a perfect evaluation. Guba and Lincoln (1981) defined evaluation as 'the systematic investigation of an evaluand's worth or merit'. Although this definition has been used extensively and is a useful guideline, it still leaves open the question of *how* we measure the worth or merit of something accurately and according to which standards and values. Fournier (2005, pp.139-140) has more recently described evaluation as an 'applied inquiry process for collecting and synthesizing evidence that culminates in conclusions about the state of affairs, value, merit, worth, significance, or quality of a program, person, policy, proposal, or plan'. This is the definition I prefer as it is indicative of the breadth and complexity of evaluation and does not disguise the problematic nature of the task in hand.

As a consequence of this breadth the literature provides us with numerous evaluation theories and models and also provides examples of practice (notably Rossi, Lipsey & Freeman (2004), Fetterman, Kaftarian & Wandersman (eds.) (1996), Guba & Lincoln (1981), Stake (1975) and Stufflebeam (2001). Evaluation practice has also been described within the German field (Kromrey (2001a), Beywl (1988, 2003), Spiel (2001), Stock (2004) and Stockmann (2004). Keiner (ed.) (2001) offered a first synopsis of the

developments that had taken place in evaluations in German Educational Science at the start of the 21st century. Struhkamp (2005) then provided an overview of the development of evaluation overall in Germany, although this was a general investigation and not specific to the field of higher education.

Over and above the actual theory of how to conduct evaluation, those wishing to work as evaluators have found useful guidance in the works compiled by Law & Bronson (1977): one for trainers of evaluators and the other for evaluators themselves. However, these were not specific to working as an evaluator in higher education. The Middle States Commission on Higher Education has also provided a number of documents pertinent to evaluators, notably the *Handbook for Evaluation Teams* (1974/2000), which is now in its sixth edition. This remains a standard work in the field (see http://www.msache.org). Patton (1982, 1987), too, has also provided guidance manuals on how to conduct evaluations. Patton has been notable in singling out the importance of the utility of evaluation and his works emphasise this aspect. Minelli et al. (2008) have also noted in their comparison of evaluations in Italian and Dutch universities that little use has been made of evaluation results.

Alkin et al. (1979) undertook an examination of five different evaluations within schools in order to glean information on evaluation utilization. Their research was spread across two years and they used non-structured interviews and literature review as part of their naturalistic enquiry. Schools were anonymised, as were the identities of their interview subjects. Although the focus of their research was different from mine, they throw some light on the ways evaluators perceive their work, and how they in turn are perceived. However, this aspect is secondary to their main research focus.

In terms of examining the types of people who work as evaluators and based on research originally undertaken by Meltsner (1976), Thompson & King (1981) expanded on the idea of evaluator types, which Meltsner had originally classified into four categories:

entrepreneur, pretender, technicians, and politicians. They looked at how school evaluators were perceived and whether these categories were appropriate. Guba and Lincoln (1981) were also concerned with this idea and developed Douglas's (1976) type model to include the following: 'sociability' (or entrée); 'quarterback'; bagman; observer; literary expert; and grant finder. Although it is not my objective to propose new evaluator types, I will reflect in Chapter 8 on whether I see similarities or differences in the evaluators in the present study.

Dressel (1982) examined evaluation work from another angle and concluded that decisions are influenced by the institutions into which evaluators go as they cannot divorce themselves completely from the values of the institution. Where Dressel's examination is small-scale, Silvers's (1982) large-scale analysis of the relationship between evaluators and their commissioning bodies reaches similar conclusions. Csizmadia, Enders & Westerheijden (2008) have taken this approach further in their investigation of how quality management (in Hungary) has been influenced by governmental policy. The present study (see Chapter 4) will also take into account the political influences from the state of Baden-Württemberg, which impacted on the way in which the evaluation was conducted.

Moving on to the question of how the evaluators might think, Schön (1983) has provided a comprehensive view of how professionals think in action, based on observations of a variety of professional areas. However, he is the first to agree that we need to inquire into the epistemology of practice. He raises questions such as: 'What is the kind of knowing in which competent practitioners engage? […] In what sense, if any, is there intellectual rigor in professional practice?' (p.viii).

Evaluators of higher education are predominantly those members of their field drawn from the upper echelons, who usually have extensive experience of management positions, and should therefore be regarded as 'elites', although they qualify equally

through their knowledge as experts. Brint (1994, p.131) provides a useful distinction between the terms 'elite' and 'expert'. He defines elites as 'those people occupying the "command posts" at the top of the central institutional domains in the society', whereas experts are 'highly trained professional staff working in these central institutional domains'. I believe that in the present case study we see a combination of 'elite' and 'expert' evaluators in action.

Morley (2003) has addressed what happens in situations in which external peers coming into institutions were *not* the equals of those whom they were examining, in other words, the evaluators were of lower status than those they were assessing. This raises the question of peer suitability and selection processes. She highlights the instability of quality assessment, depending on who is doing it and when.

Continuing the theme of the 'ideal evaluator', Stake (2004) presents detailed information on what an evaluator should be like, i.e. his personal qualities, aims, background etc., however, these remain suggestions only. The present study will therefore examine the evaluators' notions of which qualities they believe they should ideally have when working in this capacity. Evans (1999, p.147) has also discussed the roles of academics within peer reviews and concludes: […] 'peer review is manifestly not an exact science. […] Even referees in the same field and doing their best to be fair will often be in radical disagreement'. She continues: […] 'the same individual and the same body of work can be rated high or low with equal plausibility and an outcome can always be justified with a sequence of reasoning which is hard to fault.' (ibid., p.165) However, she does not offer any ideas on how this problem can be overcome.

Kalkwijk (1992), Vroeijenstijn (1992) and Kieser (1998) have provided insights into the workings of the Dutch evaluation model, which I describe in greater detail towards the end of this chapter, as this was the chosen method of evaluation in German higher education. They point to three problems with regard to peer review: firstly, not all areas

are covered equally well by everyone; secondly, subjectivity against certain theoretical approaches is apparent and thirdly, the term 'quality' itself is hard to quantify and is interpreted differently[11]. However, these suggestions are not supported with examples from the field.

Let us now examine more closely some of the considerations involved in evaluation practice within German higher education as this may be of consequence to those conducting evaluatory work there and the type of person involved.

2.6 The habitus of the evaluator

In Germany it has become standard practice for external experts to be brought in to evaluation commissions and at least one foreign expert will usually participate. Furthermore, experts are drawn from beyond the borders of the particular *Bundesland* to be examined (Tippelt, 2004). The idea here is that they will not be affiliated to the institution under examination and therefore be unbiased, although in reality some disciplines (or sub-disciplines) are so small that an entirely neutral stance cannot be adopted. Arguably, there is no such thing as an entirely neutral stance in any case, as evaluators may already be influenced in some way, albeit subconsciously. The evaluators are the peers of those undergoing the evaluation, but through their contacts with colleagues in previous work contexts or through conferences a certain amount of prior knowledge of the departmental and collegiate structures concerned may therefore be brought to the evaluation. As Harman (1998, p.354) stated a decade ago, evaluators bring 'outside values and constructs' into each evaluation and these need unpicking before consensus on evaluation criteria can be reached, which should be adjusted to the needs of each institution (Bewyl, 2003). It seems necessary for the evaluators themselves to gain some insight into the *habitus* and values of fellow evaluators before they can engage in joint analysis of the culture (field) of the institution they have been called upon to assess (cf. Bourdieu, 1988). Of course, how much each individual is willing or able to be

[11] See http://www.forschung-und-lehre.de/pdf/08_98.pdf

reflective will vary in each case. Each evaluator comes to his work with a history, an individual culture, a set of values. If we understand the mechanisms of how evaluators work and the criteria they agree upon and against which they wish to evaluate a department (which may vary and will also be dependent on the commissioning body), we may be able to discover general trends in their attitudes and work patterns that could be addressed from the outset, thereby easing the initial introductory phase and improving evaluation processes.

Experts on commissions usually cover between them all areas that will form part of the analysis. They are thus chosen for having particular knowledge. It seems also relevant to investigate the selection criteria according to which one expert and not another was asked to participate, as this might be of influence on the outcome of the evaluation. Furthermore, the functioning of the group can be hindered if, for example, one of the members decides to dominate proceedings. More junior experts could find it difficult to 'find a voice' under such circumstances. Age could play a part here, with older academics throwing more weight because of their seniority.

Another question is the extent of the contribution of those experts who have hitherto not been involved in evaluation work. They might find it necessary to adopt an observatory stance before becoming confident enough to take a more active role. Considering that the majority (perhaps all) are full-time academics with usual lecturing/research/teaching tasks, to say nothing of the conferences and guest lectures that they usually attend, what motivates them to become an evaluator? The task places a great additional burden on the individual and the financial rewards are not that great. In Chapter 5 I will show some of the underlying decisions to become an evaluator and the selection processes, as it seems relevant to trace the procedures that precede the selection of an evaluation commission.

Volkov and Volkova (2005) have described the work of professional program evaluators coming from a variety of specializations. Their approach was ethnographic and

exploratory, and their sample limited to 13 interviews with evaluators and five with stakeholders. Their work goes some way to 'inspecting the inspectors' as Alderman and Brown, 2005, p.313) says should be done. However, they make no claims of being able to generalize findings to the entire population of program evaluators (Volkov and Volkova ibid., p.2). Unfortunately, their own research interests have since moved in a new direction, and so it is not possible to engage in a discourse with them.

2.7 Professors within the dichotomy of egalitarianism and elitism

If we compare the implementation of systems of quality assurance in higher education internationally, it becomes apparent that in contrast to some countries (notably the US and the UK), the mandarin culture apparent in German academia was averse to any form of quality control (Enders, 2001), although this does not mean that mechanisms of quality control did not exist (Struhkamp, 2005) or that academics elsewhere embraced quality assurance with open arms. We see a dichotomy here, as on the one hand there was the notion of the individual expert professor who was conducting excellent research and teaching (an idea which pointed towards a notion of differentiation from one institution to the next), whereas the overall idea in Germany was that all of its HEIs were egalitarian and able to offer the *same* standards. While some universities that were founded in the Middle Ages were obviously well-known (e.g. Heidelberg or Tübingen), essentially each institution was believed to be able to offer the same services/standards in an egalitarian sense (Frackmann, 1992). Apart from the few *Numerus clausus* subjects (e.g. medicine, which required certain entry grades at *Abitur* level), the German *Abitur* per se was regarded as the entrance qualification to higher education. Even the subjects taken were not necessarily seen as a hindrance to embarking on an entirely different subject route and courses could be repeated until the student had acquired sufficient skill to progress. This system was therefore very different from the Anglo-Saxon one in which the grades achieved at A-level were held (and still are held) to be markers of a student's future level of achievement at degree level.

However, in light of the German Research Foundation (DFG) competition for 'excellence' status (officially introduced on 25 June 2005), universities in Germany are becoming more diverse as clusters of 'quality' or excellence are emerging. Altrichter et al. (1997) had correctly predicted that universities would become more heterogenous in the services they provide following the dawn of competition and creation of university profiling.

Ringer (1969) argued that the German mandarins had declined by 1933, as their values were challenged by a changing social and cultural environment, but as Teichler (2005, pp.16-17) stated more recently, it was the professor, and not the university, who was the central unit of quality in Germany. Professors in Germany were able to influence the contents of what was taught through their own personal preference and the whole department could therefore benefit or suffer accordingly. Frackmann (1992, p.51) observed that 'the individual (eminent) researcher is the core of the German research university concept', rather than the idea of a group of top-class researchers in one university or even collaborations of excellence across institutions. However, Glotz (1996, p.39) realised towards the end of the last century that 'reality is cheeky enough to disregard faculty borders; many [German] professors are not'[12], indicating that many were unable to look beyond what is directly in front of them, let alone form collaborative efforts aiming to improve overall quality. Künzel (1997) commented on the fact that leadership needed professionalizing in the academic context. Evaluation as an idea had to assert itself against this backdrop of the lone 'eminent' professor. Evaluation systems slowly started to appear in Germany from the mid-1990s onwards (Federkeil, 2004; Simon, 2003). With a generation of the German academic profession moving into retirement and an increase in academic mobility, in particular to the US, where the experience of evaluation and assessment practice is commonplace, a change in university culture may be taking place with a possibly reluctant acceptance of quality control as a given.

[12] Die Realität ist frech genug, sich über Fakultätsgrenzen hinwegzusetzen; viele Professoren sind es nicht.

The landscape has therefore changed as young academics spend terms of employment abroad where they experience different systems of quality assurance and develop possibly more self-critical stances towards their own actions within the academy. However, the scenario in Germany of the chair being the basis for decisions on research and teaching to this day nonetheless remains 'the essential characteristic of the university' (Huisman, 2003, p.39). Perhaps the notion of power which accompanies this hierarchical system is simply too attractive to young academics to dispense with entirely once they themselves come within grasp of such pinnacles.

If an innate embracing of a pan-university evaluation culture is to be achieved, the senior leadership within the academy (e.g. each individual professor) needs to encourage his team to engage sufficiently in self-reflection to discover where strengths and weaknesses lie. Over and above this, the university leadership *per se* needs the vision of and a commitment to high quality for this to become an inculcated factor within his institution. There is a strong argument that this is not always the case. Davies (2000, p.12) argues that the culture of an institution has to be conducive to enable change to occur and this has to be supported by the leadership, which will involve 'managing strategically in different cultural settings.' A sense of 'trust and respect' has to be present to enable dialogue. To some extent the hierarchical structure of German higher education could be regarded as working against establishing a culture of 'trust and respect' where dialogue between professors within one institution is commonplace. In some institutions it is obvious that internal dialogue among colleagues is antagonistic, as the experts in this commission were unfortunately witness to on at least one occasion (see Chapter 7). Goetze (2005), who provided a retrospective on his own role as an evaluator over time, has argued that while evaluations have had a positive effect, they are not used by all. This harks back to Patton's notions of utility where an evaluation report can either instigate a positive transformation or be redundant.

2.8 Influence of Dutch evaluation methodology in German HE

One major influence that had already taken partial root within German academia during the late 1990s was the evaluation system which was developed by the VSNU in The Netherlands (Association of Universities in The Netherlands) during the 1980s[13]. In turn, the VSNU had gained its inspiration from evaluation research that was making progress in the United States during the same period. Germany was fortunate to find a model of evaluation on its doorstep, and so by the late 1990s the Dutch evaluation model had become the evaluation model of choice. Briefly, this comprises the following stages:

- commission established
- external experts chosen to undertake evaluation
- institutions provide self report according to particular criteria
- experts conduct on-site visits (usually of one or two day duration) and report first findings
- final report published
- follow-up.

The evaluation usually examines strengths and weaknesses with a view to encouraging self-reflection on quality within the institution in order to achieve higher standards of learning and teaching in the future. In the knowledge that what they say can have an impact (albeit indirectly) on policy, evaluators are careful about the extent of their criticism within the final report. Kohler (2003) has also pointed out that evaluations can have an influence on the existence of courses and funding policies, and so evaluators generally deliberate over the precise wording of documents before releasing their report.

In the first part of the 21st century, the Dutch evaluation system was emulated and those experts who had spent time in The Netherlands working as evaluators were regarded as particularly valuable members of any German evaluation commission. Apart from those members who had worked in The Netherlands at the VSNU, in the case study evaluation none of the experts had been given training in evaluation theory, which supports Szanto's

[13] See http://www.tkk.fi/Misc/H3E/wg2/NLh33.pdf for brief description of the Dutch model.

(2005, p.189) view that while most of the evaluators are disciplinary experts, they 'do not have a quality assurance background and the knowledge that is necessary for a thorough review.' This remains a challenge to the organisers of evaluations, as those they recruit are subject experts, but not necessarily evaluation experts, who require a different set of skills.

2.9 Analysis of German HE evaluation landscape

Considering that the evaluation landscape in Germany was still in its infancy at the turn of the last century and is developing constantly, it is not surprising that at present there is little literature available to us on the subject. Altrichter et al. (eds.) (1997) were among the first to discuss the consequences of evaluations and the influence of the state with regard to university accountability. Mittag et al. (2003) compared and contrasted two of Germany's first higher education evaluation systems: the *Nordverbund* in North Rhine-Westphalia, and the *ZEvA*, the Zentrale Evaluations- und Akkreditierungsagentur in Lower Saxony, although this work deals predominantly with the theme of utilization. In a more recent work, Mittag (2006) refers to the current controversy raging on the validity of evaluation work. She looks at the outcomes of evaluation work in the *Nordverbund* and the *ZEvA*. Although her research touches on the role of the evaluator, her main research focus remains the utilization of evaluation in light of the present debate. Nonetheless, her work provides a useful point of reference to evaluations, albeit in other German states. While much work is currently being undertaken by the evaluation agencies in Germany (Teichler & Tippelt, 2005), research remains fairly meagre on the inside perspective of the evaluation process itself, i.e. the human dynamics of evaluation commissions and reactions to evaluations at base level. However, Stockmann (2006a, 2006b; 2007) has provided some comprehensive literature on evaluations in the German context, including a handbook for evaluators to use on site[14]. Klieme & Tippelt (eds.) (2008) have also reviewed the overall status of quality assurance in German education.

[14] Perhaps it is of note that Stockmann developed the first Masters course in Evaluation in Germany at the University of Saarbrücken. More recently, the University of Bonn has launched a similar course.

Hardly any research exists on the work of evaluators in Germany, or Baden-Württemberg in particular. During the first few years of its operation, the evaluations within this state were organised by the *evalag* agency to evaluate all of the faculties undertaking teaching and research in one discipline at a time, which meant that a fairly large group of institutions underwent the process together. Choosing to study the evaluation of *Erziehungswissenschaft* in 2003/2004 has a further advantage in that it was commissioned by the Ministry of Science, Research and Art in Baden-Württemberg, which therefore added the perspective of highlighting possible political influence in the overall evaluative task. The report of this evaluation commission is also in the public domain, so access to both the *evalag* agency and the Ministry was possibly easier than in other states. In other parts of Germany where evaluation reports are not publicly available, this kind of research might not have been possible. Although the present study is an investigation of one evaluation, it included 18 experts and one permanent ministerial representative, evaluating 15 HEIs. This can therefore be viewed as a rather large evaluation.

Alongside evaluation processes and with the introduction of modular degree courses which lead towards BAs and Masters degrees we can also identify accreditation procedures, which are required before new degree programmes can be officially endorsed. Most members of staff are therefore involved in quality control mechanisms and are implicitly aware that in an increasingly global academic market they should be accountable for what they do. However, in the sense that a national system of quality assurance is apparent, as is now the case in China (Zhou, 2005) and exists in the UK with the Quality Assurance Agency (QAA) and the Research Assessment Exercise (RAE), Germany will, for the foreseeable future, not have an over-arching system of quality assurance. Today we can see challenges not dissimilar to some of those faced by the mandarins in the 1930s: accountability, management structures in universities, and the global education market have forced professors and faculty leaders to rethink their own positions.

2.10 Political factors in association with German HE evaluation

Over and above any restrictions that may exist due to the traditional mind-set of academia, Germany is a federally-organised country. Each of its sixteen states is autonomous and can decide independently over its higher education policies, although the Standing Conference of University Chancellors (*Hochschulrektorenkonferenz* - HRK) and Standing Conference of the Ministers of Education and Cultural Affairs of the *Länder* of the Federal Republic of Germany (*Kultusministerkonferenz* - KMK) are two bodies that regularly meet to decide on policies for higher education. Turning back to July 1995, the HRK had merely agreed that evaluations with reference to teaching should take place. Agreement was also reached on establishing an over-arching body: the Accreditation Council (*Akkreditierungsrat*), which was founded in 1998 and which officially started accrediting degree programmes in 2005. This body is also responsible for the maintenance of standards within accreditation agencies, according to the standards set down by the European Association for Quality Assurance in Higher Education (ENQA 2005, §3.8). The Accreditation Council itself was subject to evaluation in 2001[15]. While evaluation agencies are themselves free to choose their own methodologies, it appears that most adhere to these standards. Many also turn to the standards developed by the German Society for Evaluation (Deutsche Gesellschaft für Evaluation – DeGEval)[16]. One easily identifiable problem is that because there is no standard system as such there can be issues with the rigour of method (Kieser, 1999).

In a move to encourage greater autonomy among HEIs, in June 2007 the KMK decided to allow within-institution quality assurance systems to be accredited, according to the standards put forward by the Accreditation Council *(Akkreditierungsrat)* which were published in October 2007. System accreditation (*Systemakkreditierung),* as this method is termed, is still in its infancy, as universities explore ways in which they can implement their own systems of quality assurance which are capable of being accredited by the

[15] See http://akkreditierungsrat.de/KMK_Evaluationsbericht.pdf
[16] See http://www.degeval.de/calimero/tools/proxy.php?id=72 for standards printed in English.

Akkreditierungsrat (see also Chapter 9). However, as Witte (2007, pp.5-6) suggests, the HEIs face numerous associated problems, which she lists as follows:

1. It needs to be acknowledged that quality assurance costs resources
2. HEIs need to develop quality management systems and a quality culture
3. The formative role of QA needs to be strengthened
4. The task distribution among actors needs to be clarified
5. The institutional interest in lean institutional quality management needs to be balanced with students' interest in transparent information on programme quality
6. The tensions between a competitive accreditation system and common standards need to be addressed
7. Curricular diversity and comparability need to be balanced

While degree accreditation processes and quality assurance mechanisms which serve this purpose have been implemented, they have been appearing concurrently in different states at different times, according to varying needs and budgets. To a large extent this process remains in flux.

Due to the need for degree accreditation, the process of evaluation as a concurrent method of quality enhancement has come under fire as university staff and administrations struggle beneath the burden of compiling reports and self-evaluations. As Hanft (2003, p.10) noted: 'evaluations are negated and denied, or else placed in the foreground if they prove useful in power politics'[17]. Power politics reveal that universities can veto participation in evaluations if they cannot find the time to engage in them, a scenario Winter (2002) had already prophesied. Although most HEIs have accepted external evaluation as a useful instrument, as evaluation cannot be enforced, universities sometimes have a lobby strong enough to put evaluations on hold if they feel this is appropriate. Indeed, the universities in Baden-Württemberg rebelled against the governance structure of the *evalag* agency, which they regarded as being too biased to

[17] Evaluationen werden negiert und verschwiegen, oder aber in den Vordergrund gerückt, wenn sich dies im Machtpoker als nützlich erweist.

allow for truly objective and neutral evaluations, and so the agency was forced to restructure itself. This is described in greater detail in Chapter 9.

2.11 Influence of PISA and the Bologna Process on German higher education

In addition to the influence of the Dutch evaluation system, the introduction of systems of quality assurance in Germany was accelerated in the wake of the first stage of the PISA studies[18] when German schools did not perform as well as peers in other countries and the spotlight was immediately thrown on Germany's education system (Gruber, 2006). Ministries and a host of other stakeholders started calling for more information on what was going on within universities and schools, with the aim of improving standards, changing curricula and altering teacher training to ensure better standards of education. So, while top-down influence was apparent, universities themselves became interested in discovering their strengths and weaknesses (bottom-up). There have even been rare instances where institutions themselves have called for an evaluation of the status quo. The teacher training college (*Pädagogische Hochschule* – PH) in Weingarten, Baden-Württemberg, is a case in point and was externally evaluated in 2005[19].

In addition to the effects of PISA, German higher education came under an even greater challenge following the Bologna Process[20]. The University Reform Law of 1998 (*Hochschulrahmengesetz* - HRG) gave German universities the legal right to start offering Bachelor and Master degrees; however, the take-up pre-Bologna was insignificant and indicative of initial scepticism towards the new degree forms (Huisman, 2003, p.20). German degrees thus usually culminated in either the *Magister*, *Diplom* or *Staatsexame*n qualifications which, although internationally recognised, were incompatible with the Anglo-Saxon three-tier degree structure, making educational mobility in the European zone rather difficult.

[18] See http://www.pisa.oecd.org
[19] See http://www.evalag.de/pdf/Kurzf/Summary_WG_Gutachten.pdf
[20] See http://www.bologna-berlin2003.de/pdf/bologna_declaration.pdf

Although the talks at Bologna did not stipulate that the German system should reform and accommodate to the three-degree structure, German academia started moving swiftly in this direction, and the majority of institutions of higher education are now engaged in this difficult process of reform with a view that by 2010 the process would be completed. This poses immense time and workload challenges as all degree programmes need accrediting before they can be implemented. Evaluation, while still in its infancy in some parts of the country, finds itself being temporarily brushed aside as institutions struggle to accredit course programmes, although the majority recognise that evaluations are fundamentally necessary and useful.

2.12 Evaluation landscape in Germany

Each federal *Land* is able to decide over curricula and budgets and has autonomy over how its quality control systems are set up and maintained. Unsurprisingly, therefore, a patchwork picture within the evaluation landscape has emerged, revealing that some of the earliest evaluation operations have already ceased activity (see LEU; *Geschäftsstelle Evaluation der Fachhochschulen in Nordrheinwestfalen* in the table below), whereas others have only recently been introduced. I have organised the layout of this diverse landscape into three broad groups (see Table 1 overleaf). Within the first group are evaluation agencies that were established for individual states. The second group consists of alliances between individual universities across state borders. The third comprises individual evaluations by single institutions and groups of states that cooperate in particular. As Schade (2004) indicates, since individual states have autonomy over their evaluation systems, it is difficult to distinguish comparisons as to what each federal state is doing and how, considering that within some of the larger states, notably North Rhine-Westphalia, a number of universities cooperate with others outside the state, whereas the majority within the state conduct joint evaluation.

Evaluations organised by individual states	
Bavaria	conducts its own evaluations internally, private
Baden-Württemberg	since 2001; evaluations conducted by *evalag* agency in Mannheim; public
Lower Saxony	since 1995; evaluations conducted by the *Zentrale Evaluations- und Akkreditierungsagentur – ZEvA*; public
North Rhine-Westphalia	from 1997-2002; *Geschäftsstelle für Evaluation der Universitäten - GEU*, private from 1997-2006 *Geschäftsstelle Evaluation der Fachhochschulen in Nordrheinwestfalen,* public, suspended operation
Hesse	individual university evaluations, reports sent to Ministry

Inter-state alliances	
Verbund Norddeutscher Universitäten	since 1994; evaluations conducted by seven north-German universities across five federal states; public/private
Lehrevaluation in der Universitätspartnerschaft Halle-Wittenberg, Jena, Leipzig	LEU - from 1999-2005; private, evaluations suspended
ZQ + Hochschul-evaluierungsverband Südwest	since 2003. Evaluations conducted according to 'Mainzer Model' in range of universities. Most reports are private
Länderübergreifende Verbund der TU Darmstadt, der Universitäten Kaiserslautern und Karlsruhe unter der Moderation der ETH Zürich	since 1999, public

Individual universities and special state co-operations	
EQ	since 2005; *Institut für Evaluation und Qualitätsentwicklung* at the University of Lüneburg, public
ENWISS	since 2001; evaluations conducted by the *Evaluationsnetzwerk Wissenschaft* in the states of Hesse; Rhineland Palatinate, Saarland and Thuringia, public

Table 1: Organisation of evaluations throughout Germany

The Berlin Communiqué of 19 September 2003 stated that 'by 2005 national quality assurance systems should include the following: […]

- evaluation of programme or institution, including internal assessment, external review, participation of students and the publication of results'[21]. However, Table 1

[21] See http://www.bologna-berlin2003.de/pdf/Communique1.pdf

reveals that this has not been implemented in its entirety throughout Germany to date (cf. also Mittag et al., 2003, p.10)[22]:

2.13 Evaluation agencies and the significance of launch date

To provide a sense of the stages at which any one group or alliance and with it the quality assurance within certain parts of Germany were standing in 2003/2004, two contrasting examples are presented. The first is taken from the evaluation agency in the state of Lower Saxony, the *Zentrale Evaluations- und Akkreditierungsagentur Hannover* – ZEvA, which introduced evaluations in 1995 [http://zeva.uni-hannover.de]. Many evaluations in this state are already in their first follow-up phase. Lessons learnt from the practice of the first have been incorporated into the second, thereby improving the quality of evaluation work. The ZEvA changed its own commission to one of institutional audit (May 2007). Instead of involving large numbers of subject experts, these audits will now consist of only five members: two chosen as subject experts with track records of leading an institution, one quality assurance expert, one representative from the field (business/industry etc.) and one student. The audit aims to identify strengths and weaknesses of internal quality assurance, thereby uncovering latent potential that can improve and develop the self-regulation of an institution. At present, this model is predominantly used to examine the quality of teaching and learning. In the future it will also address quality assurance in research. ZEvA has itself been recently accredited by the *Akkreditierungsrat* and will be therefore able in its own right to accredit system accreditation systems in other HEIs. Experiences from this agency have also been incorporated with those of another, the *Nordverbund*. Research conducted by Mittag et al. (2003) pooled information gleaned from these two agencies with the view to improving the quality of evaluations in general, and so there is a certain amount of synergy visible. It is beyond the scope of this study to provide any conclusive evidence as to whether standards have improved as a result of evaluation, although examination of recent university rankings might provide a first indication.

[22] See http://evanet.his.de/evanet/institutionen/evaluation.php

The second example is taken from the *evalag* agency which was launched in the state of Baden-Württemberg in 2000, starting active evaluations in 2001 [http://www.evalag.de]. The following section will therefore introduce *evalag*.

2.14 History of the *evalag* agency in Baden-Württemberg

The name of this agency is derived from an abbreviation of the word *Eval*uations*ag*entur (evaluation agency) and it can be located in the first group mentioned above. It was established as a public foundation to conduct evaluations on behalf of the universities, technical universities and teacher training colleges in Baden-Württemberg, which is one of Germany's largest states. It consists of a permanent Board of Governors of 11 members who are responsible for deciding which evaluation request will be accepted and when, and they are responsible for the appointment of experts.

From its inception it was the agency's task to undertake 'regular, comparative evaluation of teaching and research in the higher education system of the federal state' (*evalag* information document, April 2002). *Evalag* stated that the purposes of quality assessment were:

* to ensure that the public funding provided is supporting education of excellent quality
* to encourage improvement of standards and performance in teaching and learning at institutional, faculty, department and subject level
* to provide information for students and the general public on that education through the publication of reports
* to create transparency in the HE sector through standardised assessments, and
* to provide both HEIs and the government with data for strategic decisions.

The main principles of the assessment method were:

* regular intervals of assessment
* to combine internal and external procedures of review
* involvement of students in the process, and
* publication of results.

Evalag compiled a standard self-appraisal document to be used by institutions during evaluations. This could nonetheless be modified according to the needs of the discipline under examination.

A review of Germany's higher education evaluation agencies reveals that they tend not to specify what their understanding of 'quality' is or expand on some of the areas as expressed above. Agencies expect stakeholders to know implicitly what quality is. However, as Pirsig (1974, p.164) points out: 'But if you can't say what quality is, how do you know what it is, or how do you know that it exists?' Although quality is not determined specifically, the standards of evaluation as used by the DeGEval, the German Society for Evaluation, and the ENQA, as already stated, are guidelines that most agencies conform by.

As each *Land* and each university is autonomous, there is little need for consensus across *Länder* borders on course contents or degree structures as each state has its own concepts. This leaves issues pertaining to quality assurance rather unclear. An analysis of *evalag's* published evaluation reports shows that institutions were often unfocussed about what they should be doing to achieve high standards, which is an indicator of ongoing uncertainty about national and international benchmarks for disciplines. Those being evaluated lacked sufficient professionalism to address the process offensively (Merkens, 2004, p.20). Therefore, it seems valid to ask to what extent the experts themselves know what quality is, as there appears to be a considerable lack of clarity or consensus on this point throughout the country. This will be addressed in reference to the case study in the present study.

2.15 Evalag's evaluation methodology

During the first period of its operations (2001-2005) the evaluation commissions at *evalag* were usually quite large and it was not uncommon to find twice as many evaluators taking part as the number of institutions being examined (Harris-Huemmert, 2007b). Evaluators

were sought for their knowledge of a number of areas within a discipline in addition to specialist knowledge. They needed to have local and international reputations and proven leadership track records. They were always chosen from outside the state and included a number of international experts and one of two members from outside the discipline. The choice of evaluator echoes Habermas's (1979) notions of rational discourse that reflect a commitment to four validity claims: truthfulness; intelligibility; appropriateness to context, and commitment to proposition. However, we can argue against these notions. Part of this study will be an attempt to unpick some of these claims.

Because of their size these commissions posed real challenges to the experts taking part, in terms of logistics, exchange of internal knowledge, and finally, consensus of opinion. One key criterion for choosing experts is that they could undertake the number of required on-site visits and thereby gain first-hand experience of the institutions, which might be difficult to assess accurately from paperwork alone, as the university or institution self report might be inclined to window-dress.

Although it was more common in other countries to evaluate either teaching or research, *evalag's* first evaluations examined both aspects. Self appraisals published by the universities included publications which appeared in peer-reviewed journals; however, some institutions presented anything that had been published, peer-reviewed or not, which indicates that standards of what is perceived to be quality vary. Students, too, were given a voice as the evaluators spent as much time with them as with members of staff or the administration, hearing their opinions about the institution and their satisfaction with it. However, perhaps even more needs to be done to incorporate students into the evaluation process, as can be seen by the shift in emphasis in Lower Saxony where one student is now a permanent member of each evaluation team.

Let us now briefly turn to the effects of evaluation in Germany, as the process in the first five years of this century was new, and it was therefore not always certain that evaluation would even serve a clear purpose.

2.16 Impact of higher education evaluation in Germany 2000-2004

While HEIs undergoing audit or evaluation inevitably take centre stage for the duration of the process, in very few cases will closure of an institution be discussed or recommended in the final report. Although institutions work hard to present themselves in the best possible light and the effort that goes into the compilation of self appraisals is considerable, there is therefore possibly less anxiety about the process in Germany than in other countries. Discussions with academics in the UK reveal that the audit process in general and the RAE in particular involve a heavy workload and there are great fears that if a department/faculty is identified as having a problem, it might receive less financial aid, which in turn will weaken the institution or the position of specific members of staff. The present adjusted RAE exercise pays tribute to this negative aspect and specifies that:

> the results will be published as a graded profile rather than a fixed seven-point scale. This allows the funding bodies to identify pockets of excellence wherever these might be found and reduces the 'cliff edge' effect where fine judgements at the grade boundaries can have significant funding impacts[23].

The German evaluation system identifies weaknesses and makes suggestions for improvements, but without such a major threat of financial retribution. This can also lead to complacency, however, because if the evaluation is not finance-linked, there is no need to necessarily change existing structures. An article published by the *ZEIT* newspaper, and counter-signed by numerous professors and academic staff[24] in 2007 indicated that in spite of evaluations showing where improvements could be made, there remains an ongoing dissatisfaction with teaching standards within universities. This remains a dilemma as research is easier to assess, whereas identifying where good or excellent standards of teaching lie is far harder to pinpoint with any degree of accuracy. Evaluations at the start

[23] See http://www.rae.ac.uk/aboutus/changes.asp, accessed 09.10.08
[24] See http://www.zeit.de/campus/2007/02/aufruf-lehre, accessed 24.09.07

of the 21st century were only going so far to identify this, which was what the professors were criticising. This is an interesting point, however. If it became standard practice within universities to have consistent and spontaneous peer-review of teaching, i.e. an open system whereby it would be normal for colleagues to occasionally visit each other's lectures and seminars with a view to discussing pedagogic methods, then standards might be raised. In the current climate this would appear to be wishful thinking as professors remain reluctant to have their work examined in this manner. Arguably, therefore, any evaluation of teaching can only examine the structures in place. It does not address the actual doing of individuals. Evaluations can only go part of the way to improving the quality of HE therefore.

2.16 Background to evaluation of *Erziehungswissenschaft* 2003/2004

As a consequence of Germany's PISA results in 2000, in which its 15 year old students did not do as well in reading skills as children in many neighbouring countries and internationally, the Ministry of Science, Research and Art in Baden-Württemberg decided to involve itself in an evaluation of Educational Science originally scheduled for 2002, which would have normally been conducted as just one in a string of disciplines to be examined. However, there was now far more emphasis on this evaluation than on others as the Ministry wished to determine the status quo of *Erziehungswissenschaft*, which was regarded as key to the successful preparation of teachers and of major importance to the state's overall education system. It specifically wished to determine the contribution of *Erziehungswissenschaft* to empirical research (protocol of Ministry meeting, 28.02.03). It stated categorically that its intention was 'to identify quality and structures, not see where there was potential for expenditure reduction'. Furthermore, the evaluation was intended to investigate the extent to which young academics were being encouraged, their promotion prospects, networks and collaborative efforts within and beyond individual institutions, and also the use of innovative potential and subject-specific allocation of resources. An examination of German evaluations of *Erziehungswissenschaft* by Böllert &

Tippelt (2004, p.142-143) reveals that at this stage in Germany, most were being conducted by ZEVA, the *Nordverbund*, the Bavarian Ministry of Culture, or *evalag* (Appendix A).

The Ministry of Science, Research and Art therefore commissioned the *evalag* agency in Mannheim to conduct a cross-sectional evaluation of this discipline as it stood within universities and teacher-training colleges (see Appendix B for map of HEIs in evaluation, final report, p.14). It should be remembered that an evaluation of *Erziehungswissenschaft* had already been conducted in this state a decade earlier, which pre-dated *evalag's* active operation. The new evaluation encompassed:

1. establishing commission membership
2. agreement on evaluation criteria
3. the compilation of institutional self-reports to be analysed by the commission
4. organisation of further meetings in Mannheim
5. organisation of on-site visits with those institutions being evaluated (15 in total)
6. coordination of feedback in the form of reports on onsite visits
7. recommendations for improvement or change as presented in the final report which was published by the Ministry[25]

Nineteen evaluators were asked to take part in this commission. Four of the group were foreign experts from other European countries. There was one permanent representative from Baden-Württemberg's Ministry of Culture, Youth and Sport (responsible for schools). Two members of the Ministry of Science, Research and Art were permanent guests who attended numerous meetings and a few on-site visits. The chairman of the commission was a Dutch professor of Education. Six of the commission were female. A sub-committee was appointed early to work in close liaison with the Ministry in terms of appointing further evaluators, preparing for the first plenary meeting and assisting the other evaluators throughout the process. These people and the work they carried out form the focus of my research therefore. It was clear, however, that an evaluation of this size had not been conducted before.

[25] This is available in paper form from the *evalag* agency or can be viewed online at: http://www.mwk.baden-wuerttemberg.de/uploads/tx_ttproducts/datasheet/Erziehungswissenschaft.pdf

Summary

In this chapter I have contextualised a number of different areas: quality in higher education, evaluation of higher education, the development of quality assurance in Germany and in Baden-Württemberg in particular. I also introduced the evaluation agency which was responsible for the organisation of the evaluation of Educational Science. Evaluations at the start of the 21st century in Germany were ambitious and to some extent all-encompassing, e.g. analysing research, teaching and learning, although we will see in Chapter 4 that the experts in the evaluation of *Erziehungswissenschaft* decided that it would be beyond their capabilities and timeframe to conduct an in-depth analysis of the quality of teaching.

In my next chapter I therefore present how I conducted my research and the theoretical framework within which this is couched, before moving on to the two other sections of my study: the documentary and interview analysis.

Chapter 3

Theoretical Framework and Methodology

In this chapter I review how I conducted my research and describe the processes that I underwent as work progressed. Parts of what follows have been adapted from Harris-Huemmert (2007b).

3.1 Theoretical framework

3.1.1 Phenomenology and decision-making theory

In consideration of the fact that my research has to take so many different aspects into account, each of which have their own theoretical backgrounds, I have found myself creating a framework that attempts to mirror this. I therefore do not adhere to one main theoretical line of thought, rather this research has been influenced by two apparently different and unrelated theories and even they act only as guidelines rather than being 'golden rules'. Firstly, as the work of evaluators of higher education lies at the heart of this study I decided to adopt a phenomenological approach, which has been used in a variety of educational research case studies in different contexts: adult first-time computer experience (Howard, 1994); teaching practice in junior schools (Trotman, 1995); the use of teaching with laptop computers (Bao, 1999); educational leadership (Van der Mescht, 2004) and the lecturer as assessor (Grant, 2008). We also find phenomenological case studies used within many medical contexts: Good (1994) and Toombs (2002) are two more recent examples, although the use of phenomenology in medicine extends as far back as the first half of the last century.

There are many different interpretations of phenomenology and to use a phenomenological approach therefore seems, at first sight, fraught with problems. For example, as was stated by Heidegger (1927, lecture course on *Basic Problems of Phenomenology*): 'there is

no such thing as the *one* phenomenology.' It has been argued that phenomenology is ultimately unsatisfactory in that it can only be a starting point, a description of something as opposed to an analysis. In the preface to his famous book Phenomenology of Perception (1945/2008), Merleau-Ponty stated that phenomenology 'describes rather than explains.' How did he wish to achieve this? By employing a critical self-analysis of one's own situation within the world it was possible to start addressing what he described as the 'essence' of what one sees around one, in this case the evaluator as he examines an institution of higher education. This then would lead to a clarification and classification of these truths which could be achieved if data were examined in a reductionist manner. In 'Husserlian' phenomenology, which is hermeneutic and interpretive, truths are claimed to be 'absolute and universally true' (Van der Mescht, 2004). However, if we focus on how individuals interpret their worlds, then such absolute claims cannot be made, especially as each individual will have a slightly different interpretation of the same phenomenon. As Pring (2000, p.102) says: 'Phenomenological research uncovers subjective consciousness' and by examining what different people say about the same experience it may be possible to uncover different outlooks on the same phenomena which are equally valid, but not necessarily in accordance.

Spiegelberg (1975, p.3) describes phenomenology as

> the direct investigation and description of phenomena as consciously experienced, without theories about their causal explanation and as free as possible from unexamined preconceptions and pre-suppositions.

This indicates that the role of the researcher must continue to remain as open and neutral as possible to describe what *is*, without imposing too much interpretation on what is being described. He argues that phenomenology can allow us to pursue an avenue of imaginative self-transposal and be a cooperative encounter when we engage with others (ibid., p.47). As we are not the person of the evaluator, we need to think ourselves into the experience of the evaluation (imaginative self-transposal). Further, it is through engaging

with the different members of the evaluation that we begin to understand the experience/phenomena they went through (cooperative encounter).

Phenomenology in my use of the term then does not aim to provide *the* truth; arguably, all truths are subjective. However, as a method of distilling experience I believe that it helps me to come close to a greater understanding of the case in hand. Phenomenology inevitably focuses on description, which I undertake extensively in my chronology of the evaluation in Chapter 4. It is only once the reader has gained such insight into the 'actuality' of what took place during the process or context of the evaluation that it becomes possible to begin to understand the felt experiences of each individual evaluator and start to form a synthesis of understanding as to what they individually and jointly went through. This takes us then 'back to the things themselves'. I therefore make no excuses for the length or detail of the narrative description in my next chapter as I believe it is fundamentally important to have this knowledge available before we can progress further. My decision to focus on the context of the evaluation also goes some way to countering criticism of phenomenology which fails to look at context. This criticism is most frequently heard in conjunction with Husserlian hermeneutic phenomenology.

The evaluator's lived experience which I describe in Chapter 5 is bracketed into units of understanding as it relates to the experience. However, I do not go as far as some phenomenologists to reduce the lived experience down to a single 'unit of understanding', as I believe that we need to retain the messiness of conflicting notions and values which was expressed during the interviews. I therefore prefer to present the differing understandings of the *same* experiences. It is by joining the narrative description of the evaluation to the individual 'lived' experiences (*Eigenwelt*) of the evaluator that I attempt to present a holistic view of this particular case (Van den Berg, 1972; Giorgi, 1985; 1992, 1994). The evaluator is in context, therefore, and we explore his individual interpretation of what he was experiencing. We also understand the world in which he moved during this particular time according to those going through the same experiences: the *Mitwelt* of

evaluation. The evaluator sits within the context of German higher education, and Educational Science (*Lebenswelt*), which is his discipline situated, as it is, within the *Überwelt* of German higher education (Spinelli, 1989; Van der Mescht 2004). The evaluation's organiser, Martina Röbbecke, explained this to me in interview: 'You may be evaluating a discipline, but in actual fact you also make contact with a huge amount of university or tertiary education structures, which you have to absolutely keep in mind.'[26]

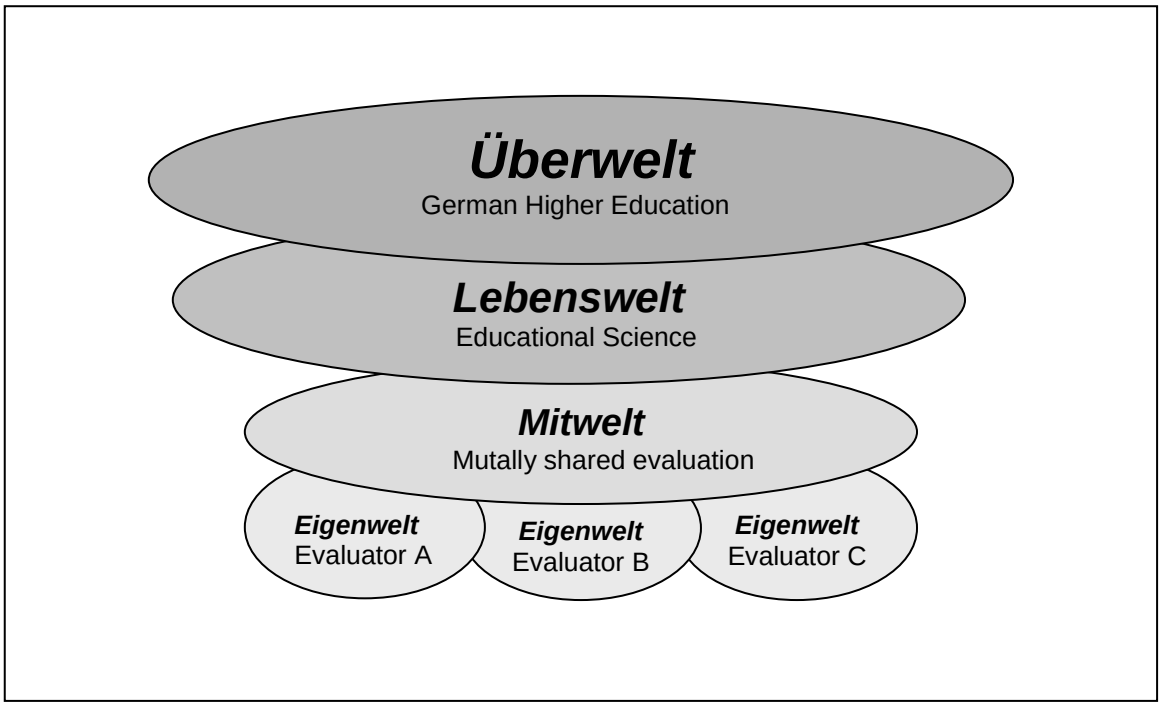


Figure 2. Holistic view of evaluator in phenomenological context

It is the merging of all four worlds therefore which will enable us to come closer to comprehending what was felt, which values existed and which judgements were made, as an evaluation is an assimilation of information and a road to conclusions on what has been seen, learnt, assessed, and described.

While we have the individual experiences of evaluators on the one hand, those participating were involved in an experience which necessitated processes of reaching mutual agreement. We therefore see a juxtaposition of the individual with the organisational. My research has therefore also been influenced by decision-making

[26] Man evaluiert das Fach, aber eigentlich berührt man natürlich kolossal viel von den Universitäts- oder Hochschulstrukturen. Das muss man unbedingt im Hinterkopf behalten.

theories which examine how people reach agreements in joint undertakings such as evaluation work (Hastie & Dawes, 2001; Raiffa 2002). Evaluation is a complex undertaking which brings people from numerous areas together to examine and judge the institutions they have been called to inspect. It is a study of human behaviour which involves gesture and communication through the spoken and written word. It involves group dynamics, rational choices, argumentation and decision-making at every step in its proceedings. We expect evaluators in particular to make rational choices. Hastie & Dawes (ibid., p.18) established four criteria by which they think a rational choice is made:

1. It is based on the decision-maker's current assets, which include physiological state, psychological capabilities, social relationships and feelings.
2. It is based on the possible consequences of the choice.
3. When these consequences are uncertain, their likelihood is evaluated according to the basic rules of probability theory.
4. It is a choice that is adaptive within the constraints of those probabilities and the values or satisfactions associated with each of the possible consequences of the choice.

However, they remind us that rational choices are not always made. Referring to Fischoff's observation (1996), they state that people have 'a general tendency to focus on a few *most salient* possibilities and consequences and ignore others, resulting in in-complete analysis' (ibid., pp.34-35). If this is true, it could be argued that evaluators, too, only focus on the *most salient* issues and that in their decision-making some areas are left unexamined. Hastie & Dawes (ibid.) have also highlighted that habit, conformity and religious principles or cultural mandates can impact on the way decisions are made. The second category in particular, conformity, refers to an inclination of individuals to imitate the decisions of those people they admire or regard as particularly successful. This can lead decision-makers to adopt a stance with which they might later be dissatisfied. My research will endeavour to examine some of the reasons for decisions and it will attempt to highlight why the evaluation took the direction it did.

3.2 Using qualitative methods

Qualitative research has been criticised as being messy, complex and sometimes unclear in terms of what it can offer. This argument seems partially justified. Results gained from qualitative research can be interpreted in a multitude of manners; they can be fuzzy and are not necessarily generalizable, although it is debatable in turn whether this should even be a criterion for research in any case. As Denzin & Lincoln (1994, p.2) state:

> Qualitative research is multimethod in focus, involving an interpretive, naturalistic approach to its subject matter. […] Qualitative research involves the studied use and collection of a variety of empirical materials […] that describe routine and problematic moments and meanings in individuals' lives. Accordingly, qualitative researchers deploy a wide range of interconnected methods, hoping always to get a better fix on the subject matter at hand.

Getting a 'fix on the subject matter' implies that it is not easy to achieve this aim. The study of any human subject will inevitably be associated with a multitude of meanings, as any subject can shift his or her position instantly and randomly, making a clear interpretation all the more difficult. Merriam (1988) points to the difficulty of reliability in social science research as human behaviour is in flux and non-static. Bednarz (1985, p.303) has even gone so far as to say that 'reliability – in the post-positivist sense – is fanciful.' Indeed, even the results from quantitative data can, in spite of best efforts to allay skew and variance, be misleading (Kincheloe & Berry, 2004). Nonetheless, quantitative methods are perceived by some to be more manageable where concrete numbers are the expression of tangible results, arrived at by apparently rigorous, transparent means. The tenor of Yin (2003), however, is that it is possible to apply the same rigour of method to qualitative research as to quantitative. He proposes that the researcher 'collect, present, and analyze data fairly' (2003, p.1). This is the perceived aim of the present study which, because it examines the work of one evaluation commission, may perhaps be described as a meta-evaluation case study (cf. Smith, 1990; Stake, 1986).

This thesis cannot make claims of generalizability and it is undeniably interpretive through the person of the researcher, although based on sound and hopefully rigorous

methods of empirical analysis. By examining a range of data and examining it phenomenologically, I can go some way to recording the 'routine and problematic moments and meanings' that Denzin and Lincoln (1994) refer to, in this instance the experiences and difficulties that evaluators of higher education in one particular evaluation commission encountered and their attitudes towards this experience overall.

3.2.1 Case study method

Yin (2003) has indicated that when 'how' or 'why' questions are raised it can be appropriate to use case studies as the method of choice. Furthermore, he defines case studies as exploratory, descriptive or explanatory. My case study attempts a combination of all three. It was inspired by Professor David Phillips, who himself had worked as one of the foreign experts in this commission. He suggested that German evaluations were a rich source for research. As a review of evaluation case studies in Germany, or indeed elsewhere, revealed a deficit in research on the work of evaluators of higher education, I decided to focus on how evaluators experience evaluations by looking in detail at the work of this one particular evaluation.

Although Mittag et al. (2003) conducted a large-scale, predominantly quantitative survey of the evaluation work in two north-German evaluation groups, the *Zentrale Evaluations- und Akkreditierungsagentur* (Central Evaluation and Accreditation Agency - ZEvA) and the *Verbund Norddeutscher Universitäten* (*Nordverbund* – Association of North-German Universities), no research has, to my knowledge, been undertaken on the work of south-German evaluations of higher education. My study therefore follows on from Mittag et al. (ibid.) who stated in their preface in their thanks to the cooperation of the two agencies that they hoped their handbook would 'contribute to bringing more transparency into the systems of evaluation of teaching and research in addition to providing useful suggestions for the manner in which evaluations can be conducted'[27].

[27] Wenn das vorliegende Handbuch dazu beitragen würde, mehr Transparenz in die Evaluationsverfahren von Studium und Lehre zu bringen sowie nützlichen Hinweise für die Durchführung von Evaluationsverfahren zu geben, ware dies sicherlich der beste Dank für ihre Mithilfe.

3.2.2 Emic and etic perspectives

According to Fetterman (1991, p.90) the emic (insider) approach is 'instrumental to understanding and accurately describing situations and behaviours.' Ethnographic research of any kind wishes to gain an understanding of social and cultural mechanisms by devoting time to observing a community or group of people. However, one of the disadvantages of this is that if research only looks at the emic view, it can lose valuable further perspectives that can be offered by those who are linked to those on the inside, but stand slightly removed from it. In this study, the emic view is provided by the evaluators themselves, whereas the etic (outsider) perspective is given by those institutions which were evaluated and had first-hand experience of the evaluators – in this case 15 institutions and teacher training colleges. In addition, two other etic perspectives are provided by the commissioning body, which in this case was the Ministry of Science, Research and Art, Baden-Württemberg, and members of *evalag*. Fetterman believes that it is only when emic and etic perspectives are brought together in analysis is it possible to state with any certainty that the knowledge gained is reliable, although others have argued that the concept of reliability is inappropriate in fieldwork settings (Wolcott, 1995). Arguably, it is only when a variety of lenses are employed in conjunction with further data from other sources that we may gain a rigorous and firm base from which answers to research questions can be found. Even so, 'it is not possible to give an exhaustive account of any object. In producing descriptions we always rely on criteria of selection and inference' (Hammersley and Atkinson (2007, p.41). My attempt at using both emic and etic perspectives on my research is limited by my own judgement and the decisions I take to focus on particular aspects of evaluator work.

3.3 Reason for not conducting pilot study

Before embarking on fieldwork it is usual to undertake some form of pilot study during which interview questions can be tried out and methods tested. However, this opportunity was not open to me in the present study. The commission in question consisted of 19 members, including one permanent representative from the Ministry of Culture, Youth

and Sport, Baden-Württemberg. As the experiences the evaluators had were specific to this evaluation and might not be similar elsewhere, it seemed inappropriate to use members from this group for preliminary testing, an idea I therefore overturned. Robson (2002) has also discussed those cases where it is simply not possible to do pilot work. It was necessary to enter the field without having been able to try questions out beforehand.

To work around this problem to some extent I decided to speak to as many professors of Education from other countries as I could, asking them about their experiences of evaluation and the quality assurance systems which were in place in their respective countries. I thought that this could familiarise me with the kinds of situations they encountered and help me to think about some of the problems they had to deal with. I therefore held talks with visiting professors in my department and informal discussions at international conferences. As a result, I gained some insight into evaluations of higher education in Canada, Mexico, and Austria. The Austrian professor in particular proved to be a rich source of information, as some of the academic mind-set in Austria is similar to that in Germany.

As Walford (1994, p.4) says, 'It is evident […] that all studies of the powerful involve an element of historical work' and in this case the history of higher education policy in Baden-Württemberg was a dimension that could not be excluded from the case study. Agendas of all kinds could play a part and I needed a keen awareness of them. A thorough understanding of the higher educational policies within Baden-Württemberg would therefore form part of my background research in order to contextualise the evaluation correctly.

3.4 Gaining overarching access

As my supervisor was himself directly linked to this evaluation, we agreed that he would approach the e*valag* agency and the Ministry of Science, Research and Art on my behalf, requesting permission for me to undertake research on this evaluation of which he had

been a part. Without his help and former participation in the evaluation access would undoubtedly have been more difficult, or even impossible. I am indebted to him for making these initial introductions. Although pleased by the ease of access, we did wonder why the Ministry had raised no objections to my research. Was I perceived by the Ministry as an 'expert', or 'critic' to whom Hammersley and Atkinson (2007, p.77) refer when they point out that the researcher may be regarded by a gatekeeper as someone who might 'sort out' an organisation? By investigating one of *evalag's* evaluations in detail, the Ministry may have assumed I would not discuss only the positive, but also the negative aspects of evaluation work. I could not answer this question at the outset of fieldwork, but as my research progressed I realised that there may have been an implicit reason for allowing a researcher into this field.

3.5 Archival research

By focusing on an evaluation that was concluded in 2004, this research is inevitably a work of contemporary history. The participants are living actors and have been and are able to share their experiences of working in this evaluation. Archival evidence has been described as 'unobtrusive data' (Hatch, 2002, p.117). It can include all manner of artefacts that people used in the course of their activities, which in this case was the predominantly organisational aspect of the evaluation. Hatch (ibid.) states further that 'documents are powerful indicators of the value systems operating within institutions' and he makes the important distinction between 'official' documents written formally and those which are provided of a more private, informal kind. Both provide insight into what people think, what problems they encounter and which strategies they develop for dealing with these. The most important use of documents is to corroborate and augment evidence from other sources. Yin (2003, p.87) puts documents into three categories:

1. Documents which corroborate and augment evidence from other sources.
2. Documents which provide other details. If details are contradictory, then interviews can follow this up.
3. Documents which help the researcher make inferences.

If all three are used, the researcher can develop a sense of history for what is being studied (Hatch, ibid.). Archival research therefore is 'an essential step' on the road to gaining a deeper impression of the case, which is why I decided to commence my fieldwork by examining the *evalag* archives.

After having negotiated a suitable date with the agency's then director, Dr Gerhard Schreier, I travelled to Mannheim on 18 July 2006 to have a first viewing. I was asked to provide the agency's employees with a brief synopsis of my research agenda so as to be informed about my study and particular points of interest. This meeting included three of the agency's evaluation organisers. The organiser who was responsible for the evaluation of education, Dr Martina Röbbecke, was not present, as she had been given leave of absence from the agency to work elsewhere.

At this stage of my research I found it difficult to say exactly what my specific points of interest were other than that I was keen to learn more about the work of evaluators and that this evaluation seemed to be a rich source of data as it was large, investigated both universities and teacher training colleges, and had the further dimension of having been commissioned by a Ministry. I was asked to sign a document of confidentiality, and in turn I asked the *evalag* employees to sign consent forms (Appendix C). In this form I declared that I would anonymise data and not quote directly, unless I had been granted permission. I will address the ethics of my research in greater detail below.

3.5.1 Method of archival research

Documentary evidence from the *evalag* archives consisted of a number of items including letters, the extensive university self appraisals, minutes of meetings, evaluator reports of meetings, and numerous occasional notes and organisational letters, for example stating which evaluators would visit which institutions. These could be divided into 'official' and 'unofficial' in the sense that some of the material was of a more complementary kind (e.g. personal notes, emails).

I examined files, noted the order of events as they unfolded, checked the minutes of meetings that were written independently by two agency employees and compared these against those of the experts who were in charge of writing down the details of the discussions. In consideration of the criteria Gottschalk et al. (1945, p.35) established, I could confirm that the documents were able to tell the truth, were 'willing to do so' and could be externally corroborated, either by parallel documents from other evaluators, or during the interviews held later. As these documents were written as standard practice and without any expectation that they would be used in future research, as far as is possible to ascertain, they seem a true and unbiased account of what happened. This is also supported by the authentic nature of the documents which do not disguise or 'dress up' anything. There do not seem to be any gaps in the material and the written evidence does not mask any of the problems of day-to-day evaluation work. In addition to the above material I was also given access to Professor Phillips's diary in which he noted down his feelings, reactions and thoughts on the evaluation process in its earliest stages. This is a particularly rich and unique source of data, which is a privilege to have been granted access to.

As Hatch (2002, p.148) says, data analysis is a 'systematic search for meaning.' By looking through the archives and reading as much about the evaluation as was possible, including the university self appraisals and the final published report, I was able to develop a sense for the evaluation, could construct patterns, identify the relationships of those involved and theorise about the decision-making dynamics of those working within the evaluation commission. This included examining relationships on four levels:

1. evaluator – evaluator
2. evaluator – evaluation agency
3. evaluator – ministry
4. evaluator – institution

All of these could, of course, be reversed in the examination of how the agency experienced the evaluators, or how the institutions perceived them. The archival work

acted as a guide, therefore, although I would not like to downplay its importance here. As I was not a direct participant during the evaluation itself, the archives reveal a depth of information on what took place which might not be as easy to obtain through interviewing methods alone. Most research is interested in what people do, rather in what they say they do (cf. Walford (2001, p.85). What people say in interviews should therefore always be treated with care as statements could be manipulated if the participant thought that the interviewer wished to hear a particular statement. The integrity of the participant is not necessarily in question, nonetheless, in certain circumstances, especially if the interviewer has asked a leading question, it might be possible to hear an answer which does not represent the true opinion of the participant.

I needed to look at the background of the evaluation before I could determine which specific research questions to ask. Tentatively, my overarching question during this stage of research was 'How does an evaluation of this kind, seen through the eyes of the experts, work?' The documentary analysis was not blue skies research, but already guiding me forwards. However, I wished to understand what the evaluators experienced from a variety of angles, using the lenses provided by those involved, including the researcher.

3.5.2 Limitations of archival analysis

As this study is concerned with the processes underlying one evaluation, one of my main interests was to examine the documents looking specifically at the distribution of roles and tasks and decision-making processes in plenary meetings and onsite visits. Although the minutes of the meetings in Mannheim and onsite with various stakeholders revealed part of this process, some of what was decided by those responsible for the compilation of the final report was not available to me. While the documents reveal concern for particular phrasing to the extent that certain tenses and expressions were debated, not all decisions could be traced. I have had to rely on what was available, augmented in part by what the individual participants described during the interviews.

3.6 Reason for conducting retrospective analysis

It could be argued that it would be more valuable to examine an evaluation as it unfolds and be witness to the process *in situ*. This would be a fascinating and worthwhile exercise which I would certainly be interested in taking up. However, this was an avenue I could not pursue on this occasion. To gain a comparative viewpoint, I asked the *evalag* agency if it might be possible to be present at the constitutional meeting of another evaluation of Educational Science they might be organising elsewhere. I was informed that there might be a possibility, as *evalag* had indeed been asked to organise an evaluation at the University of Münster in North-Rhine Westphalia in 2008. As this other evaluation was being conducted in North-Rhine Westphalia I needed the permission of that state, which unfortunately was not granted, although the subject of the evaluation would have been identical to the case study here.

One of my interview participants also recommended that I observe a one-day evaluation conducted by one of the Leibniz institutes[28]. This might have been possible in principle, however, I decided against investigating this possibility further, as the subject of the evaluation would have been quite different to what is being looked at here, and so I felt that it would not be of any great comparative value to me, even if it would certainly have been interesting to see how other evaluators make decisions about evaluation criteria and how they wish to proceed in general.

Evaluations, no matter where they are organised and held, involve a number of stakeholders. They are always political to the extent that the recommendations ensuing from the evaluation may have an impact on how the organisation/department under examination may function in the future. This can involve personnel or financial consequences. Hence, evaluations are sensitive exercises.

[28] The Leibniz Association is a non-university organisation comprising 86 institutions which conduct research in the humanities and social sciences, economics, spatial and life sciences as well as in mathematics, the natural and engineering sciences and in environmental research.
See http://www.leibniz-gemeinschaft.de/?nid=ubu&nidap=&print=0 for further details.

3.7 Method of interviewing agency employees

In addition to working through the archives, I decided to commence my fieldwork by conducting unstructured, informal interviews with three of the agency's employees (*Referenten*) to broaden my understanding of how evaluations in Baden-Württemberg are organised and structured. Informal interviews provide 'places where researchers and participants can co-construct understandings of what is happening in the research context' (Hatch, 2002, p.93). As far as I was aware, these informal interviews were the first time that the employees had reflected on their work with evaluators in this agency in a more formalised manner. Lincoln & Guba (1995, p.268) have proposed five interview outcomes:

1. Here and now constructions
2. Reconstructions
3. Projections
4. Triangulation
5. Member checking

My aim was to fulfil at least four of the above criteria in my interviews. The third point on the list was something I could only hope to put forward when all of the interviews were completed.

Although only one organiser was specifically involved in the commission on Education, being able to conduct interviews with the organisers of other evaluations increased my knowledge of how *evalag* operates, the experiences the organisers have, how they relate to the experts and the institutions involved. It also helped me to target specific questions that I could raise with the evaluators themselves. I discovered that my interviewees were willing to discuss their work; however, I was also aware that I might be drawn into any internal policy issues which may be pertinent to the participants, but which stood outside my own line of interest.

3.7.1 Method of interviewing evaluators

Having already decided against a pilot study, the next phase of my research was to start planning my interviews with each of the evaluators (18). After having conducted archival research and interviewed some of the agency's employees, I targeted a number of areas as being of particular interest as they seemed key to answering some of my research questions:

- To what extent did the ministry influence the evaluation?
- Did having prior experience of being evaluated, or evaluating, play a part?
- Was the extent to which evaluators already knew each other significant?
- Did the size of the commission make a difference?
- How did evaluators work around the mass of literature provided by the agency (15 self appraisals)?
- How did they compile the final report?

I constructed a formal semi-structured interview (Appendix D) in which I aimed to ask the same questions in the same order as nearly as possible. However, Chiseri-Strater & Sunstein (1997, p.95) refer to the ironic contradiction which interviewers of this form of interview have:

> You must be structured and flexible at the same time. While it's critical to prepare for an interview with a list of planned questions to guide your talk, it is equally important to follow your informant's lead. Sometimes the best interviews come from a comment, a story, an artefact, or a phrase you couldn't have anticipated. The energy that drives a good interview – for you and for the informant – comes from expecting the unexpected.

Although I had my list of questions, I soon realised that I could not always pursue them in the same order and that I did have to take leads from my participants where the unexpected turn occurred and needed following up during the discussion, which made me change the order of questioning. Having no comparable experience of using this precise interview before entering the field and not knowing which of the evaluators to speak to first, I selected an interview with one of the female evaluators who worked at an institution I was familiar with. This made the process slightly less daunting, although it was not clear whether the interview itself would answer the proposed questions.

During my research methodology training we had been advised against divulging too much of our own interests to any interview partners, which is something I adhered to in the first interview. However, my participants wanted to have a clear picture of who I was and what I was researching. They expressed the wish that they might prepare themselves better for the interview, focus on particular aspects and refresh their memories of the evaluation overall if they had some idea of the direction the interview might take. So, after usually approaching an interview partner via their secretary, a longer letter of introduction was sent via email. This was a method which was continued in all of the later interviews. It seemed to be a matter of some concern to interviewees whether they had been singled out, or whether all would be interviewed. I informed them either in the first email or before the interview commenced that I intended to interview everyone. The order of interviews was a matter of logistics and cost rather than a specific choice, as I needed to travel throughout Germany to complete my fieldwork and therefore tried to group together people in locations that were nearby.

To prepare for each interview I also looked up my interviewee's background as this gave me some information on their particular interests and where their expertise was situated. It was helpful to read their biographies as this could also indicate the extent to which they had collaborated with other members of the evaluation commission in the past, be that in evaluative work or in joint publications. However, such prior reading gave me a perspective on the person I was interviewing which could lead to bias on my part. To counter this or at least to document it, I kept a file containing this information and any notes on them that I had made. This formed an auxiliary part of the research data and included the context of the interview in addition to any key areas that arose from the discussion.

3.7.2 Duration, location and number of interviews

As I had been unable to pilot the interview, I was initially unsure of its duration, which made it difficult to say in advance how much time my interview partner should set aside.

This was problematic, as all of my interview partners lead very busy lives and I thought it inappropriate to be vague about the amount of time required. I therefore calculated rather roughly that we might need 75 minutes, which was a length I felt I could ask of them without being overly indulgent. This proved to be a fairly accurate calculation. The average length of interview was indeed 75 minutes, although the longest would be over 4 hours, and the shortest 50 minutes.

3.7.3 Choice of interview partner / sampling

The evaluators stand at the centre of my research. However, to illuminate their work from more than one angle I decided to conduct unstructured interviews with agency employees, ministerial members, and some institutional heads, enabling them to describe their perceptions of the evaluation and those involved. Allowing these interview partners to 'lead' the discussion lets them present things in order of their perceived importance. I intended to be of little influence in how I directed the interview, preferring instead to be more of a listener than directing. There were no sampling issues with members of the two ministries involved as I was able to speak to both of them. Equally, I was able to conduct interviews with the agency director, the organiser of this particular evaluation, and two other organisers of other evaluations (see Appendix E for list of interview partners, dates and locations).

Most of my interview partners were located in various parts of Germany. In total I held 27 interviews, fourteen of which were in nine of Germany's 16 federal states. Kiel, a port near Denmark, was the furthest north; Constance the furthest south. Two interviews with foreign evaluators were conducted in English in Oxford; the others were all conducted in German in Germany, including those with the Swiss and Dutch evaluators. Three interviews were held in Mannheim with *evalag* employees; one in Gütersloh with the former director of *evalag*, and three interviews at two teacher training colleges and one university (Weingarten, Schwäbisch Gmünd and Constance respectively).

The choice of sample is through a 'process which enabled the researcher to make assumptions about its representativeness of a wider population' (Tooley, 1999, p.169). In this case it seemed sensible to have a 100% sample and interview all of the evaluators. Equally, interviews were conducted with the two representatives from the ministries who were most involved in the proceedings. However, in terms of the 15 institutions that were evaluated, including both universities and teacher training colleges, I decided against interviewing all. Nonetheless, my sample of interviews (20% of the total number) which I conducted with institutional heads included both types of institution, large and small, civic or rural, and those which received more favourable or less favourable feedback. It seemed appropriate to form an understanding of their perceptions from both laudatory and critical perspectives as this could enhance the overall understanding of how the evaluation commission was regarded. I also wished to include both types of institution as their approach to the discipline was slightly different. I decided that a small sample would nevertheless provide me with some notion of how institutions react to evaluators and the evaluation process.

3.7.4 Interviewing elites: culture and the researcher

When trying to gain access to interview partners, problems can occur which may have an impact on the way an interviewee reacts later on. German institutions of higher education are characterised by marked and complex hierarchies (Lohnert & Rolfes, 1997, p.9). Not only is this the case, but it is a recognised fact that interviewing experts or elites in any field poses its own particular challenges (see Walford, 1994; Hertz & Imber, 1995), although some of the difficulties may have been somewhat over-exaggerated.

A first consideration when entering the field is the researcher's own background and status. In my case, the fact that I am a middle-aged female researcher from the University of Oxford, am a bilingual native speaker of German, and have lived and worked in Germany for many years in addition to having worked for a German university in the past, are all points that can and probably did have an impact on an interview situation. As Guba

& Lincoln (1985, p.90) have said: 'The inquirer and the 'object' of inquiry interact to influence one another; knower and known are inseparable'. Bailey (1999, p.34) also indicates that 'an individual's background constrains and directs his knowledge, that his interests are influenced by his training and the perceived problems of the day.' As a researcher 'you are researching yourself' (Walford, 2001, p.98). I am part of the research process and cannot distance myself or my background from my work, or indeed expect to do so. Some of the questions I raise or choose to emphasise will be different from how another person sees things. The evaluator's perception of me as the researcher can and probably does impact on the data arrived at through interview. Generally, they have wished to place me in some wider context and the fact that I can communicate with them in their own language certainly helped to put them at ease.

Being able to conduct the interviews in German immediately established a closer rapport because I was coming to the evaluators in their terms and language. My German language skills also enabled me to detect nuances of meaning or implication which might be missed if the interview partners had to communicate in a foreign language, although the majority of experts in German Education have a good, if not excellent command of English. It was quite apparent that interview partners appreciated my *Entgegenkommen* (accommodation) to them by speaking in German. It also implicitly raised my status with them as 'the interviewer has to show competence' (Walford, 1994, p.226). Furthermore, as I was coming into the field from a non-German environment I felt that this had distinct advantages for the particular area of my research. Interviewers have to be aware of the extent to which their own views are explained to interviewees (Walford, 2001, p.136). Coming from the United Kingdom I did not have a political agenda, which also made me less threatening. Evaluators therefore seemed more amenable to discussing '*Interna'* with me than they might otherwise with someone from within Germany. A danger to this, however, was that I would need to preserve anonymity if something of particular significance was disclosed. This will be discussed in greater detail below.

Although I believed it would be more appropriate to conduct the interviews in German, in terms of disseminating my research to a wider audience I have translated the examples taken from interviews back into English. As I am closest to the source of the data having conducted the interviews, I can translate into English maintaining the nuances of meaning as they occurred in the original. I have decided to present quotations in English translation, with the German original appearing in footnotes.

As mentioned above, I encountered less resistance to interview requests by being specific about my research. This became my method of choice as my interview partners appreciated hearing something about my research interests and which direction the interview might go. I was impressed that most had made the effort to refresh their memories of the evaluation by re-examining the final report before the interview. As Hillier (1999, p. 154) states:

> The interviewing technique is a particularly effective tool for reflecting on practice; it provides a focus without imposing structures determined by the interviewer, and its use generates a rich source of interpretive data that can be explored collaboratively with the respondents.

In terms of their powers of recollection, my decision to speak to all of the evaluators was influenced by Hastie & Dawes (2001, p.107) to counteract any biases from experiences more recent than the evaluation in 2003/2004. They state:

> Anchoring and adjustment can also severely affect our retrospective personal memory. Although such memory is introspectively a process of 'dredging up' what actually happened, it is to a large extent anchored by our current beliefs and feelings. This principle has been well established both in the psychology laboratory and in surveys. What we have at time of recall is, after all, only our current state, which includes fragments ('memory traces') of our past experience; these fragments are biases by what we now believe (or feel) to be true to an extent must be greater than we realize.

By speaking to each evaluator I was able to reduce the errors in memory by collaboration of information, although each and every recollection is to a certain extent unique to the individual and can, as stated above, be influenced by biases of the here and now.

Because my interview participants were not always readily available due to their heavy term-time commitments, I chose the summer vacation time as being most amenable for access when I knew that participants would not be involved with the pressures of term work. This was a good decision as invariably I was offered more interview time and interview participants were more relaxed as most did not have to rush off to other appointments. I conducted the first interview in October 2006 and the last one in December 2008. In those rare cases where an interview had to be ended due to time constraints before all of the questions had been covered, I was sent the remaining answers via email which I incorporated into the transcripts of the interview.

To encourage participants to reflect on what had been said during the interview and to give them the opportunity to delete a comment if they felt this was necessary, I returned the transcribed interviews to them for comment. I hoped that this would increase the extent to which they felt they could trust me, as they had recourse to their own words. A disadvantage was that if they felt that a statement they had made was problematic, it had to be excluded from my data analysis. However, I believed it was part of my integrity as a researcher to give my participants the opportunity to read through their transcripts. In only very few cases were text extracts removed.

3.7.5 Limitations of participant recall

Before I started my fieldwork, one of the key questions I had was the extent to which participants would be able to recall specifics from an experience which lay a few years in the past. In addition, some of my participants had engaged in evaluation work since then and might easily confuse one evaluation with another, especially if these were rather similar. This provided me with another argument for interviewing all of the evaluators and other stakeholders involved.

There was indeed a difference between those who had participated in more evaluations than those for whom this was possibly the first. Occasionally, some confusion did creep

into what they said. It was helpful to me to know which evaluations they had conducted both before and after this case study. Regarding their powers of recall, I was surprised at the extent to which interviewees were usually able to describe the minutiae of the evaluation (corroborated by evidence from other interviews).

As not all of the evaluators had participated in the same onsite visits, I had to rely on the memories of a few only (usually five) rather than all participants. It was useful at this point to have access to the list of which evaluator visited which institution to confirm who was present or not. In those cases where I was certain that an evaluator was describing something on hearsay only, I probed the questions more deeply to confirm that this was so. I was occasionally surprised when something was described in such vivid detail that it appeared the evaluator had been at an institution in person, when in fact he/she had not.

3.7.6 Ethical considerations regarding interviews

Not only was I being given access to archival documents, some of which were confidential, but I would also be speaking to extremely well-known and well-regarded members of the German Educational research community. Each evaluation of higher education has a political dimension to it (Robson, 2002, p.210) and a whole range of stakeholders are implicated. Those who choose to participate in an evaluation of this kind are aware that some of the comments/judgements they make will be critical and possibly cause controversy. There is an interesting paradox in that evaluators themselves need to be sensitive to those being evaluated while maintaining robustness in the face of criticism when judgements are declared. The same can be said of anyone researching evaluations. The methodology of the evaluation might come under criticism, the manner in which evaluators carry out their work could be found wanting, indeed, the role of the commissioning ministry might be found highly problematic, to name just some of the areas which could be exposed. I needed to consider these implications in advance of entering the field, not knowing the extent to which any of the above might be true. I therefore had to decide upon the manner in which I would treat data from the archives or

interviews and also how I would like to or be able to disseminate my research in published form later. This particular aspect has been discussed at length in the literature (cf. Burgess, 1989). Ostrander (1995, p.149) states:

> Whereas compromise in terms of timing of the release of publication may sometimes be necessary, the researcher should not compromise the integrity of the work by allowing elites to have a voice in deciding what is published or where.

Those being interviewed need to be aware that both positive and negative conclusions could be disseminated. If the researcher believes that it might be appropriate to let a certain amount of time pass until any negative conclusions are published, he can do so. Murphy & Dingwall (2001) remind us that harm should be avoided, although they state that the researcher might not always know what could actually cause harm as this is subject to the participant's perception of any given situation or context.

Fortunately, German law is unequivocal on the use of interview data. The law on individuals stipulates that no data can be forwarded to third parties without the express permission of the interviewee.[29] I therefore had to ensure that my participants knew that I would anonymise their statements so that it would be impossible to identify them from my research. In those cases where a statement was made by someone where I felt the others would immediately be able to identify the person in question, I had to refrain using this statement, even if it was particularly salient. Furthermore, on occasion and in those cases where I wished to ensure absolute anonymity I have disguised the gender of the interviewee. In these cases I have used the possessive adjective 'their', rather than 'his' or 'her'.

Regarding the archival documents, unless these were specifically confidential, I felt confident that I could use the names of those who were important to particular processes, as none of this had any implication for their personal abilities as evaluators. In Chapter 4, which analyses these documents, I have therefore used the names of those involved.

[29] See http://www.gesetze-im-internet.de/bundesrecht/bdsg_1990/gesamt.pdf for further details.

It would certainly have been more straightforward to discuss the data on every occasion if I could refer to specific participants; however, this would possibly implicate them in a manner with which they were not happy. Total anonymity, however, was something that I could not offer as the results of the evaluation are in the public domain and its participants are well-known. In conformity with Hatch (2002, p.52), I wished to treat my participants with 'respect, concern, and consideration', and so I decided on a compromise which I hoped all participants would be in agreement with by assuring anonymity except in those cases where participants had granted me permission to be quoted. This would be especially important in those cases where I would be quoting from instantly identifiable participants such as the ministerial members, the chair of the commission or certain institutional heads. My consent document broadly set out the aims of my work and informed participants that they could stop taking part at any time without needing to disclose their reasons. It also informed them of the following:

> Anonymity of those interviewed would be upheld, except in those cases where permission for quotation had been granted.
>
> Voice recordings would be held secure and material would only be used for the purposes of doctoral research.

With regard to data gained from interviews, the majority was quoted in 'coded' form, disguising the identity of the interviewee. I have randomly assigned letters of the alphabet to my interview participants. I believed that this was an appropriate compromise which would allow me to report fully on what was said, but maintain the integrity of those involved by not implicating them.

3.7.7 Method of interview data analysis

Data analysis poses problems of its own. We should ask questions such as: How can I be sure that what I am looking at or have listened to is an accurate reflection of what took place? Is it authentic? Before we engage in analysis (unless we have a grounded theory approach where we wait for things to happen), we usually already have preconceived ideas about what it is that we might be looking for. We do not simply plunge in sightlessly

and soundlessly. When patterns start to emerge and begin to confirm the hunches that we may have had, what we need to do is to try to find data that confound this by using the Popperian idea of refuting a conjecture. Is there anything we can hold up to the hunch that might bring it down again or does it withstand our criticism? Have the data been acquired through rigorous means? Does the sample of those interviewed stand up to criticism? If we can answer these questions with a yes, then any data that support our questions can be accepted. Sadler (1981, pp.27-30) has provided an informative list of frequent problems that can be encountered when qualitative material is analysed. This is a useful point of reference to avoid some of the pitfalls that can occur.

The basic premise that data analysis involves 'mind work' (Wolcott, 1995, p.233) is deceptively straightforward. The mind of the researcher is obviously involved, but this also means that his values, judgements and feelings are implicated at every stage of the research process and also need reflecting upon and describing, or, to put this in phenomenological terms, reducing. A qualitative case study of this kind may be duplicated in its methods; however, its interpretation will be unique to the researcher. I believe that another researcher using the same methods might arrive at the same conclusions.

My data analysis involved examining the written evidence as provided by the archives and the spoken word provided by the interviews. I used the archives to contextualise the evaluation and present a chronology of events and provide interpretations of what was happening as the evaluation unfolded. This was more descriptive than analytical, although I did not refrain from offering my own interpretation of events if I thought there was a particular aspect of the evaluation methodology or process that warranted highlighting (Robson, 2002). In many cases it was possible to corroborate evidence from minutes taken by different people at the same meetings (e.g. Röbbecke and her assistant).

Data from interviews were transcribed either by me or a German transcriber who also signed a document of confidentiality. The data transcription took well over one year to complete. Once I had completed the transcription I listened to each interview to ensure that the wording was accurate. Anything not transcribed by myself was also cross-checked for accuracy. Although I was not conducting a discourse analysis, I included all additional details such as pauses, coughs, smiles, gestures and so forth to remind me not only of what was said, but the manner in which it was said, as this is also part of the interpretation of experience.

Although many of my interview participants expected me to 'tidy up' what they said, as spoken German is different from academic German, I did not pursue this route as I wished to present their statements authentically. Their belief that I would 'tidy' the interviews reminded me that their academic *habitus*, or what was expected from those analysing interviews in Germany, was different from expectations in the UK context where it is normal for interview transcriptions to be identical to what was actually said. In those cases where a direct quotation which I wished to use was particularly 'rough', I did tidy it up very slightly to allow for smoother reading. This did not in any way, however, change the meaning of what was said. My translations have been cross-checked by other German native speakers to ensure their accuracy.

When I started the interview coding using MaxQDA software, the first few interviews I coded contained fewer codes than those I coded later as I was creating units of understanding as I progressed through the transcripts. After I had coded a sufficient number and felt satisfied that I would not create more codes, I went back over the first interviews to ensure that they were coded in the same manner as the later ones. The number of codes originated was immense (2644 in total). Some codes such as 'Reason to do evaluation', 'Personal Qualities of Evaluators' or 'Attitude to evaluation' had numerous sub-codes. The computer aided me in terms of organising my data as I could immediately group together quotations from similar codes. It also helped me to identify

the weighting of different codes, allowing me to present some aspects in order of importance to the evaluators. The coding as such became quite complex and discrete as I grouped together themes and sub-topics that recurred, in addition to coding themes that stood out from the rest as being singular. It is always the researcher's decision as to which examples to choose and this is a subjective decision, although where possible I attempted to select quotations that were representative. Those quotations which might be regarded in quantitative terms as outliers were also allocated and I provided reasons for why these singularities might occur.

As I did not use the same semi-structured interviewing technique with institutional heads, ministerial members or evaluation agency members as I did for my interviews with the evaluators, I entered the transcribed documents into MaxQDa, but used a different, less rigid method of coding for each of these groups, creating sub-sets of codes for these. In particular, the interviews with institutional heads were very specific to their organisations, which made it almost impossible to apply the same codes to the interviews. I therefore created sub-codes specific to each institution in question.

3.8 Summary

In this chapter I have outlined my theoretical framework which uses phenomenology and some aspects of decision-making theory to aid our understanding of the individual and organisational aspects of evaluation work and the stakeholders involved. I have described the way I conducted my research and the decisions I took before I entered the field, and during the first stages of my fieldwork. By examining the literature on the evaluation and speaking to the stakeholders concerned I aim to present a complete picture of the evaluation so that the reader feels as close to what took place as is possible from a distance.

In what follows in the next chapter, we shall learn about the political background to this particular evaluation and understand more about the individual backgrounds of those

participating in the evaluation, their specific roles and the work they undertook. We will then be equipped with sufficient information to understand what the experts described to me in interview, from Chapter 5 onwards.

Section 2: DOCUMENTARY ANALYSIS

Chapter 4

Chronology of evaluation as revealed through documentary analysis of *evalag* archives

4.1 First steps into documentary research

To start my research with archival research seemed a sensible decision as 'documents are powerful indicators of the value systems operating within institutions' (Hatch, 2002, p.117). They provided me with insight into the various stages of the evaluation and indicated some of the accompanying problems or challenges which the evaluators came up against. I did not know from the outset the extent to which the archive at the *evalag* agency in Mannheim would yield results, a problem frequently encountered by researchers as they do not initially know how rich the data will be. However, as *evalag* is an organisation designed to support a public system (HEIs), I felt confident that records of the evaluation were kept, although I did not know the extent of coverage. Fortunately, *evalag's* archive was well organised and comprehensive, although the agency had not conducted that many evaluations up to this point. After scanning the archive I could group the documents which were examined for this thesis as follows:

1. Ministerial correspondence and legal documents
2. 15 self reports from HEIs
3. Minutes of meetings
4. Various drafts of sections of final report
5. Correspondence with various stakeholders (participants and HEIs)

In addition to the sources at Mannheim I was also given a diary which had been kept by Professor Phillips. In this he recorded the first stages of the evaluation. Information from this source is particularly valuable as it reveals the manner in which he approached the

evaluation, his feelings about how the commission worked together and about its development as work progressed.

In order to understand this evaluation and those who were responsible for the decision-making, my chosen modus operandi is to present the documents in the order in which they appear, rather than grouping these into categories or themes. This chapter will reveal how decision-making processes were steered by certain people at particular times, in addition to examining the overall debates that were conducted during the various plenary and sub-committee meetings. Supplementary supporting statements from those HEI heads who were interviewed, or from members of the Ministry, also appear in this chapter and help to triangulate the data.

While reference has already been made in Chapter 2 to developments in evaluation theory and practice, the first section of this chapter will describe some of the events in German education during the last four decades of the twentieth century which led towards the evaluation of Educational Science in Baden-Württemberg in 2003/2004. It is not the purpose of this first section, however, to provide a comprehensive overview of all recent higher educational reforms in Germany as this would go beyond the limits of my study. The chapter then moves on to examine the different stages of the evaluation. It will become apparent from what follows that the process in which the experts were involved was challenging and messy. Some issues, such as which evaluation criteria would be suitable for use in Educational Science, were not clear when work commenced. The group would keep needing to readdress its own values on certain aspects of this discipline and prepare numerous drafts of different sections of the report before they were satisfied that what they were presenting was final (see Chapter 7).

It might be easier for the reader if I had grouped certain events, such as plenary meetings, together and discussed these as one unit. However, by so doing I would be disguising the complexity of the evaluation and removing part of its untidiness. As Ball (1990, p.9) states:

> Abstract accounts tend towards tidy generalities and often fail to capture the messy realities of influence, pressure, dogma, expediency, conflict, compromise, intransigence, resistance, error, opposition and pragmatism in the policy process. It is easy to be simple, neat and superficial and to gloss over these awkward realities. It is difficult to retain messiness and complexity and still be penetrating.

I am faced with the challenge of attempting to present the evaluation in its messiness, but still provide 'penetrating' analysis of what was taking place. Some parts of this chapter might appear to be repetitive or messy; however, this mirrors exactly the revision processes that the experts underwent as they established their evaluation criteria and refined them.

Although this section of my thesis predominantly presents evidence from documents, these only tell part of the story of how the evaluation proceeded and how the evaluators felt about what they were doing. Especially in consideration of how the evaluation commission was established, or who suggested whom and why, the documents alone are not sufficient as evidence. The oral history from the interviews with the experts, which I present from Chapter 5 onwards, will further augment our understanding of the evaluation and the more emotive aspects of evaluation work. As McCulloch & Richardson (2000, p.116) state: 'Oral sources are presented as a suitable means of discovering the experiences, outlooks and alternative agendas of […] groups'.

4.2 Historical background preceding evaluation 2003/2004

As Germany continued with its reconstruction following the Second World War, it became clear that with increasing numbers of children being born at the start of the 1960s at some point schools would struggle to educate them if universities did not provide sufficient numbers of qualified staff (Führ & Furck (eds.) 1997). Equally, some regions of Germany were not as well endowed with HEIs as others. Baden-Württemberg revealed notable gaps in the distribution of its HEIs, although it did harbour some of Germany's most famous: Heidelberg, Tübingen, Freiburg, to name but a few examples. Furthermore, equality of opportunity was not pervasive, as the three-tier system of schooling favoured university

entrance for those with more academic parents (Dube, 2006). Against this backdrop, the Heidelberg professor Georg Picht produced a number of articles in the Protestant conservative weekly *Christ und Welt,* which was published in 1964 as a book entitled *Die Deutsche Bildungskatastrophe.* Picht called for a radical restructuring of education to provide enough teachers for the two million additional children expected by 1970. Although his ideas were not entirely new, Picht drew the link between economic performance and education, stating that insufficient investment in education could lead to an economic crisis or even standstill. He also said that some parts of society (women, children from rural areas, and Catholics) were being discriminated against (Hahn, 1998, p.121). A number of federal states reacted immediately and Baden-Württemberg, as just one example, wished by 1970 to double the number of children leaving school with the *Abitur*.

As the 1960s proceeded, higher education therefore became accessible to more students than ever before. As previously mentioned, the *Abitur* entitled university entry and although more students were seen as desirable, Germany's politicians struggled to accommodate to ever-increasing student numbers. Many new institutions were therefore founded during the 1960s and the early 1970s, the majority of which were non-specialist in that they provided for all disciplines of study. Two examples of such new HEIs in Baden-Württemberg are the teacher training college in Freiburg founded in 1962 and the University of Constance, which was founded in 1966 to improve opportunities in the region. Owing to Germany's federal structure, however, higher education reform in general was by contrast slow, which was something students in particular were calling for. They wished HEIs to be less authoritarian, engage in new forms of pedagogy, and generally be more welcoming to new curricula (Hahn, ibid., p.124).

Institutions originally conceived for small numbers of students now catered for the masses. Professors still tried to maintain their idea of an elitist ivory tower, but found they were now dealing with unaccustomed workloads and a very different, less elitist student body with a broader representation of the wider public (Enders, 1997). The student revolution of 1968 saw the relinquishment of long-standing traditions of academia, for example, lecturers in gowns, or graduation ceremonies, which were overturned in favour of a new found modernism. Although professors were reluctant to welcome these changes, German higher education had undeniably entered a new era which they needed to come to terms with.

Baden-Württemberg is one of Germany's largest states and some of its HEIs are famous far beyond its borders (e.g. Tübingen, Freiburg, Heidelberg). However, owing to its size and the fact that some parts of the state are geographically rather isolated and lack a decent infrastructure (the so-called *Oberland* in the area of Ravensburg, Weingarten and Schwäbisch Gmünd) means that some HEIs are quite regional in character and cater for students in the direct vicinity only, whereas others, such as Heidelberg, draw students from a far greater area. The nature of each institution is therefore very different, with tiny institutions of only a few thousand students surviving alongside the traditional giants with vast student populations. Considering that the charters of each institution (university or teacher training college) are basically similar, with an emphasis on teaching, research, student support and the provision of training of future academics through doctorates, it might seem appropriate to compare like with like in terms of quality of teaching, research and research agendas.

Concerned with the quality of its HEIs and fairly quick to react to changing circumstances or challenges, the state of Baden-Württemberg had a long history of self-reflection and

engagement with standards. It was therefore not particularly surprising that following an initial recommendation by the Ministry of Science, Research and Art, two ministerial meetings in December 1990 and July 1991 put forward the creation of a commission *Pädagogische Hochschule 2000* to determine the future of its teacher training colleges. In their coalition agreement in May 1992, both main political parties, the CDU and the SPD[30], voted for the inclusion of the question as to how teacher training overall would be conducted. The title of the commission was therefore altered to 'Structural Commission Teacher Training College 2000' (*Strukturkommission Pädagogische Hochschule 2000*)[31].

It was recognised that teacher training colleges were in a comparatively weak position in comparison with those institutions offering teacher training throughout the rest of Germany. Their staff members were predominantly older and moving into retirement; there were limitations in the extent to which research could be conducted as professors had large teaching loads; student numbers fluctuated quite dramatically on an annual basis, which caused organisational problems; and lastly, their libraries and administration were insufficiently equipped. The commission therefore suggested a number of possible future scenarios. They proposed that alliances with other institutions might be formed, or that the teacher training colleges be incorporated into the universities, which was a decision that the other *Länder* had already taken. Alternatively, they could be developed and enlarged, or grouped together. Although the commission proposed this diversity of ideas, the one they favoured was for teacher training to be integrated into the universities by 2001. The proportion of professors to lecturers (2:1) should be reversed, with a minimum of two professors for each key subject area. Overall, research should be strengthened and more finances provided to improve the position of the HEIs so that they might be able to

[30] The CDU (Christian Democratic Party) is a conservative party, whereas the SPD (Social Democratic Party) is more left-wing. German politics have been dominated by these two parties, with the FDP (Free Democrats) providing a third more liberal direction.
[31] See *Lehrerbildung in Baden-Württemberg. Kurzfassung des Abschlussberichts*, Stuttgart 1993.

compete on equal footing with the rest of Germany. However, the *Land* decided against merging teacher training colleges with universities and chose the support option with a view to strengthening their position overall. Baden-Württemberg would therefore become the only state in Germany to continue with this system of teacher training for those wishing to be employed in *Grund-*, *Haupt-*, *Real-*, and *Sonderschulen*, whereas *Gymnasium* teachers would be trained in universities.

Four years later, in 1996, Baden-Württemberg conducted a further review of its nine universities to investigate whether they were working efficiently and to identify what might be required to ensure a good tertiary education in the future. Although the 1980s had been characterised by overcrowded HEIs, student numbers in the 1990s were declining again and Baden-Württemberg identified that Germany in general was not a particularly attractive venue for international students. In 1998, after 18 months of intensive work conducted by an independent panel of scientists, industrial bosses, representatives from the HEIs and representatives from organisations outside the state, a report was submitted to the then *Ministerpräsident* Erwin Teufel (see *Abschlussbericht der Hochschulstruktur-kommission*, MWFK, June 1998). In brief, although the quality of Baden-Württemberg's HEIs was deemed to be good, it was recognised that the available resources needed restructuring, although no HEI would be closed. The overall tenor was a recommendation for more streamlining and profile development, which inevitably meant that some universities would be obliged to renounce some of their breadth in favour of focussing on particular strengths, a move that pre-dated the later DFG *Exzellenzinitiative* during Chancellor Schröder's government in 2002 by four years[32]. The HEIs had been incorporated in this investigation from the beginning and perhaps unsurprisingly had been

[32] Schröder's party, the SPD, was predominantly egalitarian and left of centre in its political outlook. The idea of excellence was therefore novel and came unexpectedly. Schröder had recognised that German universities were struggling to maintain their international reputation, and so he put forward the idea of creating an initiative whereby individual research groups, departments (or HEIs) could be given the badge of excellence if their work was judged to be of such a high quality.

initially hostile to the process until they were informed that their finances were secure for a period up to 2006. Nonetheless, they were required to lose 9% of their staff within a ten year period. A so-called 'solidarity pact' between the universities and the state ensured that reforms would be ongoing and that the recommendations would not be put on the shelf. Rudolf Böhmler, who chaired the commission, concluded the introduction to the report by stating that he believed Baden-Württemberg's universities would continue to maintain their excellent position in Germany, especially as science and research were given such a high political priority.

In the wake of the first PISA study in 2000, in which, as I have already mentioned, German schoolchildren did not perform as well as many of their counterparts in other industrialised countries, or indeed even in some which were not as well developed, and also in consideration of the Science Council's recommendation (*Wissenschaftsrat* - WR) that empirical educational research be expanded and applied, politicians immediately expressed concerns about the manner in which teachers were being prepared for their role. Empirical educational research was seen as key to the success of improving standards in schools. In 2002 the SPD politician Günter Kaufmann put forward a request to the Baden-Württemberg parliament (*Drucksache 13/1160, 9.7.02*) calling for an examination of the situation of empirical educational research in the state. The request was comprehensive and divided into two main areas:

Part I

1. To see which institutions within and without Baden-Württemberg's HEIs were conducting empirical educational research and which areas of focus they were covering.
2. The extent to which the institutions had developed their personnel and material resources over the past 10 years.

3. To what extent and for which research projects money from funding bodies had been invested in empirical educational research, and more specifically in teaching and learning research and the ongoing development of subject didactics.
4. To discover how much the overall amount of financial support from above was in proportion to the total of research funding in the arts and social sciences.
5. To discover which empirical scientific surveys had been conducted by HEIs in Baden-Württemberg over the past 5 years on the basis of questions on
    a. Teacher-learning relations
    b. The use of applied knowledge
    c. The capability of teacher education
    d. The evaluation of school work
    e. School autonomy
6. To discover how the results of empirical educational research might be systematically integrated into the training and further training of teachers and school development and to what extent they were being considered.
7. To discover how the structural and development plans of BW's HEIs took account of the empirical educational research being conducted and to identify where cuts or developments might occur.

Part II

To develop a concept for developing empirical educational research and to optimise the results of research into school practice.[33]

[33] Der Landtag wolle beschließen, die Landesregierung zu ersuchen I. zu berichten: 1. welche Einrichtungen innerhalb und außerhalb der Hochschulen sich in Baden-Württemberg mit empirischer Bildungsforschung beschäftigen und welche Arbeitsschwerpunkte diese Einrichtungen jeweils haben; 2. wie sich an den unter Ziffer 1 angesprochenen Einrichtungen, die aus dem Staatshaushalt finanziert werden in den vergangenen 10 Jahren die Personalstellen und Sachmittel entwickelt haben; 3. in welchem Umfang und für welche Projekte Einrichtungen der Forschungsförderung in den vergangenen 10 Jahren den Hochschulen Fördermittel für die empirische Bildungsforschung insbesondere die Lehr-Lern-Forschung und die Weiterentwicklung der Fachdidaktiken zur Verfügung gestellt haben; 4. welchen Anteil die unter Ziffer 3 genannte Forschungsförderung am Gesamtumfang der Forschungsförderung für den Bereich der Geistes- und Sozialwissenschaften ausmacht: 5. welche wissenschaftlichen Untersuchungen der Hochschulen des Landes sich in den vergangenen 5 Jahren auf empirischer Basis mit Fragen * des Lehr-Lern-Verhaltens, * der Verwendbarkeit des Ausbildungswissens, * der Leistungsfähigkeit der Lehrerbildung, * der Evaluation schulischer Arbeit und * der Schulautomomie beschäftigt haben und welche Schulen des Landes dabei als empirische Felder in die Untersuchungen jeweils einbezogen wurden; 6. wie die Forschungsergebnisse empirischer Bildungsforschung systematisch in die Aus- und Weiterbildung der Lehrerinnen und Lehrer und die Schulentwicklung einfließen können und inwiefern sie dort berücksichtigt werden; 7. wie die Struktur- und Entwicklungspläne der baden-württembergischen Hochschulen die Einrichtungen der empirischen Bildungsforschung behandeln und insbesondere, ob und wo Streichungen oder ein Ausbau geplant ist. II ein Konzept vorzulegen, mit dem in Zukunft die empirische Bildungsforschung in Baden-Württemberg verstärkt weiterentwickelt und die Rezeption der entsprechenden Forschungsergebnisse durch die Schulpraxis optimiert werden kann.

The state reacted quickly. By September the Minister of Science, Research and Art, Dr. Frankenberg, had produced a detailed report (25 September 2002 Nr.21-610/55) in which his own Ministry and the Ministry of Culture, Youth and Sport presented a joint statement. Descriptions of each institution in which empirical educational research was being conducted were given and answers provided on each of the above questions. Notably absent from the report was a clear picture of what funding was being provided for such research. Absent, too, was an overall vision as to how this research might be developed within the structural and material planning of the HEIs. It was apparent that the HEIs generally wished to maintain the existing status quo which they regarded as sufficient.

With regard to the second main area, the state was unequivocal. It viewed empirical educational research as significant and believed that results gained from this kind of work were fundamentally important. Although empirical educational research was seen as being of a good standard, the *Land* wished to ensure that this would remain an important contributor to helping schools maintain or improve their standards. Baden-Württemberg would therefore continue to give empirical educational research high status and be willing to provide support to ensure its success.

As previously mentioned in Chapter 2, the *evalag* agency was created to provide an independent evaluation service for the HEIs in Baden-Württemberg. At this stage in its history it occupied a perhaps strange position as it was an independent body, but funded by the state. The HEIs were therefore sceptical and voiced their concerns unreservedly. The agency was regarded as an instrument of state control and a cost factor. University presidents told Dr. Gerhard Schreier, the first director of *evalag*: 'you are using up our money'[34] (Interview, 19.12.07).

The governing body of *evalag* consisted of six university presidents, one representative from the Ministry of Science, Research and Art and three so-called externals, consisting of

[34] Sie verbrauchen unser Geld.

one former minister of state, one representative from industry, and a Dutch university president. The externals were not given voting rights, however. The board was chaired by the General Secretary of the Swiss University Chancellors Conference (see *Qualität für Lehre und Forschung, evalag,* p.27). It was this combination of politics and university chancellors which led to *evalag*'s precarious position from the first day of its active operations in 2001, as HEIs were acutely aware that politics were closely involved and that *evalag* might not always be able to operate autonomously. Institutions were therefore immediately suspicious of evaluations because they felt they went 'against our autonomy'[35] as Schreier reported. Schreier had come to the agency following engagements with the Standing Committee of University Chancellors (*Hochschulrektorenkonferenz*, hereafter HRK) and the *Wissenschaftsrat* during which he had already identified that the higher education arena in Baden-Württemberg was 'something special, a difficult terrain'.[36]

4.3 Organisational steps towards evaluation of *Erziehungswissenschaft*; composition of expert committee

The first three evaluations *evalag* conducted in 2002 were of Architecture, Geography and Social Work. In the preface to each of these evaluations, the then head of *evalag's* governing body, Dr. Mathias Stauffacher, wrote that the agency was still in its infancy, finding its feet and that the HEIs being examined were also treading new ground as evaluation was still a new phenomenon, but that most HEIs had encountered experts who were committed, willing to support the institutions and develop new ideas with them to improve what was already there, echoing Guba's comment at the start of this thesis.

An evaluation of Educational Science was scheduled for 2002. However, the PISA debate and the state's concerns about the extent to which empirical educational research was being conducted and applied brought about a parliamentary decision to commission a cross-sectional evaluation of this discipline. The published report of the evaluation reveals

[35] […] gegen unsere Autonomie […].
[36] […] was besonderes, eine schwierige Landschaft […].

that the Ministry commissioned the evaluation in the spring of 2003. However, work towards establishing a commission had of course commenced early in 2002. In advance of the already scheduled evaluation, Schreier had started corresponding with Professor Ewald Terhart in Münster to talk about which experts might be included and their discussions included mention of Terhart's experiences working as an evaluator in Bavaria and Lower Saxony (letter, 30.03.02). Earlier Terhart had written to Schreier (letter, 14.02.02) saying that he would recommend Professor Heinz-Elmar Tenorth whom he described as 'highly competent in his subject and experienced in the politics of higher education'[37], which already gives an indication of which qualities were possibly being looked for in the commission membership. It seems that at this point both Terhart and Tenorth were envisaged as co-chairs of the *evalag* Education evaluation (interview with Tenorth, 16.07.07)[38].

In preparing for this earlier planned evaluation, the HEIs offering Educational Science had been asked to put forward one spokesperson who in turn should nominate three people to act as external evaluators, which was the system for providing suggestions for evaluation commissions. The *evalag* agency in turn then usually chose an internationally recognised expert from the field to take the chair and make further recommendations for the commission. The chair needed to get an initial idea of the status quo within these institutions, then organise the remaining experts accordingly and ensure that the commission did not get too large. Ideally, the chair should have good contacts in higher education, have a broad spectrum, and be familiar with the political environment.

Thus, after dossiers on various individual experts had been sent to both Terhart and Tenorth, Schreier asked them on behalf of *evalag's* governing body (letter, 02.08.02) to prepare a list of experts and return their recommendations by 16 September 2002. The following points were given for their consideration:

[37] […] fachlich hoch kompetent und hochschulpolitisch erfahren.
[38] Bevor ich zu dieser Kommission kam, gab es Monate vorher schon einen Versuch, eine Kommission einzusetzen, der Herr Terhart und ich hätten vorsitzen sollen.

1. To ensure that the different subject areas were sufficiently covered or represented twice, as the evaluation needed to look at 14 HEIs (which was the status quo at the time).
2. The experts should be sufficiently qualified in both teaching and research to preclude any doubts about the independence of their judgements.
3. Independence should be directly maintained to preclude conflicts of interest if possible.[39]

Professor Jürgen Baumert, director of the Max Planck Institute for Human Development in Berlin, had also been approached (letter, 17.07.02), but initially declined the invitation because of time commitments, but also because he felt that as the director of a non-university research institution he should not be on the commission (letter, 19.07.02). Around the same time Professor Hans-Uwe Erichsen, an emeritus professor of Law and former Director of the DFG whom Schreier had earmarked as a possible non-specialist, was approached, although Schreier told him that the final decision on who would become a member of the commission would be made by the governing body of *evalag* (letter, 11.07.02).

Meanwhile, it seems that in its preparation for its Education evaluation the Ministry was conducting its own search for experts. We therefore see developments towards compiling a list of experts for a standard evaluation of Education on the one hand, while the Ministry was concurrently initiating its plans to ask *evalag* to do an evaluation on its behalf and had commenced its own search for experts on the other. An overlap of interests was therefore inevitable. When Schreier received notification from the Ministry of these plans he was therefore in a difficult position as he had already recruited Tenorth and Terhart as co-chairs

[39] Die fachlichen Teilgebiete sollen hinreichend abgedeckt sein, ggf. auch doppelt, da das Verfahren nach derzeitigem Stand 14 Standorte umfasst und damit gerechnet werden muss, dass einzelne Fachvertreter nicht an allen Standorten mitwirken können.
2. Die als Gutachter mitwirkenden Wissenschaftler sollen in Lehre und Forschung hinreichend ausgewiesen sein, nicht zuletzt, damit an der Unabhängigkeit ihres Urteils kein Zweifel entstehen kann.
3. Unabhängigkeit soll auch in dem direkten Sinne gewährleistet sein, dass direkte Interessenkollisionen nach Möglichkeit ausgeschlossen sind.

and approached Erichsen, but now he had to start over again, this time taking the considerations of the Ministry into account. He had to explain the new situation to them and find out whether they would still be willing to be involved in the new evaluation, especially as the role of chair would be taken by a Ministry-appointed foreigner. As the Ministry was now the commissioning body, *evalag's* governing body asked the Ministry to also take responsibility for the selection of experts.

Once the ministerial contract was signed and in contrast to the existing practice of selecting a professor from Germany for the role of chairperson, it was obvious that a strategic decision to appoint a foreigner to the chair had been made. The Ministry of Science, Research and Art asked *evalag* to contact Professor Ype Poortinga of Tilburg University in The Netherlands to take over this role (letter, 11.11.02). However, Poortinga was unavailable, and so the Ministry approached Professor Eric De Corte from Leuven in Belgium. De Corte was an experienced evaluator who had worked on numerous evaluations both in his own country and in many others. He declined to take the chair, however, as he felt that his spoken German was not good enough to warrant allowing him this important position (email, 16.11.02). Nonetheless, he replied that he would be willing to be involved as an ordinary expert. Schreier informed two ministerial members, Knorr and Herberger, that De Corte had declined, but that Professor Wynand Wijnen from the University of Maastricht in The Netherlands, an equally experienced and highly regarded evaluation expert, had indicated he would accept (email, 19.11.02). The Ministry accepted this recommendation and Wijnen was duly appointed. Both of these experts had also been recommended to Schreier by Jürgen Baumert (email, 02.12.02).

As the chair was a foreigner, the Ministry elected to have two people from within Germany to work closely with him and support him throughout the proceedings. Although Baumert had declined Schreier's initial offer in July, he was therefore re-approached, but this time by the Ministry directly. An expert on PISA and an empirical researcher, he had worked with the evaluation's organiser, Dr. Martina Röbbecke, when she was also employed at the

Max Planck Institute for Human Development, and so this time he decided to accept and became a member of the sub-committee to support the chair. The Ministry also put forward the name of Professor Marianne Horstkemper from Potsdam for the sub-committee as she was an expert in teaching. The precise reasons for these recommendations are not known; however, the geographical proximity of Potsdam to Berlin might have been a factor, as it was possibly thought that Baumert and Horstkemper could perhaps more easily organise meetings in advance of the main evaluation to discuss methods and contents.

Following his appointment Baumert went through the list of some of the other experts suggested by *evalag* and put forward names of additional female members, who he felt were underrepresented at this point: Professor Fölling-Albers, Regensburg; Professor Hijting (a Dutch, but German-speaking expert); Professor Scheunpflug, Nürnberg, Professor Horstkemper, Potsdam (therefore recommended from two different quarters) and Professor Faust, Bamberg. As Professor Manfred Prenzel, director of the Leibniz Institute in Kiel, said during our interview on 16 August 2008:

> I had spoken to Jürgen Baumert about the procedure a long time in advance, before the whole thing was set up. We were in contact and thought whether it might make sense to talk about certain constellations and people we could invite. So there were always discussions going on in the background.[40]

By 10 February 2003 (*evalag* document on recruitment status), the list of experts who had accepted included the professors Wijnen, Baumert, Tenorth, Prenzel, de Corte and Fölling-Albers, with new acceptances from Professor Ingrid Gogolin, Hamburg and Professor Heinz Mandl, LMU Munich. Professor Horstkemper's appointment was still unclear, as was that of Professor Fornefeld, who had been approached, but had yet to accept. Konrad Horstmann was down on the list as the representative from the Ministry of Culture, Youth and Sport. At the bottom of this document in handwriting are the following subject areas, which were still not represented in the group: social pedagogy, a non-subject expert (although we recall that Schreier had already asked Erichsen in the previous recruitment

[40] Ich hatte im Vorfeld relativ lange mit Jürgen Baumert über das Vorhaben gesprochen und wir hatten lange im Vorfeld, bevor die ganze Geschichte etabliert war, hatten wir schon Kontakt und haben überlegt, ob das sinnvoll ist, irgendwelche Konstellationen und zum Teil über Leute gesprochen, die man einladen sollte, also, da gab es auch immer Gespräche im Hintergrund.

round), and a so-called ‘Praktiker’, with the name of Bolius in brackets. Frau Bolius was a member of the Ministry of Research, Science and Art who would provide the name of someone in this category. Terhart did not appear at this point, whose name had just been put forward to the Ministry by the evaluation’s organiser, Dr. Martina Röbbecke (letter, 06.02.03) as being a specialist in school education.

In a document fourteen days later (*evalag* document, 24.02.2003) the recruitment list was looking more firm and comprehensive. It now included Wijnen as chair with Baumert and Horstkemper as members of the sub-committee, and the following others: De Corte, Fend from Zurich (still pending acceptance), Fornefeld, Fölling-Albers, Gogolin, Klemm, Mandl, Prenzel, Simmel-Joachim, Tenorth, Terhart, with Erichsen as the non-specialist.

An initial pre-constitutional meeting between Wijnen, Horstkemper, Knorr, Herberger, Bolius, Schreier and Röbbecke was held on 26 February and recorded by Bolius (minutes, 28.02.03). Baumert was not present at this meeting during which the motives for the evaluation, overall desiderata and purpose were discussed. The commissioning body, the Ministry of Research, Science and Art, stated categorically that the evaluation was not to concern itself with cost-cutting matters, rather it should examine quality, resources and structures. The tension between Educational Science as an independent discipline and its service role in teacher raining should also be scrutinised. At this meeting it was decided that the following should also be approached: Beck (St. Gallen); Altrichter (Linz); Phillips (Oxford); Kaase (Bremen) and Zimmerli (Witten-Herdecke). The commission was asked to consider the role of Special Education, which was a particularly strong field in Baden-Württemberg. Economic and vocational education and social work should also be included.

Beck, Phillips and Altrichter quickly accepted, although the latter had to later step down from the commission because of conflicting engagements. Professor Frank Achtenhagen accepted to fill in the gap in expertise in business and human resource education.

4.4 Ministry's criteria for selection of experts

At this stage it seems useful to examine what the Ministry was looking for in its appointment of experts in this particular evaluation. If we consider who was selected for the task, then a few factors appear relevant. Those chosen were selected for their reputations in terms of both teaching and research. Without exception they were at the forefront of their respective areas of expertise. Furthermore, the majority, if indeed not all, were experienced in university administration. Most had occupied positions such as dean, chancellor or pro-vice chancellor; many had been involved directly with the DFG, which is responsible for the allocation of funding to research projects. Additionally, many had experience of working abroad and quite a few, although not all, were already experienced evaluators. Some even had international reputations as evaluators (e.g. Baumert, De Corte, Phillips, Wijnen). The average age of these evaluators was late fifties, early sixties, and a few, including the chair, were officially emeritus and in retirement, although still highly active in the university environment in various capacities.

I asked Frau Bolius about the optimum desiderata when looking for members for such an expert committee and what the Ministry's selection criteria were. She replied as follows (interview, 10.04.08):

> First of all we try to find someone for each of the larger subject areas. […] Then we have to make sure that the commission does not get too large. As Ministry of Science we have a certain amount of experience in who has profiled themselves in Baden-Württemberg and elsewhere within the research landscape. Then we ask professors whom they might recommend and then of course we try to put the committee together according to principles of equity so that it is as balanced as possible in its subject areas. And of course to ensure that we don't have people from Baden-Württemberg to judge each other. So we also need a regional balance.[41]

[41] Zunächst mal versucht man, für jede größere Fachrichtung jemand zu finden. Also, dann muss man schauen, dass das Gremium nicht zu groß wird […] Dann haben wir als Wissenschaftsministerium natürlich auch eine gewisse Erfahrung, wer sich in der Baden-Württembergischen, also, wer sich in der Forschungslandschaft auf bestimmten Gebieten hervorgetan hat. Dann fragt man natürlich auch Professoren, wen sie empfehlen könnten und dann versucht man natürlich auch, so ein Gremium nicht zuletzt auch unter Gleichstellungsgesichtspunkten zusammen zu setzen. Dass es halbwegs ausgewogen ist, fachlich ausgewogen. Natürlich auch, dass nicht Baden-Württemberger Baden-Württemberger beurteilen. Also, dass man auch eine regionale Ausgewogenheit bekommt.

However, as this evaluation would be an examination of both teaching and research she went on to describe the further criteria which they were considering in this specific case:

> Research itself needed to be judged according to international criteria and so we required an international participation in the commission. And we also had the Ministry of Culture involved, Herr Lübke, who was responsible for teacher education.[42]

Over and above the fact that the experts needed to have subject-specific credentials, it was important to be considerate of the people involved:

> The composition is also sensitive because one knows that […] there can be animosities. […] You always have to be careful to consider if the experts have ever been employed by any of the institutions that are to be evaluated… [43]

She was also clear about the fact that the concept of the teacher training college retained in Baden-Württemberg was one which not all evaluators might be sympathetic to. A balance of opinions therefore needed to be guaranteed. As the process of being evaluated was a recent phenomenon in the state the Ministry also wished to ensure maximum credibility in terms of who was appointed to the commission: […] 'we did think that people with high reputations in their fields would have greater acceptance'[44] (see also Patton, 2002).

The Ministry had a particular opinion regarding the function of the non-specialist expert in the commission:

> We expect the non-specialist expert to break through subject blindness a little, so to speak. That by looking in from the outside they can ask uninhibited questions, questions which the subject experts might no longer ask because they see them as taken for granted and so on. […] And that depends largely, I think, on the individuals themselves, if they can fulfil this expectation, or if they withdraw and say, I can't judge that, I can't help in that case.[45]

---

[42] Die Forschung selber sollte ja auch unter internationalen Gesichtspunkten beurteilt werden und deshalb auch eine internationale Beteiligung im Gremium und dann hatten wir ja auch noch das Kultusministerium, den Herrn Lübke, im Gremium drinnen, da er ja für die Lehrerbildung zuständig ist.

[43] Ich meine, die Besetzung ist auch sensibel, weil man natürlich weiß, dass es […] aber auch Animositäten gibt. […] Man muss auch darauf achten, ob die Gutachter irgendwann mal in den zu begutachtenden Institutionen tätig waren und solche Dinge.

[44] […] wir [sind] schon davon ausgegangen, dass fachlich renommierte Leute eine höhere Akzeptanz haben.

[45] Von der fachfremden Person erwartet man, […] dass sie quasi die Fachblindheit ein bisschen durchbricht. Ja, also, dass sie einfach quasi mit dem Blick von außen unbelastet Fragen stellt. Fragen, die sich Fachleute oftmals gar nicht mehr stellen, weil sie es schon so selbstverständlich haben und so. […] Und da hängt das, denke ich mal, ganz stark an der Person selber, ob diese Erwartung erfüllt werden kann, oder ob sich jemand zurückzieht und sagt, das kann ich nicht beurteilen, da kann ich nicht helfen. Ich denke, das hängt sehr stark von der jeweiligen Person ab, ob das klappt, oder nicht.

There is the implicit suggestion here that a non-specialist expert would need to be possibly even more vocal in his contributions than ordinary members of the commission. At very least he should be someone who had shown himself to be forthright and not afraid of asking occasionally 'naïve' questions. We will later learn in Chapter 5 what the non-specialist thought of his own role.

In terms of how the Ministry located the names of experts in this field, Frau Bolius confirmed what the archives had already revealed in that names were put forward from a number of sources: the Ministry, institutions themselves, *evalag*, but also from some of the experts already appointed. She would gather the names and coordinate the appointments. One of the problems encountered was the fact that while many might be suitable, not all were able to take part in the commission. She stated that in order to put together a commission of this kind it was necessary to have at least one third more good suggestions than could be used.[46]

4.5 Context preceding constitutional meeting 16 April 2003

Although *evalag* had already organised and conducted a few independent evaluations before 2003, due to the fact that this commission had been appointed by a Ministry, the evaluation had a different tenor from the outset, both in terms of its organisation, but also how it was perceived by the institutions to be evaluated. Additionally, there was a certain amount of tension between *evalag's* governing body and the Ministry, as the governing body wished to retain its independence, although in this case it was felt that the agency was bound to do the Ministry's bidding. One of the main points of contention was that the Ministry wished to attend meetings and even accompany the commission during their onsite visits, something which the governing body regarded as unacceptable. Problems of conflicting interests among stakeholders in evaluation work have been addressed in the literature (Brandt, 2007, p.179; Rossi et al., 2004, p.374 and Stockmann, 2006b, p.270).

[46] Deswegen muss man eigentlich, um so ein Gremium zu besetzen, also mindestens ein Drittel gute Vorschläge mehr haben, als dann nachher umgesetzt werden.

The experts on the commission were not aware of these problems when they were appointed, nor were they aware of the tensions between the governing body and the Ministry. However, when asked about how they viewed the role of the Ministry, the responses were mostly favourable, although we should remember that the Ministry did not give them much choice in the matter as it remained adamant about its accompanying role. This comment from Expert E is representative:

> [It was] exciting and critical at the same time as they were the commissioning body and had decided to engage two of their members. They participated as listeners, but also sometimes took part in the discussions, which was regarded very critically at first, by myself included. However, as the evaluation progressed, I have to say that this idea of being passively accompanied by representatives of the commissioning body gained ground with me. I was very sceptical at first as I feared that independence would not be guaranteed. That wasn't the case. In fact, it even helped to firm up the contract in those cases where it wasn't so clear and I experienced that in a very positive manner. So, my initial great scepticism made way for an acceptance of the status quo […], but I kept having problems where they restricted the scope of our evaluation.[47]

There was a certain degree of scepticism about the Ministry's role within the evaluation, and this scepticism also existed in the institutions being examined. As Schreier mentioned (interview, 19.12.07): 'The mistrust of the institutions was fuelled by representatives from the Ministry'[48], an understandable point, as the HEIs felt that they were being threatened in their very existence. Indeed, such was the extent of *Angst* that when the experts later arrived at the PH Freiburg, for example, they would be 'welcomed' by banners and large hordes of students proclaiming that they wished to save their institution, as they were under the impression that the evaluation had come to close their college, which was, of course, completely wrong. Morley (2003) has addressed aspects of mistrust and fear among various stakeholders in evaluation processes and reflects that because evaluators

[47] […] spannend und zugleich kritisch fand ich, dass es ja aufgrund dessen, dass sie den Auftrag gegeben haben, und für sich ausbedungen haben, selbst mit zwei Mitarbeitern beteiligt zu sein. Und zwar als Zuhörende beteiligt zu sein, zum Teil aber in den Diskussionen beteiligt zu sein, das war am Anfang, ja, ist das sehr kritisch aufgenommen worden; auch von mir. Im Verlauf der Evaluation muss ich sagen, dass diese Idee der Begleitung, aber der passiven Begleitung dieser Evaluation, durch Vertreter des Auftraggebers... das hat bei mir an Boden gewonnen. Am Anfang war ich sehr skeptisch, weil ich befürchtete, dass da die Unabhängigkeit nicht gewährleistet sei. Das war aber nicht der Fall, sondern es hat eher dazu beigetragen, dass man den Auftrag immer dann, wenn er nicht so klar war, konkretisieren konnte und das habe ich sehr positiv erlebt. Also meine anfängliche große Skepsis wich nachher eher einer, ja, einer Akzeptanz durch die Gegebenheit der Umstände, dass man Sachen, die unklar waren vom Auftrag her, im Verlauf der Evaluation klären konnte. […] ich habe immer dort Probleme gehabt, wo sie unsere Evaluationsperspektiven eingeschränkt haben.

[48] Das Misstrauen der Hochschulen wurde genährt durch Vertreter des Ministeriums.

can be perceived as being the intermediaries of commissioning bodies (in this case the Ministry), HEIs may feel 'cheated or defrauded'.

On behalf of the commission and in their capacity as members of the sub-committee Baumert and Horstkemper therefore wrote to the Ministry asking for clarification on the precise nature of ministerial participation as they feared that the commission could be manipulated to compile an acquiescent report (*Gefälligkeitsgutachten*). In interview (17.07.07) Horstkemper said:

> […] it was very important to guarantee and secure the independence of the group in the initial stages. And then there is the question of when the commission needs times during which the Ministries and their associated members do not take part, allowing the commission to be on its own, but where we can still ensure exchange of information.[49]

The argument for having ministerial members involved was explained by Ulrich Lübke, the permanent representative from the Ministry of Culture, Youth and Sport as follows (interview 20.4.08):

> The evaluation of Educational Science in the universities and teacher training colleges concerned to a very large extent the provision of Education within the training of teachers. And we had to expect that because all of the evaluators came from outside Baden-Württemberg they would have many questions. What does teacher training look like here? To what extent is Educational Science incorporated? To what degree is the contribution of Educational Science steered? And by which Ministry?[50]

It was therefore apparent that there were still numerous aspects that needed clarifying before the evaluation could proceed any further.

4.6 Constitutional meeting 16 April 2003

The first plenary meeting that took place in Mannheim was attended by the following members (see Table 2 overleaf):

---

[49] […] diese Unabhängigkeit des Gremiums zu garantieren und zu sichern das fand ich in der Anfangsphase ganz wichtig. […] Also, das ist dann die Frage, wo braucht die Kommission beispielsweise dann auch mal Zeiten, wo die Ministerien, ministeriellen assoziierten Mitglieder nicht teilnehmen, sondern wo die Kommission unter sich ist. Und wo dann […] aber gleichzeitig […] der Informationsfluss gesichert [ist].

[50] Die Evaluation der Erziehungswissenschaften an den Universitäten und an den Pädagogischen Hochschulen hat ja in einem ganz großen Umfang auch den Beitrag der Erziehungswissenschaften im Rahmen der Lehrerausbildung betroffen. Und man musste damit rechnen, da die Evaluatoren alle nicht aus Baden-Württemberg kamen, dass sie viele Fragen haben. Wie sieht die Lehrerausbildung hier aus? In welchem Umfang sind die Erziehungswissenschaften einbezogen? Wie wird der Beitrag der Erziehungswissenschaften gesteuert? In welchem Ministerium?

Professors	**Ministry**	***evalag***
Dr Wynand Wijnen (chair)	Tania Bolius	Dr Jörn Alphei
Dr Jürgen Baumert (sub-committee)	Dr Klaus Herberger	Dragana Molnar
Dr Marianne Horstkemper (ditto)	Ministerialrat Dr Heribert Knorr	Dr Martina Röbbecke
Dr Herbert Altrichter	Ministerialrat Gerber	Dr Gerhard Schreier
Dr Erwin Beck		
Dr Barbara Fornefeld		
Dr Maria Fölling-Albers		
Konrad Horstmann (Ministry of Culture, Youth & Sport)		**Excused were:**
Dr Heinz Mandl		Dr Frank Achtenhagen
Dr David Phillips		Dr Erik De Corte
Dr Manfred Prenzel		Dr Hans-Uwe Erichsen
Dr Monika Simmel-Joachim		Dr Ingrid Gogolin
Dr Heinz-Elmar Tenorth		Dr Ewald Terhart

Table 2: Participants in plenary meeting 16 April 2003

Professors Renate Walthes and Andreas Schelten were yet to be asked to participate. Konrad Horstmann as the permanent representative from the Ministry of Culture, Youth and Sport would soon be replaced by Ulrich Lübke.

The constitutional meeting was extensive and lasted the entire day (9.30am-4pm). The schedule was as follows (*evalag* document, 16.04.03; minutes 09.05.03/AZ: MR 132-03): Schreier made the introductions on behalf of *evalag* outlining the supporting role of his agency, but emphasising that on this occasion they were working on behalf of the Ministry. Heribert Knorr introduced the commissioning Ministry and briefly described the history leading up to the evaluation by stating that even before the PISA results were known, it

was apparent that the capabilities of Educational Science and its possible contribution to teacher training and empirical educational research were being critically debated. PISA had only served as an additional motor to drive the debate forwards more rapidly. He stressed that it was not the Ministry's intention to make financial cuts, since the finances of Baden-Württemberg's HEIs were fixed by a 'solidarity pact' until 2006. The Ministry did, however, wish to learn more about the university structures in both teaching and research and check up on its teacher training capabilities overall. Resources might be reallocated, but they would not be reduced. Knorr also made it very clear that the commission could not recommend a fusion of *Pädagogische Hochschulen* into universities as the *Land* had made a firm decision to keep the existing structure in place.

Professor Wijnen then introduced the experts and asked whether the evaluation should evaluate Educational Science as an independent discipline and its contribution towards teacher training, or whether teacher training per se should be evaluated. It was agreed that the contract could not encompass an examination of course contents as this would increase the extent of the evaluation considerably. In order to determine how far Educational Science was contributing to the training of teachers it would, however, be necessary to examine its relationship with subject didactics and look at the structures in place within the discipline.

Baumert discussed the criteria according to which they would evaluate Educational Science in both universities and teacher training colleges. He differentiated three functions of the subject:

1. the present status and extent to which Educational Science could be a researching discipline
2. what it was achieving in the main courses (Diploma/Masters)
3. its achievements within teacher training, differentiated according to university or teacher training college.

The main criteria in terms of research (using criteria developed in The Netherlands) would be quality, productivity, relevance, and feasibility. The group would also examine if the central sub-disciplines of Educational Science were properly represented within the *Land*. Quality would be the key criteria in terms of evaluating teaching. To achieve this they would look at process and result indicators (examination results and graduates' destinations). What the commission would not be able to do was determine the quality of individuals' teaching in the HEIs.

The group agreed on the great importance of the state's legislation with regard to teacher training, details of which the evaluators would receive in due course. Courses would be examined with regard to their relevance, their modularisation, and where possible, statements should be made about the competences and quality of what they offered. The criteria for these would be determined at a later stage.

Horstkemper led a comprehensive discussion on the kinds of questions which the group should ask the HEIs, which stimulated numerous changes in the existing questionnaire (the basis of the HEI self report) which *evalag* had provided as a template from its previous evaluations. They determined that the questionnaire should be directed at the following groups: university heads, deans and professors of Education. Specifically, it would ask about the significance of Educational Science within the institution, the extent of its organisation and its perspectives, and finally, what significance Educational Science as a main course of study had in comparison with other disciplines and which perspectives the HEI could see for it.

The questionnaires would be addressed to the heads of each HEI, who would then forward these to deans to complete the general sections encompassing research, main study course, supplementary course of study, and teacher training. Professors would then complete the subject-specific entries including details on publications, funding, and how far they were

involved in examining. The questionnaires/self reports were to be completed and returned to *evalag* by 1 August 2003.

After lunch Wijnen addressed the logistics of the evaluation, speaking about its duration, how onsite visits would be conducted and interpreted, and what the individual responsibilities of the experts would be and how *evalag* would support them. Each institution would be visited for one day only within three one-week blocks, allowing five institutions to be visited per week (now 15 HEIs in total). On completion of the onsite visits there would be a meeting with representatives from the HEIs (14-16 January 2004) in which the results of the evaluation would be discussed with each in turn, a procedure which Martina Röbbecke described as 'a very unusual instrument' (interview, 09.01.07)[51]. However, Stockmann (2006b, p.289) has referred to the fact that so-called concluding workshops (*Abschlussworkshops)* in which evaluators meet with the other stakeholders involved at the end of an evaluation can be very effective, because by speaking to them in person evaluators increase the likelihood of evaluation results and recommendations being accepted and implemented. Guba & Lincoln (1989, p.54-56) also believe that involving stakeholders at the last stage of an evaluation increases the likelihood of findings being implemented.

Wijnen then moved on to discuss future dates and the preparatory meeting of the expert group before their onsite visits and the initial composition of these groups. The next meeting would take place in Mannheim on 25 September during which the onsite visits would be prepared and groups for the various HEIs allocated. *Evalag* would help to put the visiting parties together which would each consist of five members in total. Some experts would function as the 'report writers' and would therefore play a greater role in the compilation of the final evaluation report. Their individual HEI reports from each HEI should be ready one week in advance of the plenary meeting in which their findings would be discussed (10-11 December 2003). As the Universities of Karlsruhe and Stuttgart

[51] […] ein ganz ungewöhnliches Instrumentarium.

offered Business Education and vocational teacher training and Technical Education respectively, the group decided that an expert in Technical Education should also be appointed to the commission.

Date	**Evaluation phase**
16 April 2003	Constitutional meeting
Early May 2003	Questionnaires to be sent to HEIs
By 1 August 2003	Questionnaires to be returned to *evalag*
25 September 2003	Preparatory meeting for evaluator group
20-24 October 2003	Onsite visits
3-10 November 2003	Onsite visits
17-21 November 2003	Onsite visits
10-11 December 2003	Plenary discussion of onsite visits and structure of final report
14-16 January 2003	Final plenary meeting with representatives of HEIs
Spring 2004	Compilation of final report

Table 3: Evaluation schedule (as of 6 May 2003) (Anlage 1, AZ: MR 132-03)

How did the experts feel about the day's events? Did they think that they had been given enough information about what lay ahead of them? What did they think about each other following this first meeting? If we turn to Phillips's diary entry of 16 April 2003 we can learn at least how one expert felt. It is indicative of the number of unanswered questions that were still pending at this early stage in the evaluation and it also shows that it was only now that the group learnt about the extent to which their perspective should be directed towards the role of empirical educational research, something which many had not been aware of when they accepted the task. This extract also reveals something of the problematic nature behind the developing nature of the evaluation's methodology, as agreement still needed to be reached on some of the questions relating to the organisation of the work (cf. Guba & Lincoln, 1989, p.263):

> Herr Schreier outlines the work of *evalag*, and reminds us that we have been appointed by the Minister. Herr Knorr, from the Ministry, tells us about the imperatives which here led to the commission being formed. Not just the PISA results, it seems (though that is certainly one point), but decisions taken by the *Land*

> parliament to do with the structure of higher education and with the need to develop empirical research in Education. We are told that financial support for the universities and PHs is guaranteed.
>
> There is a lot of discussion about the commission's brief. The main concern seems to be that it is too large and complex and that any investigation will have to differentiate between the universities and the PHs. We move then to a detailed consideration of the draft questionnaire and decide on some basic restructuring so that discrete sections deal with 'education' and teacher training. It looks as though the text will have to go through a few more drafts before it is ready. Then the institutions will have three months or so in which to provide what will be a great deal of complex information. Who will provide it? The presidents/rectors? The deans? The professors? Probably a mixture of everybody. We decide to collect a lot of data (and publications, for example) from individuals, and I propose the British model of asking for the best four publications from each person. […] Much is made, too, of seeing that which figures largely in the RAE assessment – the need to have coherent research strategies.

This last point would be one which the evaluators would repeatedly return to when making suggestions to HEIs about the specific direction or focus they might like to take in the future. This also raises the point that German HEIs were beginning to move away from the 'comprehensive' notion of equal standards and equal courses, to one of differentiation and increasing specialisation.

4.7 Compilation of questionnaires A and B for self report

When the evaluators received a copy of the minutes from the first plenary meeting they were also sent the standard *evalag* questionnaire which had been used in all of its previous evaluations. However, for the purposes of this evaluation the existing questionnaire needed substantial modification before it could be distributed to the HEIs. In advance of the constitutional meeting Röbbecke had already started working on the new questionnaire together with Baumert, Horstkemper and Wijnen, and she had a first draft ready by 10 April, which everyone received at the first plenary meeting. Substantial feedback was immediately provided by Phillips. Within 24 hours he had compiled a comprehensive list of suggestions on points which he felt might be considered in a revision of the questionnaire (email, 17.04.03). It is useful to examine the kinds of considerations he had, as this throws light on one expert's notions of quality, which although not necessarily representative of the entire group, does provide us with a fascinating insight into the manner in which these points were debated. The archives have not revealed further

documentary evidence of feedback from the other experts, and so we will have to rely on his suggestions alone here. The following examples are taken from three sources: the original draft questionnaire (10.04.03), Phillips's suggestions to Röbbecke (email 17.04.03), and the final questionnaires A and B, both of which are provided in translation in Appendix F.

4.7.1 Professor Phillips's comments on questionnaire draft of 10 April 2003

Although he did not discuss each of the points contained in the draft questionnaire, Phillips's comments show us how far this particular expert involved himself in the minutiae of the wording and therefore his understanding of notions of quality in HE. Furthermore, his suggestions highlight the kind of information that he, as one of the foreign experts, was able to feed into the proceedings, coming as he did from the British university environment with a slightly different tradition. I shall go through the suggestions and reveal the extent to which some of his ideas were adopted and incorporated into the final questionnaires that were distributed to HEIs. The following section moves from point to point in the order in which he chose to comment on them, and so the subject matter is quite diverse and not necessarily logically connected.

Phillips's initial reaction to the questionnaire was one of consideration for the institutions involved. Question 1.1 in the draft requested a one-page summary of each HEIs' historical development, which he felt would be 'off-putting'. In view of the fact that each institution usually had some kind of written documentation on its own history, the final questionnaire asked them to enclose appropriate material instead of summarising themselves in this manner. With regard to student participation on international programmes, he thought that draft question 3.5 needed to be more specific, hence his suggested inclusion of the internationally highly-respected Erasmus and Socrates exchange programmes. By including reference to these, HEIs would see these as benchmarks and something that they should be thinking of for their students in the future if they had not already done so. This

was incorporated into questionnaire A 3.12., which now included questions about the precise number of students involved in such international programmes.[52]

Phillips's understanding of professional development was inevitably influenced by his experience of the British system in which it had become standard practice for universities to offer further development programmes for their staff. Regarding draft question 3.6, which asked which didactical approaches and innovations in teaching were being used in the HEIs' teaching[53], he commented as follows:

> […] it might be helpful to find out if there is a notion of 'professional development' among teaching staff. Do they have opportunities to develop professionally in terms of being informed about the latest developments in 'Hochschuldidaktik' and related issues? We have a lot of pressure to do this in England, though there is much resistance. I spent a whole day recently being 'trained' how to conduct interviews for academic posts. In future only those trained to conduct such interviews will be allowed to take part in them.

This suggestion was later integrated (see Questionnaire A 4.8, in Appendix E)[54]. Draft question 4.4 had only asked when internships would take place and how they were supervised.[55] Phillips believed that this question was not comprehensive enough:

> I think we should find out a lot more about the nature of the 'Praktika', what partnerships are established with schools and with the officials responsible, what involvement the teaching staff has with the students' practical experience – and most significantly, what exactly do the institutions mean when they keep saying that students' experience is 'praxisnah'? This is a word that needs a lot of unpicking.

Questionnaire A, 3.11 in the final version now appears in a far more differentiated format: 'In which phases of different courses are internships envisaged? Are internships obligatory? What is your concept and how are these internships supervised?'[56] By asking an institution to describe its concept on internships, the evaluators would find out far more

[52] […] Wie werden die Studierenden darauf vorbereitet (Stichwort Fremdsprachenkenntnisse)? Wie viele Studierende haben daran in den letzten fünf Jahren mit welchem Erfolg teilgenommen? Wie viele Studierende haben in den vergangenen fünf Jahren an EU-Programmen (z.B. Socrates, Erasmus) teilgenommen?
[53] Welche hochschuldidaktischen Ansätze und Innovationen in der Lehre werden in Ihrem Fach praktiziert?
[54] Welche hochschuldidaktische Ansätze und Innovationen in der Lehre werden praktiziert? Werden Ausbildungsprogramme in der Lehre angeboten, und wie sind sie gestaltet?
[55] In welchen Phasen der verschiedenen Studiengänge sind Praktika vorgesehen? Wie werden diese Praktika betreut?
[56] In welchen Phasen der verschiedenen Studiengänge sind Praktika vorgesehen? Sind die Praktika vorgeschrieben? Welches Konzept haben sie, und wie werden diese Praktika betreut?

about the HEIs' approach to this highly relevant area in a student's training and the weight they gave to a student's participating in internships overall.

In terms of absences from institutions which might have an impact if a professor or member of staff was away for a considerable amount of time, the draft asked HEIs to explain longer absences, but did not specify what it meant by 'longer'. By leaving the question open in the final version 'Explain absences' the HEI would now be required to provide information on *any* absences, thereby providing a clear picture of the level of engagement of its staff members.

Although female student participation rates are lower in some disciplines, the reverse is true in Educational Science in which there are far more female students than male, a point which Phillips picked up on and asked to include. The final questionnaire took this into account in question 5.3.1 by asking 'What possibilities do you see to make Educational Science more attractive to men?

Draft question 11.1 asked how the HEI rated its own standing in teaching, study and further education. In a similar vein draft question 21.4 asked how the HEI rated its research in international comparison[57]. Phillips commented: 'There might be ways in which the institutions could be guided as to how they would categorise their achievements on a scale of some sort.' Although this was a useful suggestion, this was not taken up as a point for consideration in the final questionnaire.

In the draft there was no reference to how much formal research methodology training was being provided. Phillips wrote: 'Given the interest of the Ministry in promoting "empirische Forschungsmethoden" this seems an important question.' Again he referred to the British context as a point for comparison:

[57] Q.11.1 Wie schätzt das Fach (das Institut) die eigenen Leistungen in Lehre, Studium und Weiterbildung ein?
Q.21.4 Wie schätzen Sie die Forschungsleistungen Ihres Faches (an Ihrer Hochschule) im internationalen Vergleich ein?

> We have a lot of experience of this in England, since the principal research grant-giver in the social sciences (the Economic and Social Research Council – ESRC) requires that universities provide research training courses with specified minimal content, otherwise they will not provide studentships.

As there is no direct equivalent to the ESRC in Germany and there is no stipulation that HEIs provide such research methods training, this particular suggestion was not adopted. It could be argued that this is an area which German HE needs to address, however.

Phillips made a particularly important point regarding publications, as publication output is frequently held to be an indicator of quality in HE. The draft merely asked for the number of publications over the last five years and was not specific about the type of publications or whether these had been peer-reviewed or not (draft question 17.1). Phillips felt that this in itself was not very indicative of the quality of the work being produced, and so he commented:

> I personally believe that the number of publications is not very revealing. Instead it would in my view be much better to have returns from individuals, and for each person to list his/her 'best' four or five publications over the past five years. They might be categorised (article in international refereed journal, article in refereed German journal, book chapter, monograph, etc.). I can let you have the categories used in England if that is helpful to the group.

Question B 2.1 appeared in far more specific format in its categorisation of published material (see Appendix E for all categories), allowing for a more differentiated understanding of what was actually being produced by each HEI. Here we see categories which cover the entire range of publication type from monograph, to internationally peer-reviewed paper, to non-reviewed papers.

Phillips referred to one of his colleagues on the commission, Baumert, who had indicated during the plenary meeting that an institution should have a 'properly defined and coherent research strategy'. Question 17.2 in the draft asked whether there was a common discussion about strategies for publication.[58] Phillips commented: 'This is among the most significant criteria used in our research assessment (RAE) exercise in England.' This

[58] Stellen Sie gemeinsame Überlegungen zu Publikationen an?

would appear in a far more expansive form in the final Questionnaire A (Appendix E, points 2.1.-2.1.5.).[59]

In the draft there was no mention of sabbaticals, which might give an indication as to individual involvement with creating or developing networks, increasing individual knowledge in a particular area, or creating and maintaining partnerships with other HEIs. Phillips asked as follows:

> Could the question of sabbatical leave be included here? What is the entitlement? What are the difficulties in taking such leave? More important – do the people granted such leave have to report afterwards on what they have done? And if they have done not very much, does this affect decisions about their entitlement to leave in the future? This is an important matter of accountability.

This suggestion was incorporated as point 4 in the final Questionnaire B (Appendix E).

One of Phillips's points referred to how research might be judged. While it did not feed into the final questionnaire directly it was picked up in later discussions when the evaluators had to think about the quality of the research output in the HEIs under investigation. Regarding point 21.4 in the draft which asked HEIs to rate their research in international comparison he wrote:

> This is one of the most significant questions. It might be useful to provide categories for the response, again along the lines used in England. I can provide the exact categories, but they would be something like: 5. most research of international significance; 4. all research of national significance; 3. some research of national significance; 2. no research of national significance; 1. insignificant research record. The actual categories […] are rather more sophisticated than this, and they are for the evaluators to use, rather than for self-evaluation purposes. But something along these lines might be useful here.

The fact that categories of this type existed meant that if they were fed back to the HEIs, then these might consider in future the significance of their research (or research proposals) against these criteria, notably the most important one: international impact and significance. At very least it would make the HEIs think about the research they had been conducting hitherto and whether this met any of these criteria.

[59] An welchen wichtigen in- oder ausländischen Einrichtungen war das hauptberufliche wissenschaftliche Personal in den vergangenen 5 Jahren zu Gast (differenziert nach Person und Institution)? Bitte nennen Sie alle in den Fragebogen B genannten Einrichtungen.

Point 23.1 in the draft asked 'How has your faculty discussed PISA and its consequences (in organised events, with students), and which consequences should result in terms of teacher training?'[60] Phillips indicated that to focus on PISA would make it appear to be the driving force behind the question. He suggested adding something like 'and other international studies of the performance of education systems'. This point was accepted and appeared in Questionnaire B 5.7.1 as follows: 'How has your discipline discussed international studies which judge the capabilities of educational systems and their consequences […]?'[61]

Phillips's final comment added a new dimension, something which had not been addressed or appeared so far. He wrote:

> Finally, might there be a question on the relationship between 'Bildungspolitik' and 'wissenschaftliche Forschung'? That is, should the institutions with which we are concerned be trying to assist, criticise, generally contribute to, the discussion of educational issues at a political level and thus to influence the future of education in the *Land* and beyond?

Although this went beyond the remit of this evaluation it reflects upon the ongoing debate on the extent to which research in education should be applied, thereby directly impacting on political decision-making. Saunders (ed., 2007) suggests that although much educational research can be applied, part of the problem is that the policy makers sometimes do not grasp the meaning or consequences of educational research, or, in worse cases, ignore findings if these are not suitable to the prevalent party line. It is perhaps surprising that this point was not adopted considering that it was the Ministry's precise instruction that the evaluation examine the quality and extent of empirical educational research, as it believed that results from this could indeed be of use to determining policy regarding schools and/or university programmes, course contents and so on.

[60] Wie wurde in Ihrem Fach über PISA (bzw. TIMSS und Iglu) und die Folgen diskutiert (in organisierten Veranstaltungen, mit den Studierenden), und welche Konsequenzen sollten daraus im Hinblick auf die Lehrerbildung gezogen werden?

[61] Wie wurde in Ihrem Fach über internationale Studien, die die Leistungsfähigkeit von Bildungssystemen bewerten, und ihre Folgen diskutiert (in organisierten Veranstaltungen, mit den Studierenden)? Welche Konsequenzen sollten daraus im Hinblick auf die Lehrerbildung gezogen werden?

Between 16 April and 5 May (the precise meeting date is unclear from the archives), Röbbecke and the sub-committee met in Berlin to discuss the contents and wording of the questionnaire and as can be seen from the above information, much of what Phillips had commented on would be incorporated in the final version. Once the questionnaires were finalised, they were dispatched in early May to the HEI chancellors with the request that they be forwarded to the relevant deans of the Faculty or Heads of Department who should gather collective information from members of staff. In other words, the questionnaire was not to contain individual responses, but those gained through group discussions within each department (draft letter Röbbecke (MR) to university chancellors, 05.05.03).[62] Röbbecke included some guidelines as to the length of the self reports which should not exceed a maximum of 50 A4 pages (excluding the statistical data). The HEIs should return 20 copies, which would then be immediately forwarded to the experts. This guideline was, however, not adhered to very rigorously and the experts would later find themselves reading considerably more information than had been called for.

4.8 Interim organisational details

Röbbecke sent the minutes of the first plenary meeting to the experts on 9 May 2003 (AZ: MR 132-03). Wijnen had asked her to change the evaluation schedule to allow for more time to discuss the onsite visits as he felt that two days would not suffice. As a result, both he and the sub-committee suggested turning this meeting into a three-day event. The first two days would be devoted to reporting on individual HEIs; the third would be a discussion of the report which they would then start compiling.

On 15 May 2003 Röbbecke dispatched Baden-Württemberg's examination regulations to the evaluators so that they could familiarise themselves with the legal aspects of the state teaching examinations (AZ: MR 145-03). Each federal state in Germany has its own regulations regarding examination procedures, contents, duration of examinations, etc., and

[62] Bitte leiten Sie den Fragebogen möglichst rasch an den Dekan des Fachbereichs bzw. an den Leiter des erziehungswissenschaftlichen Instituts weiter, der gebeten wird, ein kollegiales Votum der erziehungswissenschaftlichen Professorinnen und Professoren einzuholen. Der Fragebogen soll also nicht individuell beantwortet werden, sondern das Ergebnis gemeinsamer Diskussionen im Fach sein.

so it was necessary for those evaluators coming from different *Länder* to know the procedures within this state. In addition, one section of the examination regulations pertaining to the qualification of junior school teachers and teachers in the *Hauptschule* was in a process of modification and the expectation was that the draft the evaluators had received would probably be officially passed on 1 October 2003, around which time the onsite visits would be taking place. The information forwarded by the Ministry also contained the examination regulations for students wishing to become teachers in the *Realschule* or *Gymnasium* and those wishing to become Special Needs teachers, which was an area of particular emphasis in Baden-Württemberg.[63]

The HEI self reports were returned by August and immediately forwarded to the evaluators. As mentioned above, the amount of material was immense, with each HEI providing as much information as they felt the evaluators might need, which in many cases far exceeded the 50 pages requested. It was a challenge to read all of the data as the first group meeting was scheduled for the 25 September. Phillips commented in an email (08.09.03): 'This is indeed a mountain of reading! We don't have in English the proper equivalent of "Papierkrieg" […], but I can see now why this term needed to be coined…' This point will be expanded on in Chapter 7. Although a vast amount of information had been provided, various pieces were still missing. Röbbecke listed these as follows:

- Chancellor's statement from the University of Stuttgart
- Self report from the University of Karlsruhe (General Education und vocational training, replies to questionnaire A, compilation questionnaire B and addenda)
- Extra comments from the Chair of Business Education, University of Mannheim, on the answers in questionnaire A and
- A new compilation of questionnaire B from the above chair in Mannheim. Please exchange your two versions of this document, as the former has proved incomplete.
- Our data gathering is almost concluded. We are still missing the compilations to questionnaire B from the University of Tübingen, which will probably arrive by the time of our next meeting. (Letter, 19.09.03, AZ: MR 353-03)[64]

[63] The education of those with special needs actually finds part of its roots in this state. Establishments for those with mental or physical handicaps, and where the first pedagogic steps towards their treatment could be identified, were set up in the 19th century. These therefore served as models (cf. p.19, final report).

[64] Die Stellungnahme des Rektors der Universität Stuttgart; den Selbstreport der Universität Karlsruhe (Allgemeine Pädagogik und Berufspädagogik, Antworten auf Fragebogen A, Zusammenfassungen Fragebogen B und Anlagen; einen Nachtrag des Lehrstuhls für Wirtschaftspädagogik, Universität Mannheim, zu den Antworten auf Fragebogen A; und eine neue Zusammenfassung der Fragebogen B des Lehrstuhls für

Some HEIs gave the evaluators little time to acquaint themselves with the data and those who had been assigned to visit Tübingen in particular would only have a few days to read the missing information and prepare for their visit. Stuttgart sent its final documents to *evalag* on 12 September (archive number 265-03) in which it described the significance and integration of teacher training within the institution and how Educational Science as a main subject compared with other main subjects. In order to gain an overview of how Educational Science was situated and perceived within this institution, the evaluators needed to have their questions answered before going onsite.

4.9 Preparations in advance of plenary meeting 25 September 2003

The sub-committee and *evalag* worked together closely in the run-up to the plenary meeting on 25 September. Röbbecke had sent out a provisional list of topics which the sub-committee would discuss in their meeting scheduled for 11 September (document: *Tagesordnung*, 25 September 2003). She had also prepared a 10-page internal document which summarised why this evaluation was being conducted and further points which should be brought to the evaluators' attention before they conducted onsite visits. It seems relevant to examine this document, as it provides further information on the background of the evaluation which the evaluators might not have been privy to before this point (document: *Vorlage für die Sitzung der Gutachterkommission* am 25.09.03 (zu Top 6).)

Section 1 summarised the reasons for the evaluation: the ongoing debate on the quality of teacher education; poor PISA results; a lack of high quality empirical educational research and a general lack of personnel to support young academics. Röbbecke drew on information provided by the HEIs to summarise some of findings already available. In section 2, entitled 'Educational Science and its contribution to teacher training', she described some of the reform suggestions made by the HRK in 1998, the KMK in 2000

Wirtschaftspädagogik, Universität Mannheim. Bitte tauschen Sie diese Version gegen die Ihnen vorliegende Version aus, da sich diese als unvollständig erwiesen hat.
Die Befragung ist somit nahezu abgeschlossen, es fehlen nur noch die Zusammenfassungen der Antworten auf Fragebogen B von der Universität Tübingen. Voraussichtlich werden und auch diese Unterlagen bis zur kommenden Sitzung in Mannheim vorliegen.

and the WR in 2001 She focussed in particular on the points regarding the function and achievements of Educational Science as these indicate how this field was viewed by others. For example, the WR had viewed Educational Science as being not well integrated into the first phase of teacher training, subject didactics and the science of the subject overall. Röbbecke also reported that Educational Science displayed a certain lack of commitment with regard to the training of teachers and that the WR had noticed a 'distancing from training duties within the framework of teacher training' (WR 2001: 25)[65]. It was also difficult to distinguish the relationship between depth and breadth within the discipline. Overall, the self concept of individual subject didactics was unclear and the achievements of these insufficient. She also mentioned that the HRK had criticised a lack of organisational and contextual liaison within teaching courses, a point which was also supported by the KMK.

The HRK desired more connectivity between the discipline and the aims of teacher training. It described course contents as being not particularly systematic in this respect. Both the HRK and the WR stated that owing to the lack of liaison between Educational Science, subject didactics and individual subjects there was a lack of institutional identity which was responsible not only for a large drop-out rate, but also for longer study durations and a lack of orientation among students in general. The HRK even went so far as to describe the two phases of teacher training as being entirely separate and insular, which the WR agreed with, saying that there was not sufficient communication between the two phases of initial study in the first and teaching practice during the second, where students spent most of their time in schools (*Vorbereitungsdienst*). More particularly, there was no scientific control of the quality of the second phase, or clear standards on what this should be achieving. The KMK also confirmed these points in its own analysis.

[65] Der Erziehungswissenschaft wird attestiert, dass ihr Bezug auf die Lehrerbildung an Verbindlichkeit verloren habe und eine „Distanzierung von Ausbildungsaufgaben im Rahmen der Lehrerbildung" festzustellen sei.

The WR highlighted the importance of lifelong learning, although it felt that further training was presently characterised by whim and lack of cohesion. Knowledge transfer was equally insufficient between schools and universities. This point was also taken up by the HRK, which criticised that further training seemed to be totally detached from the state institutes for research and teaching in HEIs. As a final point, both subject didactics and Educational Science were characterised by a massive deficit in the quantity and quality of research, a point which the WR strongly emphasised needed to be altered.[66]

Each of these groups, the WR, HRK and KMK, provided a number of proposals for change which can be summarised as follows:

1. The disciplines needed more integration and better communication overall;
2. A better agenda for empirical research, much of which should be of international quality;
3. An improved system of further and ongoing training;
4. Improved liaison between different phases of study.

There was also discussion about the introduction of a BA/MA degree structure for teacher training. The HRK, for example, thought that the BA could provide first subject knowledge and then the Masters would include subject didactics and Educational Science, although it did not preclude the idea that Educational Science could be taught earlier. Modularisation also formed part of the discussion, with the idea put forward that a foundation course could be similar for all teacher trainees, allowing a final decision on whether to pursue a teaching career until after the intermediate examination stage (*Zwischenprüfung*). Finally, the KMK requested an external evaluation of teacher training throughout Germany in order to improve understanding of the status quo and quality of different teacher training systems in the sixteen *Länder*.

[66] Sowohl in der Fachdidaktik als auch in der Erziehungswissenschaft bestünden massive Defizite in der Qualität und in der Quantität der *Forschung*, die nach Ansicht des WR „mit Nachdruck“ behoben werden müssten.

In her description of the key topics for discussion, Röbbecke now turned towards the situation in Baden-Württemberg and she stated that the number of students taking Educational Science as a secondary subject alongside their teacher training had vastly increased during the last 30 years, although only a few took it as a main degree course. Only 1.5% of all students in 2001/2002 took this course in Baden-Württemberg compared with 2.5% in Germany overall. This contrasted with high numbers of students enrolled on teacher training courses, i.e. c.12% of all students.

She highlighted the special situation in Baden-Württemberg. Tables revealed the numbers of students registered in teacher training courses in both universities and teacher training colleges, with significant differences in numbers of students in the individual HEIs, ranging from PH Schwäbisch Gmünd with 1189 students in 2001/2002 to 3890 in PH Freiburg in the same year. We would see increases in student numbers in 2002/2003 in both HEIs (see Table 1, final report). The proportion of teacher trainees registered in teacher training colleges as opposed to universities was 60:40.

Regarding research in particular, additional tables were provided to give information on the distribution of funding, especially those projects financed by the DFG. Röbbecke highlighted the University of Mannheim and the PHs Freiburg and Ludwigsburg as having received the highest amounts of funding from this body. With regard to the numbers of students completing doctorates, the universities had produced a total of 131 over the last five years, whereas the PHs produced 79. The University of Tübingen was identified as producing the highest number of doctorates over the past five years (66), with the University of Heidelberg in second place (38). The other universities produced only few doctorates between them (Table 9a, final report). The PH Freiburg produced the greatest number of doctorates (23), followed by the PH Heidelberg (16) (Table A6, final report). In total 15 people wrote their *Habilitation* in the universities, with only four completing theirs in the PHs. In each case, although more women were studying Educational Science, significantly more men were completing both doctorates and the later *Habilitation*.

Regarding the allocation of staff in both institutional types, it was apparent that there were large imbalances, with only 30 C4 and C3 professors in the universities compared with 112 in the PHs, which also had marginally higher numbers of C2 and C1 professors than the universities (28.5 to 23.5) (Table 3, final report).

Although comprehensive and informative, Röbbecke's initial overview of the data she had received was still not complete as parts of the data had yet to arrive. In her letter to the sub-committee in advance of their meeting on 11 September she said that this 'text was a just-in-time production'[67] and that there was the ongoing problem of gathering complete data sets from the institutions. Furthermore, the Ministry had requested each of the 15 self reports, a point she wanted to discuss at the meeting. In a similar vein she also wanted to address the Ministry's wish to be present during the onsite visits and quoted the experience of the WR where representatives of the government or from the *Länder* attended onsite visits, but did not participate in internal group discussions.

Regarding the responsibilities of those experts chosen to write the onsite reports, Röbbecke said that Wijnen had suggested each of the evaluators should send in the questions they believed the institutions should be asked. These would be bundled by the agency and distributed to the group.[68] The 'reporters' would also present their onsite reports in the December meeting and then again to the HEIs in the January meeting. One logistical detail was that the onsite plan, which *evalag* had compiled, did not suit every HEI, and so it would need adjusting in those cases where two subjects, for example, needed evaluating. She suggested that in these HEIs the evaluation group would have to split up.

In the last point Röbbecke highlighted some specific requests that had arisen from particular HEIs. PH Karlsruhe, for example, wanted its chancellor to be present at the meetings with all groups (staff, students, etc.). The University of Constance posed a special

[67] Dieser Text ist ein ‚just in time' Produkt.
[68] […] Herr Wijnen schlägt vor, dass alle Gutachter/-innen ihre (schriftlichen) Fragen zu dem jeweiligen Standort an die *evalag* schicken, sie sollen dann gebündelt dem/der Berichterstatter/-in zur Verfügung gestellt werden.

logistical problem since the Educational Science part of its teacher training took place at PH Weingarten, and so it was unclear with whom the evaluators should even conduct their meetings. Wijnen therefore suggested that three staff from Weingarten attend meetings in Constance. Another problem was caused by PH Heidelberg, where Röbbecke suggested the team be divided to cover Educational Science plus Educational Psychology, and Special Education. However, this would cause an imbalance regarding PH Ludwigsburg, which also placed emphasis on its Special Education, which was taught in Reutlingen. We see from these examples that the logistics of organising the onsite visits needed careful consideration before the evaluation team could be sent out into the HEIs, most of which needed resolving before the plenary meeting on 25 September. Let us therefore now look at how far the group had progressed when it came together again for the second time.

4.9.1 Plenary meeting 25 September 2003

The meeting took place from 10am to 4pm and included the following participants (protocol of meeting, 25.09.03, Mannheim):

Professors	**Ministry**	***evalag***
Dr Wynand Wijnen (chair)	Tania Bolius	Dragana Molnar
Dr Jürgen Baumert (sub-committee)	Dr Klaus Herberger	Dr Martina Röbbecke Dr Gerhard Schreier
Dr Marianne Horstkemper (ditto)		
Dr Frank Achtenhagen		
Dr Herbert Altrichter		
Dr Erwin Beck		
Dr Maria Fölling-Albers		
Dr Barbara Fornefeld		
Dr Ingrid Gogolin		
LMR Ulrich Lübke		**Excused were:**
Dr Heinz Mandl		Dr Maria Fölling-Albers
Dr David Phillips		Dr Erik De Corte
Dr Manfred Prenzel		Dr Hans-Uwe Erichsen
Dr Ewald Terhart		Dr Monika Simmel-Joachim
Dr Renate Walthes		Dr Heinz-Elmar Tenorth

Table 4: Participants in plenary meeting 25 September 2003

If we compare this list with that of the constitutional meeting we see that De Corte and Erichsen were unable to attend either meeting. Walthes and Schelten were new to the commission and attending for the first time. Herberger's place as the Ministry representative would soon be taken by Lübke who would assist the evaluators during their onsite visits. We might ask whether those who came to the second plenary meeting found it a disadvantage having missed the first. When I later interviewed those members and asked them about this (see Chapter 5), I discovered that there was no consensus of opinion regarding this fact, which was slightly surprising as I had imagined that they would all regret missing being part of the evaluation from the outset, for example, when in actual fact they displayed more pragmatism.

The first point on the agenda was made by Wijnen regarding the extent of the Ministry's involvement in the proceedings. As we already know from the archives, there had been an interim meeting between him, Baumert, Horstkemper, Herberger and Bolius to discuss the Ministry's engagement in the commission. It had been agreed that the Ministry would receive all documentation and its representatives would attend the onsite visits as observers. The experts would retain their independence and be able to have meetings after each visit without the presence of ministerial members. Röbbecke described this problem (interview, 09.01.07):

> Someone from the ministry was always present. It wasn't clear to me or the commission from the beginning that it was going to be like this. […] it was only as a result of the energetic protests of the commission's chair, Mr Wijnen, who, as a Dutchman, had definitely expected the process to be more distanced from the state, that process elements were introduced which created a kind of arrangement. These process elements were adapted from the Science Council, and so they had a high legitimacy. They envisaged that representatives of the MWK would take part in onsite visits, but the commission would always have opportunities to take time out to confer without a ministerial representative being present.[69]

[69] Es was immer jemand vom Ministerium dabei. Dass das so sein würde war mir von Anfang an nicht klar. Der Kommission auch nicht. […] nur aufgrund der energischen Proteste des Leiters der Kommission, Herr Wijnen, der als Niederländer ein deutliches staatsferneres Verfahren erwartet hat, sind dann Verfahrenselemente eingezogen worden, die […] so eine Art Ausgleich hergestellt haben. […] diese Verfahrenselemente wurden vom Wissenschaftsrat adaptiert, also, hatten insofern eine hohe Legitimation. Sie sahen vor, dass der Vertreter des MWK an der Begehung teilnahm, aber die Kommission hatte jederzeit die Möglichkeit eine Auszeit zu nehmen, […] also zu beraten, ohne dass ein Vertreter des MWK dabei war.

Professor Phillips's diary entry (25.09.03) also discusses this problem and describes the manner in which the rest of the meeting was organised:

> This morning we assembled at *evalag*. […] Our chairman tells us at the outset that there has been discussion with the Ministry people about their role in the commission and that it has been resolved that they will not necessarily play such an important part – they will attend a lot of meetings […] and attempts will be made to avoid the impression that they are influencing us. I am glad that this matter has been raised, since I was really struck by the extent to which the ministry representatives contributed to our last meeting – while they did not dominate, exactly, they certainly said rather more that I would have expected.
>
> The meeting proceeds at considerable speed. We discuss the arrangements for the site visits, the role of the reporters on the situation in each institution, and the shape of the final report.
>
> I wonder how it is going to turn out. There is much detail which I can't hope to come to terms with: the aim[s of] *Prüfungsverordnungen* [examination regulations] and complicated course arrangements. It is possibly best if I concentrate on questions of principle. But what are we going to recommend? And what chances are there of German professors, jealously guarding their academic freedom, taking any notice of it?
>
> A few things are already striking in the responses we have received. Most people say they are overworked and have insufficient resources generally; research income – even for the more famous institutions – is very low; there are few publications in journals with international standing. The PHs in particular, however, seem to have reasonable conditions, certainly in some cases in terms of staffing – and so why is the research effort not more impressive? Baumert says some very perceptive things and makes me think in terms of looking at the potential of the existing resources – i.e. not being negative about the obvious failings, but positive about what might be achieved.

This last comment is indicative of the overall tenor of how the evaluation team approached their task, namely that they did not wish their evaluative statements to be of a punitive or negative nature, but that they wished to highlight what was working well and suggest what else could be done to improve existing structures. Professor Phillips's entry also shows the difficulty in managing such large amounts of information. I will be addressing this point in my next chapter, where I examine what the evaluators said about their work.

Röbbecke reported that the HEIs had been given two questionnaires, A and B, which were divided into two sections. The first (Section A) reported on the specifics of individual subjects, whereas Section B looked at the output of individuals, data from which were returned anonymously to protect the identities of those responding. She reported that the data collection was incomplete and that the HEIs would be asked for further data regarding

the number of students registered, the number of those taking examinations, and the average length of their courses. They should also provide information on the extent of staff hours in the subjects under examination.

The second main point on the agenda discussed a number of areas pertaining to the logistics of the onsite visits. Evaluators would make their own way to Mannheim, but *evalag* would organise hotel accommodation and travel details from then on. The group would always meet at 4.30pm on the Sunday preceding the onsite week to prepare for the week's business. Röbbecke outlined the duties of those allocated the responsibility of writing onsite reports as these would describe the specifics of the HEI and collate the most important observations and experiences from oral feedback in the evenings after the visit. Those experts who were writing reports would not be burdened with the additional task of leading discussions in HEIs, a role which Wijnen offered to take over. *Evalag* would take the minutes of the onsite visits and return these to the respective report writers to assist them in the compilation of the onsite reports, the first drafts of which should be complete by early December. All of the group members were asked to return the questions which they wished to pose to HEIs to *evalag*, where they would be bundled and distributed in advance of the onsite visits.

The third agenda point presented a first draft structure for the final report, divided into a general section in which structural points and quantitative data concerning the subjects Educational Science, Special Education, and Business and Technical Education would be described, followed by a section containing reports on each HEI. The motion for this suggestion was passed, although the group felt that it might be necessary to amend the format of the report once the onsite visits had been concluded. In particular the group held that there should be more discussion on the differences and commonalities between the universities and the teacher training colleges, the role of subject didactics and the evaluation criteria. Onsite reports should be no longer than 4-6 pages (the average number

of pages in the final report for each HEI including tables on each was six). The group asked *evalag* to prepare tables on the most important data of each HEI.

4.10 Towards determining evaluation criteria for *Erziehungswissenschaft*

The fourth point on the agenda regarding the evaluation criteria was summarised thus:

> Following intensive debate on the present draft of evaluation criteria, it was agreed that Herr Baumert would rework the evaluation criteria. HEIs should, if possible, be informed of these criteria before the onsite visits take place.[70]

This brief statement belies the struggle which the group had to determine exactly what the criteria for their decision-making might be. If we examine the two-page draft of these criteria (document, *zu Top 4 der Kommissionssitzung am 25. September 2003*) which the group had been given during the meeting, we discover a two-page document which was still far from final. The criteria were divided into five points. If we take research in Educational Science, for example, then this should be evaluated according to quality defined in terms of DFG funding, relevance and scientific significance of research output, coherence and continuity, productivity (publications); and relevance of research projects, in terms of their contribution to solving educational problems and those in teacher training. However, it was still unclear which publications could be included; nor was quality defined in further detail other than those areas already specified. Chapter 6 will describe evaluator attitudes to this problem in greater detail.

Defining quality in educational publications remains a point of contention as some materials (school textbooks, for example) which have considerable impact and are probably regarded as being of high quality will not necessarily appear within lists of recommended publications, which predominantly consist of peer-reviewed journals. It is therefore an area in which determining quality criteria becomes particularly thorny, as determining quality factors does not necessarily work within Education as well as in other

[70] Nach einer intensiven Diskussion über den vorgelegten Entwurf der Evaluationskriterien wird verabredet, dass Herr Baumert den Kriterienkatalog überarbeitet. Die Kriterien sollten den Hochschulen möglichst noch vor den Begehungen zur Information zugehen.

fields. The publication-quality-impact problem persists to the present day and will in all likelihood persist in the future, unless more differentiated criteria are established for awarding 'quality points' to materials published within Educational Science that, while undeniably good, do not necessarily 'fit'.

The second point on the agenda dealt with teaching and learning and was sub-divided into six further sub-headings:

1. cohesion of teaching aims and teaching offers
2. study organisation (modularisation, examination organisation, advice and support)
3. research relevance for teaching
4. internal quality assurance
5. academic development (doctorates)
6. success (pass rate, length of study, graduate employment)[71]

The third detailed criteria on teacher training divided into how the HEI regarded the organisation of teacher training, and its course contents with regard to the schools it was training for; the other points listed emulated the five remaining points from the list in point 2 above (e.g. research relevance for teaching, etc.). The fourth point referred to academic development and its uses, while the fifth point looked at the criteria for examining the structure of the subject in terms of its material and personnel resources, cooperation, internationalisation, and how far sub-disciplines were represented. Returning to the meeting of 25 September, the fifth agenda point passed the suggestion for a meeting from 10 to 12 December during which the onsite reports and key points of recommendation would be discussed. The group also requested that their January meeting with the heads of HEIs be shortened into a two-day event.

The sixth agenda point referred to the various recommendations for reform which had resulted from the meetings of the HRK, KMK and WR. The group stated that it would

[71] […] Bezug der Ausbildungsorganisation und –inhalte auf das Berufsfeld Schule; Studienorganisation (z.B. Modularisierung, Kerncurriculum, Prüfungsorganisation, Beratung und Betreuung); Forschungsbezug der Lehre; Förderung beruflicher Kompetenzen; Interne Qualitätssicherung; Studienerfolg (bestandene Prüfungen, Fachstudiendauer).

refer to particular points in the introductory text of its final report. Some of the experts had enquired about the systematic differences between the universities and teacher training colleges, which they felt needed answering with regard to teacher training and research.[72] They were keen to see the extent to which both institutional types were able to fulfil both teaching and research requirements. They also noted the high numbers of staff involved in Special Education, a point which would be dealt with in greater detail at the next meeting.

In preparation for the meeting with representatives from each HEI to be held in January we see a very tight schedule developing, with two teams of evaluators speaking to five HEIs on one day, ten therefore in total, with 90 minutes for each HEI. The day was scheduled to commence at 8.30am and the last discussion with HEIs would finish at 6.30pm. Friday 16 January would be slightly less intense, with two teams working with three and two HEIs respectively, concluding their discussions at 1.30pm. This would leave the rest of the day for discussions, and the meeting would conclude at 5pm.

4.11 Interim correspondence and proceedings prior to onsite visits

As soon as Röbbecke had been provided with the missing data from the University of Tübingen (questionnaire B, publication list), she forwarded the information to the group (29.09.03, AZ: MR 374-03) so that they now had the complete set of materials required.

On 6 October 2003 she forwarded a comprehensive and detailed letter (7 pages, plus appendices) which had been sent to the *evalag* agency from a university professor (06.10.03, AZ: MR 386-03) in which he commented on the evaluation and the situation regarding Educational Science in his institution in particular. In the first part of his letter he painted a bleak picture of the current state of Educational Science in Baden-Württemberg and referred to the historical reasons for its present deficits (e.g. some chairs being vacated and not replaced, or reallocated). He stated that the universities and teacher training

[72] […] mehrere Gutachter (stellen) die Frage nach den systematischen Unterschieden zwischen Universitäten und Pädagogischen Hochschulen, die sowohl im Hinblick auf die Lehrerbildung als auch auf die Forschung beantwortet werden müsse.

colleges did not have a disciplinary structure in the sense of representing all of the important fields of development and research in pedagogy and Educational Science, nor did research projects form any cohesive links. Chair holders could do as they wished, regardless of political or research-linked guidelines and in total disregard of the changing needs of teacher education.[73] He stated that the evaluation presented an opportunity (if the results became public) to bring about a discussion of these problems, which might lead to a quality improvement overall. The evaluation should consider three points: to let a subject continue or be restructured; discontinue it or expand it. He criticised the questionnaires sent out to HEIs as being insufficient, which he put down to the fact that the evaluation commission might not have been given very clear guidelines as to what needed evaluating. We shall later see from the interview data that this hypothesis was indeed true. He specifically reminded the commission to be aware that they could serve a political function 'to delay the necessary decisions, which should have been taken ages ago by those responsible in politics and administration, and bind themselves into a 'political process', which is presently running idle.'[74] Evaluations can inevitably serve a political function and, as Pollitt (1998, p.215) has indicated, they can serve as 'baubles or bolsters'. Evaluations can provide governments with more effective controlling mechanisms than any other form of control (Kieser, 1999b, p.57). In other words, if a political decision has been taken before an evaluation, an evaluation can add the empirical evidence to reinforce a political argument.

The letter furthermore described in detail the efforts whereby this professor had established a Pedagogy seminar to provide a unit which would offer the best conditions for a university which itself did not offer a broad spectrum of subjects. He told the commission

[73] […] die Universitäts-Pädagogik in Ba-Wü (hat) keine disziplinäre Struktur in dem Sinne ausgebildet […], dass alle wichtigen Entwicklungs- und Forschungsfelder der Pädagogik/Erziehungswissenschaft im Lande vertreten sind, gar nicht zu reden von Forschungsprogrammen und –verbünden; vielmehr folgten die Lehrstühle und Institute ihren eigenen selbstdefinierten Interessen, größtenteils auch völlig unabhängig von Politik- oder Forschungskonjunkturen der Universitätspädagogik in der BRD und ohne Berücksichtigung der sich wandelnden Anforderungen an die Lehrerausbildung.

[74] Es sei erlaubt, darauf hinzuweisen, dass die Evaluierungs-Kommission […] Bedacht nehmen möge, dass sie auch die politische Funktion hat, die nötigen Entscheidungen, die die Verantwortlichen in Politik und Verwaltung längst hätten treffen müssen, zu vertagen und in einen „politischen Prozess" einzugeben, der im wesentlichen Leerlauf ist.

that he needed additional chairs in Pedagogy, Educational Psychology and subject didactics. He summarised his efforts thus: 'they failed both on an institutional and structural level, both in terms of teacher training overall and the desired structures in his institution in particular.[75] Against this backdrop he asked the group to define the following:

1. their aims and their criteria
2. how they wished to implement recommendations and their criteria
3. their results and their criteria
4. their perspectives for the teaching degrees in [name of HEI] and their criteria.

It might seem somewhat strange for an individual to forward such a wide-ranging amount of information to an evaluation commission, as he knew that his HEI was going to be evaluated. However, this particular professor had nothing to lose by sending this comprehensive report as he informed the group that he would soon be vacating his university post. His letter was passionate about his attempts to introduce better quality mechanisms for those in teacher training at his institution and the handicaps that he had encountered in his efforts to raise the quality within his HEI. If we compare his letter with some of the comments in the final report, we see that a number of his suggestions in terms of defining aims, results and perspectives were indeed included. The group had therefore taken much of what he said very seriously.

4.12 The Ulm Declaration 2001

As one of the appendices added to his letter, the writer included the so-called Ulm Declaration on the training of *Gymnasium* teachers in Baden-Württemberg's universities, which none of the evaluators had been given as background material prior to their work in this commission (see Appendix G). It should be noted here that this was signed by a number of professors, both within and outside of the state and two of the names appear also in our list of evaluators in 2003/2004: Tenorth and Terhart. The Declaration referred to a political decision in Baden-Württemberg whereby the university training of

[75] Sie sind institutionell und strukturell fehlgeschlagen, sowohl bezüglich der Lehrerbildung im ganzen als auch hinsichtlich der angestrebten Struktur in [name of HEI] in besonderen!

Gymnasium teachers was to be reformed. There would be a greater percentage of courses in Pedagogy and School Pedagogy, Educational Psychology, obligatory subject didactic courses and a practical semester. Although more money was to be made available for chairs in the above areas, it came as a complete surprise that the University of Constance, the PH Weingarten and the seminar in Rottweil had signed an agreement whereby Constance would restrict itself to the subject-specific courses, with the pedagogic, educational-psychological and subject didactic courses being offered by the teacher training college and the seminar for students in more advanced terms on their course. Constance had even reallocated its only chair in Pedagogy. The University of Karlsruhe and the PH Karlsruhe had signed a similar agreement. The signatories of the Ulm Declaration were concerned that if these models were copied any further then the aforesaid and state-endorsed reforms would come to nothing and the quality of teacher training for the *Gymnasium* would suffer as a consequence. Their words were unequivocal:

> The proceedings in Constance and Karlsruhe stand in gross contradiction to the most recent recommendations from the KMK and the expert recommendations from a number of *Länder* […] (*Ulm Declaration*, p.2)[76]

Although some of the other German experts may have been aware of these events, it is debatable whether the foreign experts in particular had, up to this point, been informed of the political background and present situation in Constance and Karlsruhe. The minutes from meetings do not record that Tenorth or Terhart spoke directly about the fact that they had signed the above Declaration. However, the fact that they were both signatories on the Ulm Declaration meant that their viewpoint regarding the above-named institutions was inevitably somewhat politically influenced. It might therefore be appropriate to let other evaluators conduct the onsite visits at these institutions. In order to clarify this point, we can examine the lists of onsite allocation to determine whether *evalag* had, in the organisation of the onsite teams, taken this into account to keep these particular experts away from these HEIs for the reasons stated above.

[76] Die Vorgänge in Konstanz und Karlsruhe stehen im krassen Gegensatz zu den jüngsten Empfehlungen der Kultusministerkonferenz und der Expertenempfehlungen in mehreren Bundesländern […].

4.13 Allocation of evaluators to HEIs for onsite visits

Wijnen had already offered to be present at all of the onsite visits and he took over the role of chair at most of the meetings. The protocols for the visits were written by those named in italics in the following tables which show the lists of participants for each of the onsite weeks. Three evaluators attended the whole week, whereas others were only available for some days. We therefore see Achtenhagen, Mandl and Wijnen attending week one, Fölling-Albers, Terhart and Wijnen week two, and Erichsen, Beck and Wijnen week three. Unfortunately, Altrichter had withdrawn from the onsite visits and therefore the entire evaluation, which is why the allocation of evaluators for the first week was changed and others took over the role of report-writing for which he had initially been assigned.

Monday 20.10	Tuesday 21.10	Wednesday 22.10	Thursday 23.10	Friday 24.10
Uni Mannheim	**Uni Hohenheim**	**PH Weingarten**	**Uni Constance**	**PH Freiburg**
Mandl	*Prenzel*	*Prenzel*	*Mandl*	*Mandl*
Horstkemper	Horstkemper			Horstkemper
			Horstmann	Horstmann
Mandl	Mandl	Mandl	Mandl	Mandl
			Phillips	Phillips
	Prenzel	Prenzel		
Wijnen	Wijnen	Wijnen	Wijnen	Wijnen
Achtenhagen	*Achtenhagen*	Achtenhagen	*Achtenhagen*	Achtenhagen
				Simmel-Joachim
		Walthes		Walthes

Table 5: Allocation of experts for first onsite week 20-24 October 2003 *(report writers in italics)*

Monday 03.11	Tuesday 04.11	Wednesday 05.11	Thursday 06.11	Friday 07.11
Uni Ulm	**PH Schwäbisch Gmünd**	**Uni Tübingen**	**Uni Stuttgart**	**PH Karlsruhe**
Gogolin	*Terhart*	*Baumert*	*Terhart*	*Fölling-Albers*
		Baumert	Baumert	Baumert
Gogolin	De Corte	De Corte	De Corte	De Corte
Fölling-Albers	Fölling-Albers	Fölling-Albers	Fölling-Albers	Fölling-Albers
Terhart	Terhart-	Terhart	Terhart	Terhart
Tenorth	Tenorth			
Wijnen	Wijnen	Wijnen	Wijnen	Wijnen
		Simmel-Joachim		
	Walthes	Walthes	*Schelten*	*Fornefeld*

Table 6: Allocation for second onsite week 3-7 November 2003

Monday 17.11	Tuesday 18.11	Wednesday 19.11	Thursday 20.11	Friday 21.11
PH Ludwigsburg	**Uni Heidelberg**	**PH Heidelberg**	**Uni Karlsruhe**	**Uni Freiburg**
Report writer: *Beck*	Report writer: *Tenorth*	Report writer: *Gogolin*	Report writer: *Gogolin*	Report writer: *Horstkemper*
Beck	Beck	Beck	Beck	Beck
Erichsen	Erichsen	Erichsen	Erichsen	Erichsen
Gogolin	Gogolin	Gogolin	Gogolin	
				Horstkemper
			Lübke	Lübke
		Phillips	Phillips	Phillips
Tenorth	Tenorth			
Wijnen	Wijnen	Wijnen	Wijnen	Wijnen
Fornefeld	Fornefeld	*Walthes*		
		Simmel-Joachim	*Schelten*	

Table 7: Allocation for third onsite week 17-21 November

These tables reveal the varying engagement of different evaluators. Those experts chosen for very specific knowledge in a particular area, for example Schelten and Fornefeld, were assigned those institutions in which their particular expertise would be needed, e.g. Special Education, or technical and vocational education. They were therefore usually called upon to compile the reports on the HEIs they visited. The only experts who did not compile reports were the non-subject specialist, Erichsen, and two of the foreign experts, Phillips and De Corte. These members were allocated four or five HEIs each and Erichsen, the non-subject specialist, was allocated five. The majority of experts were able to visit either four or five HEIs, which was what was expected of most participants.

The allocation of HEIs to evaluators was a mixture of personal interest, availability and logistics. Each evaluator had been asked to indicate when they might be available and which HEIs they could/might like to visit, although in this letter the HEIs were listed A-O, thereby trying to avoid bias in the selection of institution to visit. As Expert I told me: 'It's always a combination of interests and pragmatic considerations.'[77] The organisation of onsite allocation was left in the hands of the *evalag* agency. Remembering the engagement

[77] […] Und so etwas ist ja immer eine Mischung von Interessen und pragmatischen Überlegungen. Es muss sich dann auch einigermaßen terminlich vereinbaren lassen und die Kompetenzen müssen vorhanden sein. Quotations from individual interviews are being anonymised from here onwards.

of Terhart and Tenorth in the Ulm Declaration, we discover that Terhart visited the PH Karlsruhe, otherwise neither he nor Tenorth were involved in visiting any of the institutions which they had criticised. So it seems that this aspect, which could have caused problems with regard to the way the HEIs perceived the evaluators, was indeed taken into account when decisions on the allocation of experts were made, therefore helping to maintain the integrity of the evaluation commission and to avoid any tensions that might have arisen had they been present at the University of Constance, for example. That said, the fact that the Ulm Declaration made a very clear statement and had been read by the other evaluators on the team might mean that the other experts were more inclined to be against the so-called Constance Model. Indeed, this was the case and the commission's negative opinion towards this model would only shift after Professor von Graevenitz, the rector of Constance, had held an emotive speech on the model's advantages during the last meeting of HEI heads with the commission in April 2004. This finally swayed the commission to accept this rather unusual form of cooperation. As Professor von Graevenitz told me (30.10.08), he realised that if he hadn't spoken up in defence of this model, the commission would have recommended its termination:

> […] I discovered afterwards that the majority of the commission were in favour of recommending that this model be terminated in Constance. And I then held a pretty passionate and long speech in front of the commission. I don't recall the details of it any more. All I know is that some of the experts started looking rather pensive and the end result wasn't to have the model stopped, but to let it continue under certain provisos and then re-evaluate it after a period of time[78].

4.14 Developing evaluation criteria for judging quality of research

As indicated earlier, Baumert had told the group that more work needed to be done to establish the criteria according to which the commission would judge the quality of research in HEIs and he had offered to invest more of his time in working on this specific task. He also believed that something similar should be done with regard to teaching,

[78] Es war dann so, das habe ich hinterher erfahren, dass im Grunde das Gremium mit seiner Mehrheit entschlossen war zu empfehlen, diesen Versuch in Konstanz abzubrechen. Und ich hab dann eine ziemlich flammende, lange Rede, auch sehr energische, vor diesem Gremium gehalten. Deren Inhalt weiß ich jetzt im Einzelnen nicht mehr. Ich weiß nur, dass dann einige der Gutachter sehr nachdenklich geworden sind, so dass das Ergebnis dann eben nicht die Empfehlung war, diesen Versuch in Konstanz abzubrechen, sondern ihn mit bestimmten Auflagen zu versehen und nach einer bestimmten Zeit wieder zu evaluieren.

although he explained that currently there were no research indicators of a comparable nature for teaching quality.[79] Baumert therefore asked Horstkemper for her support; he believed she had far greater knowledge in this field to compile criteria that could be used to evaluate teaching.

In early October Baumert wrote to the evaluators (letter, 08.10.03) describing his progress in determining which evaluation criteria they should be using. He proposed two drafts (a longer version for use by the experts, and a shorter one to be sent to the HEIs) on which he asked for feedback. He recommended that the experts fill in their judgements independently after each onsite visit, and this would then lead to a joint discussion on which final judgements could be made.[80]

The HEIs had at this point already provided classification lists for their publications, which, as Expert B stated, was 'an interesting discovery'[81] in itself. This was said ironically and by this he was indicating the somewhat idiosyncratic inclusions that some university staff believed should be on the publication lists: 'They included local leaflets amongst peer-reviewed publications. That was impressive.'[82] Another evaluator provided me with a detailed explanation of what was meant by 'Bäckerblume' (The Baker's Flower), an expression which would repeatedly come up in the interviews I conducted and was regarded as a standing joke by the whole commission:

> Everything that had gone to print was included in the self report publication lists. Regardless. And at some point a joke appeared that if a pedagogue published an article on 'How to raise my child properly' in the journal for master bakers, "The Baker's Flower", which is distributed in bakeries, then this colleague would include it as a scientific publication. It was the standing joke in the commission which flagged up the missing criteria for quality. The oddest local regional papers were listed as scientific publications which was quite endearing on the one hand, but quite dreadful on the other if one considered the fact that these were German institutions of higher education trying to sell

[79] Rindermann (2001) had in fact been working in this area.
[80] Ich [schlage] vor, dass jedes Kommissionsmitglied, das an einer Begehung teilnimmt, diesen Bogen unabhängig voneinander ausfüllt. Daran kann ein gemeinsamer diskursiver Prozess der Urteilsfindung anschließen.
[81] […] ein interessanter Befund […]
[82] Die haben dann Bäckerblumen-Publikationen unter referierte Zeitschriften eingeordnet - war schon eindrucksvoll.

> this as research and not even noticing how much they were embarrassing themselves in the process by listing them.[83] (Expert R)

This particular quotation shows the naiveté of some university staff and the lack of rigour that was still apparent in the lists which HEIs were sending out to evaluation commissions. It also shows the dilemma the commission was facing in how to classify publications, as within Educational Science this seemed a particularly undefined and difficult area. For example, one evaluator said:

> What does it mean if a subject specialist writes a schoolbook? Is this an achievement or an escape from work? To begin with we all looked sceptically until the subject specialists in the commission, on our first evening, when the problem cropped up (which was on the first day), started to tell us how difficult it was to write a book for schools. How much creativity you need to write exercises, how you need to be able to break down exercises according to level of difficulty, how you need to envisage a whole programme and maintain it over the duration of a school life. And that this isn't perhaps research in one sense, but that it is an incredible achievement, relating to construction of knowledge in applied contexts. And as a result of this we all became more understanding.[84] (Expert R)

A well-written, theoretically-based schoolbook could have far more impact overall, for example, than a peer-reviewed paper written for a small audience. How could this be judged correctly? This problem is still ongoing in Education and indeed during the interviews a number of the evaluators said that in retrospect they believed that the criteria they implemented would need readjusting now (if they were to repeat the exercise), as they had since realised that some of their criteria were possibly too narrow and not totally relevant to this discipline, even though they themselves came from within it. This difficulty

[83] In den Veröffentlichungslisten in den Selbstberichten wurde im Grunde alles, was irgendwie zum Druck gekommen war, als Veröffentlichung gewertet. Unterschiedslos. Und irgendwann kam dann der Gag auf, wenn die irgendeinen Artikel in der Zeitschrift der Bäckerinnung schreiben, die Bäckerblume, die in Bäckereien auf der Theke verteilt wird und da ein Pädagoge schreibt, "wie erziehe ich mein Kind richtig", dann gibt der Kollege das auch noch als fachwissenschaftliche Veröffentlichung an. […] es ein Running Gag der Kommission, um die fehlenden Gütekriterien für Veröffentlichungsorte sichtbar zu machen. […] Die kuriosesten, lokalen, regionalen Blättchen wurden als wissenschaftliche Publikationsorte interpretiert. Und das war, irgendwo rührend, irgendwie. Auf der anderen Seite ganz entsetzlich, wenn man sich gedacht hat, dass das deutsche wissenschaftliche Hochschulen sind, die das als Forschung verkaufen und es nicht merken, dass sie sich blamieren, in dem sie es nennen.

[84] Was bedeutet es, wenn ein Fachdidaktiker ein Schulbuch schreibt? Ist das eine Leistung, oder ist das Flucht vor Arbeit? Am Anfang haben wir alle skeptisch geguckt, bis uns dann die Fachdidaktiker in der Kommission, am ersten Abend, als das Problem auftauchte und das war gleich am ersten Tag, dann mal erzählten, was das für ein Mühsal ist, ein Schulbuch zu schreiben. Was man an Aufgabenerfindung haben muss, wie man auf Lernniveaus runter brechen können muss, wie der einen Lehrgang konzipieren muss, und da über ein ganzes Schulleben hinhält. Und dass das vielleicht zwar nicht Forschung in dem Sinne ist, aber eine ungeheure Leistung, bezogen auf Konstruktion von Wissen im Verwendungskontexten. Und da sind wir auch alle verständiger geworden dabei.

aside, it was generally agreed, however, that there should be a greater emphasis on attempting to write for international journals. Expert N said that some academics in the HEIs under examination thought it was disconcerting to suddenly see the inclusion of English titles on the list of publications which they should be aiming for, but that they had now come round to the idea[85].

The creation of the publication lists was a particularly controversial point of debate within the commission and not all members were satisfied with the way it was compiled. Expert Q felt that they were working according to strictly conventional university criteria for the judgement of quality. However, in this discipline they felt that within the field of pedagogy, when relating to questions of practice such as scientific school research, improvements in teacher training and the improvement of social pedagogues, these criteria did not do justice to describing the worth or quality of the result of this practice and training.[86]

It was therefore apparent that Baumert had a difficult time establishing the list of publications to be finally included as those regarded as 'quality'. Working together with colleagues from the DFG and the president of the umbrella commission for all subject didactics, he started organising an online questionnaire in which associates were asked to list the journals they believed should be included on a list of acceptable publications in Educational Science, as something of this kind specific to Educational Science did not exist in Germany at that point. The Max Planck Institute for Human Development would provide assistance in the compilation of the data and deliver a list to the commission when completed. Furthermore, as there was no equivalent way of classifying teaching quality Baumert was liaising with Horstkemper on assessment criteria. As he stated, they would

[85] Und für die war das schon unheimlich, dass plötzlich die englisch sprachlichen Publikationen gezählt wurden. Aber letztlich haben dann doch alle mitgespielt.

[86] Ich hatte das Gefühl, dass diese Kommission sehr stark an konventionellen universitären Kriterien gearbeitet hat. Sehr konventionellen Kriterien. […] Und bei einem Fachbereich Pädagogik scheint mir das, zumal wenn er sich auf praxisorientierte Fragestellungen bezieht, wie schulwissenschaftliche Untersuchungen, Verbesserung von Lehrerausbildung und auch, in diesem Fall, Verbesserung von der Sozialpädagogenausbildung, oder Sozialpädagogenausbildung in der Wirksamkeit scheinen mir das Kriterien zu sein, die nicht der Güte, ja, nicht die Qualität des Ausbildungsganges beschreiben.

only be able to relate to quality indicators of the organisation within departments, as it would not be possible to assess the quality of individuals.

If we examine the two drafts of Baumert's criteria for assessing the quality of research, we discover that only very few changes to the wording were made to the document for HEIs entitled 'Assessment criteria for research in Educational Science', as he had already compiled comprehensive documents which took international criteria for assessing research into account. The other draft for internal evaluator use was similar to the HEI document, although it was slightly more differentiated. However, comparison with the final version which appears in the published report (pp.184-187) reveals that here, too, most of the wording remained the same. Rankings were divided into six categories ranging from 1 (unsatisfactory) to 6 (excellent), allowing evaluators to rank each category as they progressed with the evaluation. Although Baumert was still working on the list of publications to be included in the evaluation, he forwarded a list from the Journal Summary List, which had been ranked according to merit/status (2001) so that the evaluators could identify the extent to which HEIs were producing publications which appeared in these internationally known journals.

4.15 Typical timetable for HEI visits; onsite reflections

Whilst work was still ongoing in terms of establishing the assessment criteria, *evalag* was organising the practicalities of the onsite visits. Below is a typical timetable (document, *Evaluationsverfahren Erziehungswissenschaft, Begehung der PH Ludwigsburg*).

9.30	Rail transfer from Mannheim to PH Ludwigsburg
10-10.45	Meeting with Rector, Dean, and Female Student Representative
11-noon	Meeting with students
12.15-13.15	Meeting with junior staff
13.15-14.00	Lunch
14-16.15	Meeting with professors of Educational Science and Educational Psychology
16.15	Transfer to railway station
16.31-17.28	Return rail journey to Mannheim and taxi transfer to hotel
18.45-19.45	Post onsite discussion in *evalag* HQ
20.00	Evening meal in restaurant

The schedule was tightly organised and allowed for over five hours of discussion time per day. The evaluators were given 45 minutes with the institutional heads, one-hour meetings with both students and junior staff members, and in this particular case two-and-a-quarter hours with the professors, although this usually depended on the time it would take the group to return to Mannheim, in which case the meeting with professors might be slightly shorter. The order of meetings starting with HEI heads was maintained throughout, as the evaluators would then be able to feed back some of the information from other groups to the professors, whom they visited last in order to stimulate debate on what might be done differently if problems or specific areas of discussion had arisen during earlier meetings.

Following return transfer to Mannheim, the groups would come together to discuss their findings/experiences before ending the day with a meal. We can assume, however, that the report writers would spend an additional hour or two compiling a brief evaluation report which they would have more time to expand on after the onsite week was concluded. If we include the group meeting in the evening and only look at the timetable, we can say that on average the evaluators would spend a total of six-and-a-quarter hours actively engaged in their evaluation task, although this figure should be expanded to include the pre-visit meetings over breakfast in which particular points of focus would be described, in addition to the more informal discussions that would take place during lunch, transfer to and from the institution, and over the evening meal. It is therefore more realistic to say that the number of hours spent on an average onsite visit would be nearer eight or nine hours, if not indeed longer depending on the time travelling to and from institution.

If we examine Phillips's diary in which he describes his visit to one HEI, we see a similar pattern emerging.

> Up early for a post-breakfast meeting to discuss how to approach the sessions planned at the University, a 'new' foundation of the mid-60s. It is gratifying to realise during the session that we have mostly focussed on the same issues. Mandl and Achtenhagen will take the lead, and I will join the *Erziehungswissenschaft* group rather than that for *Berufswissenschaft*, which will be chaired by Frank Achtenhagen. We had been asked to submit questions for particular institutions, and mine, together with those of Professor Walthes, were circulated.

> We travel in taxis to the university […]. Our first meeting is with the Rector and other senior people. There is much talk of cooperation with the PH Weingarten (which the *evalag* is to investigate separately) and with the *Lehrerseminar* in Rottweil. I ask who is gaining and losing from the arrangement, since I have already heard that people in Weingarten aren't so sure that the cooperation is working. The general impression transmitted to us is that all is working well. But then we meet with the students, who – whatever their course – are uniformly critical of pretty well everything, especially the unhappy arrangements with Weingarten and Rottweil.
>
> The *Mittelbau* is represented by only three people, who look rather unimpressive, but speak generally positively about their experience. Later we confront the *Hochschul-leitung* with the views of the students, and one Prof. […] who had chaired the *Studenten-kommission*, says there must be some kind of mistake and sweeps the criticism aside as if it is of no consequence. Appalling arrogance. When we later discuss our general impressions I argue that the students' views were heartfelt and serious and that we cannot allow such a reaction to pass unnoticed.
>
> I really feel for German students who remain quite anonymous to those who teach them, who are crammed into over-filled lecture rooms, who are given little advice of a personal kind and who get scant information generally. I wonder if we are going to find similar dissatisfaction among the student body elsewhere. We have a good summary discussion, during which we seem to hit on the same points for inclusion in the report; then we pile into a bus – in the rain – for the two-hour journey to […] during which we pass through quite a bit of snow.

This extract reveals another problem that the commission would repeatedly encounter: the student voice described its institution in a rather different manner from that portrayed officially through the HEI. This emphasises how important it is for evaluators to spend some time with students to hear at least something of their views, even if there were indications that in some cases the students were not necessarily representatively chosen and might present a biased view. The problem of disempowerment of various stakeholders invariably exists (e.g. HEIs choose students who are sympathetic to the HEI, rather than letting a broader, possibly more critical sample, speak) and it is important to give those in less powerful positions (e.g. students) opportunities to express their views (Guba & Lincoln (1989, pp.52-53).

Following the onsite visits, each of the report writers finished their work and sent their comments back to *evalag* in advance of the December meeting. Phillips, who had not been charged with writing an onsite report, meanwhile sent Wijnen a letter containing his reflections in which he summarised his thoughts on the overall status of Educational Science in Baden-Württemberg. Although this letter was not intended for wider

distribution (and Phillips told me in interview that had he known his letter would be disseminated to everyone, then he might have formulated it rather differently), Wijnen nevertheless wished the rest of the group to see his comments. Röbbecke therefore forwarded the letter to the others (email, 08.12.03).

The tenor of Phillips's letter was to highlight the positive aspects of what was working well within the state, but pick out the problematic areas which needed support e.g. the support of young academics; the increasing use of cooperations; a reduction of legislation and increased transparency; and more support of the teacher training colleges to reduce overcrowding and staffing problems.

4.16 Post-onsite plenary meeting 10-11 December 2003

Once the extensive round of onsite visits had been concluded, the evaluators returned to Mannheim for two days (10-11 December) in which to have an initial discussion of their findings. On this occasion all of the evaluators were present and the meeting was also attended by Bolius, Herberger, and Knorr from the Ministry.

Baumert reported first on a special fact-finding meeting he had just held in Stuttgart on 5 December, attended also by Knorr, Herberger and Bolius, which had been set up to discuss some of the structural problems Educational Science was suffering from and which the committee members would be addressing in their reports. Knorr reiterated the ministerial viewpoint that there would be no change to the particular Baden-Württemberg structure of universities and teacher training colleges, even if the commission wished to recommend otherwise (Point 1, Protocol of commission meeting, 10-11 December). Baumert had also asked for clarification on the Ministry's view regarding the minimum requirements that should exist in each HEI, their opinion on models for successful recruitment of young academics, perspectives for the development of Special Education, and the infrastructure for research. Furthermore, Baumert requested more information on how the Ministries viewed the implementation of the Bologna Process. Knorr stated that all degrees should be

restructured into BAs/MAs by 2010. However, this stood in marked contrast to Lübke's reply when he stated that the Ministry of Culture had *no* plans to reorganise teacher training into a BA/MA structure.

Attention in the group now turned to a discussion of how they would report on each HEI. They decided to divide the reports into two sections: the first would describe the status quo in the HEI; the second would be an evaluative statement of its status quo and its capabilities. This structure would be maintained throughout to provide a brief evaluative statement after each descriptive section. Furthermore, it was agreed that the topic 'Quality assurance' would address both research and teaching and not be divided into two discrete sections. The group also agreed on a template for the HEI reports, which would include a brief historical synopsis of the subject in the institution, followed by points regarding:

1. Development and profile
2. Aims of training and provision of courses
3. Staff
4. Materials and rooms
5. Research
6. Educational Science as a main study programme, divided into sections on organisation of courses, relationship of teaching to research and young academic development, success rates
7. Teacher training, divided into sections on courses and course organisation; relationship to schools; relevance of research to teaching; success rates
8. Quality assurance
9. Further academic development and training, divided into structure and number of courses available; scientific background and relevance
10. Recommendations by commission

Evalag would assist in the compilation of these reports and organise material provided by the HEI self reports accordingly.

The evaluation criteria for research and teaching which Baumert and Horstkemper had been working on were now distributed, although Baumert's list of publications was still not finished and could not be discussed at this meeting. The group decided that the evaluation

criteria would from now on be called reference framework (*Referenzrahmen)* using a term familiar within the Dutch evaluation system (see also final report, pp.184-190). The group also agreed with one of Terhart's proposals to differentiate the criteria for research according to the different HEI type being evaluated, which would make the assessment fairer. We might like to remember here that Point 3 (P3) of the German Society for Evaluation standards concerning propriety advise the following:

P3 Complete and Fair Investigation

The evaluation shall undertake a complete and fair examination and description of strengths and weaknesses of the evaluand so that strengths can be built upon and problem areas addressed.

The group were certainly concerned that the criteria they applied in their assessment of the teacher training colleges could not be identical to those used in assessing the universities, as this would not have been fair. Whether or not they believed they were successful in this undertaking will be addressed in my following chapters.

If we return to what should be included in the institutional reports, they agreed to add 'Cooperation and Networking' as an additional point within the research section, and in the section entitled 'Curricular Structures' aspects such as modularisation and credit systems in all degree programmes should be named as a further aspect of the 'quality dimension'. The revised criteria would be sent out to the HEIs forthwith. Wijnen said that he would provide information on the Dutch system of networks supporting doctoral students.

The group now reached an important decision (agenda point 4). They were unanimous that the schedule for meetings with HEI heads, originally planned for 15 and 16 January 2004, was untenable, and so they agreed that this would be deferred until 1 and 2 April. This would allow for more time to compile the evaluation reports, including the HEI data which

evalag was incorporating. Following a final review of all HEI evaluation reports in a plenary meeting now planned for the end of February, the HEIs would receive their individual reports in advance of the final meeting in April. They agreed that each HEI could send their chancellors/rectors and up to four additional representatives, who the HEIs would select.

The group also decided that the following structural and organisational questions would form the basis of their guidelines for recommendations:

1. The developmental possibilities of teacher training colleges to become so-called 'professional schools', a specific organisational type located somewhere between the polytechnic (*Fachhochschule)* and the university
2. The limitations of the Baden-Württemberg model of teacher training for the *Gymnasium* (including the question of minimum resources)
3. The minimum requirements and necessary infrastructure for efficient research
4. The principle of regionality
5. Models for the successful support of early career academics
6. Long-term restructuring of staff
7. The insufficient profiling of HEIs and
8. The developmental perspectives of special pedagogy (see *agenda point 5: Konzeption der Empfehlungen*).[87]

Baumert offered to write the recommendations text, for which *evalag* would provide a template. Furthermore, the following evaluators would provide brief texts on specific topics: Prenzel (innovations in teaching); Erichsen (teacher training); Gogolin (teacher training colleges) and Fornefeld (special education). The non-specialist, Erichsen, would contribute on teacher training. In his former capacities as both Rector of the Westfälische Wilhelms-Universität in Münster (1986-1990) and President of the HRK (1990-1997) he had gained great insight into the structures of teacher training and the legal aspects of a

[87] Die Entwicklungsmöglichkeiten der Pädagogischen Hochschulen, einem spezifischen Organisationstypus zwischen Fachhochschule und Universität, zu „professional schools"; die Tragfähigkeit des baden-württembergischen Modells der gymnasialen Lehrerausbildung (einschließlich der Frage der Mindestausstattung; die Mindestausstattung und die nötige Infrastruktur für eine leistungsfähige Forschung; das Prinzip der Regionalität; Modelle einer erfolgreichen Nachwuchsförderung; der langfristige Umbau der Personalstruktur; die unzureichende Profilbildung der Hochschulen; und die perspektivische Entwicklung der Sonderpädagogik.

qualification which was granted not by the university or teacher training college, but the state in which the HEI was located.

4.17 Meeting of sub-committee 16 January 2004

After the Christmas vacation, the sub-committee met with Schreier and Röbbecke in Mannheim in January to discuss the interim developments (protocol, 16.01.04). This meeting preceded another scheduled for later in the day in which they would meet with Ministerpräsident Prof. Dr. Frankenberg. This five-page protocol is interesting for a number of reasons. Firstly, it is written in less descriptive language than is usually the case with protocols. The points are listed more in bullet form and not always in the chronological order of agenda points to be dealt with. Secondly, it is indicative of the evaluation being very much 'work in progress' at this stage, where a number of points still needed thinking through and clarifying before being addressed to the whole group. I would therefore like to examine this protocol in greater detail.

The first agenda point referred to the evaluation criteria and it was agreed that a page should be included on criteria for research which were similar to those used for international research, the notion being to raise the bar of standards. It was also agreed that they should refer to cooperation models and networking in research. Equally, curricular structures, modules and credit systems as ongoing criteria for quality should be included.

Agenda point 3 discussed some of the problems that had arisen during onsite visits, notably at the Universities of Ulm and Constance. In Constance, only one professor was in post, as the other chair had been vacant for a few years. The commission believed that two professors were the minimum for maintaining teacher training at *Gymnasium* level within the university. At Ulm the difficulties lay in the conflicting interests of the different chairholders. The tenor was that the commission should recommend either closure or expansion in both of these institutions.

An additional problem lay with the data provided by the universities. The sub-committee decided to ask the HEIs to check the allocation of third-party funding and ensure that the figures were accurate for whole years. The commission believed that in terms of identifying the quality of what HEIs were doing it was essential to stress the importance of funding applications and their success. Third-party funding would have to be defined accurately to ensure that like could be compared with like.

Horstkemper stressed another problem which lay with the individual subject didactics and the lack of research therein. Her suggestion was for more cooperation between institutions and networking, which might improve the research situation. Furthermore, it was suggested that universities should create specialist areas. As the Ministry would be reading the report closely, the sub-committee decided to be particularly sensitive in their evaluative statements regarding subject didactics and the problems in this area, putting this in the main section covering the whole evaluation: 'Subject didactics go into the main section, should be treated constructively and with care.'[88] The format of the reports was also decided during this meeting: the evaluators would provide their comments at the end of each section on main subject, teaching and research.

As the commission was unable to recommend changes regarding the basic structure of teaching training in both universities and teacher training colleges, their recommendations for structural changes were limited. Nonetheless, the view was that the preparation of teachers overall in teaching practice and subject didactics in the teacher training colleges lacked a more fundamental research base. The universities, on the other hand, often lacked the more practical aspects of teacher training. The creation of 'professional schools', which would combine practice, training and research, was therefore proposed. This recommend-dation should be aimed at individual HEIs where the commission thought it especially necessary. Two professors per subject area were held as the absolute minimum to ensure quality of teaching and research:

[88] Fachdidaktik in den Manteltext, sollte konstruktiv und vorsichtig behandelt werden.

> Schwäbisch Gmünd and Weingarten: can these sites be maintained? But what they offer is important to the region, and so the Ministry will have to ask itself if it wishes to maintain the site for regional reasons, but then the minimum standards should be established in the foreseeable future (above all Schwäbisch Gmünd).[89]

One of Wijnen's recommendations regarding research projects that were drawing only very limited funding (also in agenda point 3) was for the establishment of cooperations and networks between HEIs. Programmes could be used by various different HEIs, rather than individually. The group acknowledged that the DFG was generally in favour of seeing teacher training colleges becoming research-intensive institutions, if the state of Baden-Württemberg were to support this idea.

The group discussed Phillips's suggestion (agenda point 5, referring also to agenda point 2, which was about a previous discussion with the Ministry) that people who had experience in other areas should be allowed to become chair holders, providing that they were given the accompanying support and would be able to supervise doctoral students themselves. It should be mentioned here that the federal law on the employment of professors of Educational Science stipulates that they have three years of teaching experience in schools before being allowed to work as a professor in an HEI, which was therefore prohibitive for those without this qualification.[90] In the final report, the commission recommended that the three-year rule be waived if a candidate had suitable experience in another field (final report, p.45). However, the legislation remains in place at present, so no changes have been made thus far.

Regarding Special Education (agenda point 6), the sub-committee acknowledged that although it was in a generally healthy position and had the potential to make an impact internationally, its present research output was not very high and expansion was recommended. The group agreed that this was not just a question of resources, but of the

[89] Schwäbisch Gmünd und Weingarten: kann man den Standort aufrechterhalten? Aber Angebot ist regional wichtig, da muss sich das Ministerium fragen, ob sie den Standort aus regionalen Gründen erhalten wollen, aber dann muss der Mindeststandard in absehbarer Zeit hergestellt werden (v.a. Schwäbisch Gmünd).
[90] Each *Land* has its own legislation on the employment of professors. §49 HochSchG (3) in Rheinland-Pfalz is one such example of the three-year teaching clause.

culture within HEIs and developmental perspectives that could occur in cooperation with others, which could happen across state borders, not just within the state.

In terms of the publications lists (agenda point 3), the sub-committee decided to send HEIs revised publication lists which they could comment on. If the HEIs did not agree with what the commission had decided to remove from the publication lists, then the HEIs had the opportunity to prepare a new list if necessary. The commission would accept this revised list.

At the end of the meeting it was agreed that Baumert would work on structural recommendations and the Ministry discussions. Horstkemper would revise the recommendations. An abstract of two pages should be written to precede the final report, and overall any descriptions should relate to specific problems, which Baumert would do a final edit on.

4.18 Plenary meeting 27 February 2004

The commission convened again in Mannheim at the end of February for a day-long meeting (10am-4pm) (protocol, *Sitzung der Gutachterkommission*, 17. Februar 2004). All of the experts were present, except for Fölling-Albers, Gogolin, Prenzel, Simmel-Joachim, Tenorth and Walthes. Four of the six female members were therefore unable to attend. There were six agenda points:

1. Reports and information
2. Discussion on draft for general recommendations
3. Discussion of onsite reports
4. Preparation for discussions with HEI representatives on 1-2 April
5. Timeline
6. AOB

Baumert reported on the meeting which the sub-committee had held in Mannheim with Ministerpräsident Frankenberg on 16 January (*Anlage 1*), in which the evaluation's most important findings had been discussed. Eight points had been debated:

1. The situation of teacher training colleges and productive developmental perspectives
2. Possibilities and limitations of the Baden-Württemberg model of *Gymnasium* teacher training: minimum resources, scope for cooperation and optimisation
3. Coordination and concentration of Educational Science research
4. Support for young academics throughout all HEIs
5. Recruitment of qualified staff in school education and in teacher training colleges
6. Differentiation and supply of Special Education
7. Courses in Educational Science: types and minimum requirements
8. Security of regional availability and ensuing problems

Baumert reported that the ministerial meeting had been conducted in a pleasant and constructive atmosphere and that he was sure the Ministry would be receptive to at least some of the suggestions made by the commission where these might have painful consequences for various HEIs. He also felt that funding would be made available for the support of graduate networks, for example. The Minister also informed the sub-committee that he intended to introduce the trimester term structure into teacher training colleges in his state. However, as this had not yet been officially passed by state legislation, the commission would be unable to comment on the proposal. Rector Ossner (Weingarten) told me in our interview (20.02.2007) what the Minister's motivation for this suggestion had been, and what had happened to the idea since 2003/2004:

> […] it was the Minister's idea to let teacher training colleges operate for four more weeks than other HEIs. So, initially he thought of a trimester structure and then moved away from that idea. What was left was the so-called February block and so we have winter and summer semesters, plus this block. And we then said that we would use this February block in a really useful manner. […] To prepare students for examinations. To prepare students for re-sits. But above all we use this for inter-disciplinary work groups who work in centres and incorporate our students into these. [91]

[91] Das war eine Idee des Ministers, dass die Pädagogischen Hochschulen vier Wochen mehr arbeiten als andere Hochschulen. Und deswegen, also am Anfang dachte er an eine Trimesterstruktur und dann ist davon weg gekommen und dann ist übrig geblieben der so genannten Februarblock, also, dass es Wintersemester, Sommersemester gibt, plus diesen Block. Und wir haben gesagt, dass wir den Februarblock […] wirklich sinnvoll nutzen wollen. […] Für Studierenden also, dass Prüfungsvorbereitungen gemacht werden, dass Studierende, die durch Prüfungen durchgefallen sind hier vorbereitet werden, die wiederholen. Aber dann vor

Agenda point 2 regarding the draft of the general recommendations discussed what should be included in the final report. *Evalag* had already prepared a draft on the basis of previous evaluations; however, this needed revision to suit the subject of this evaluation. It was agreed that *evalag* would present a revised draft in time for the meetings on 1 and 2 April. The commission asked *evalag* to include an expansion of the remit of the evaluation to include the contribution of Educational Science to teacher education; to consider the topic of teaching-learning more deeply; and to refer to the minimum resources necessary for each subject. Descriptive and summative text sections should be more clearly differentiated, and so the report would need re-organising.

There had been a controversial discussion within the commission on the minimum number of chairs required for Educational Science departments, but they had finally agreed that two should be the minimum standard[92]. However, they were undecided on how many lecturing staff should be provided to support the professors. Professors Achtenhagen and Schelten would prepare a draft for the chapter on economic and vocational teacher training, while Fornefeld offered to edit the recommendations on Special Education.

The next point is indicative of the conciliatory nature of the evaluation. One of the evaluators wished individual chair holders to be named if the quality of what they were doing was below the expectations of the commission, similar to the 'name and blame' culture introduced by the Blair government in Britain, where failing schools were publicly identified. This expert felt that the commission was too institution-oriented rather than focussing on where most of the 'quality dilemma' lay, namely with particular individuals. However, most of the other commission members voted against this suggestion. Instead, they felt that agreed developmental perspectives should be expected of staff members,

---

allen Dingen auch, dass interdisziplinäre Arbeitsgruppen auf Zentrumsebene hier arbeiten und die Studierenden auch mit einbeziehen.

[92] Die Kommission hält nach einer kontroversen Diskussion an der Auffassung fest, dass zwei Professuren im Fach Erziehungswissenschaft die Minimalausstattung für eine universitäre Lehrerbildung sind.

otherwise the spectrum of research could disintegrate into whimsical research topics.[93] This point is a good example of the overall tenor of the evaluation to be conciliatory rather than aggressive, even when it was acknowledged that some academics in the field were not working as well as they should.

In agenda point 3 Röbbecke reported that the HEIs had been given the onsite reports and had already returned their corrections. However, some of them had indicated that they were finding it hard to see a differentiation between the evaluation of Educational Science on the one hand, and that of Educational Psychology and teacher training on the other. The individual report writers should be aware of this and return any corrections to *evalag* by 13 March, allowing the HEIs enough time to respond one final time before the meeting in Mannheim at the start of April. HEIs should also be made aware that some further changes to the texts might be made after the April meeting. The report writers would anonymise individuals and ensure the cohesion of recommendations in their texts with the general text at the start of the report and a clear line between the evaluative statements and the following chapter on recommendations. There was therefore considerable exchange between the evaluation commission and the HEIs, with ample opportunity for comment and feedback at each of the stages.

In advance of the meeting with HEIs in April (agenda point 4) each HEI would be given 90 minutes with members of the commission, including those who had conducted the onsite visits. The meeting would be chaired by someone other than the report writer. In particular, three points should be taken into consideration:

1. HEIs should be given an opportunity to correct statements if these were deemed wrong and point to facts which the commission may have overseen.
2. HEIs could state their reactions to the commission's recommendations.

[93] Mehrheitlich wird die Auffassung vertreten, man müsse abgestimmte Entwicklungsperspektiven der Hochschulmitglieder erwarten können, weil das Forschungsspektrum sonst zu leicht in eine Beliebigkeit der Forschungsthemen zerfalle.

3. HEIs should illustrate their developmental perspectives and cooperation possibilities against the wider context of national developments.

There would be no discussion of the viewpoints and judgements of the commission. At this point the group decided to situate their evaluative statements at the end of the reports, not at various points within the text. However, this decision was later overturned (see final report). All edits and revisions were to be finalised by 28 May, allowing more time after the April meeting with HEIs, as the original date of 23 April was seen as untenable.

4.19 The Achtenhagen/Schelten statement 23 March 2004

Prior to the meetings with HEIs in early April Röbbecke distributed the most current version of the reports on the 15 HEIs (letter, 23.03.03, AZ: MR 117-04). She also distributed a document compiled by Achtenhagen and Schelten, who were experts in vocational training, which presented their arguments against a model of cooperation between teacher training colleges and technical colleges (*Fachhochschulen*) regarding the 'higher teaching certificate in vocational training', which was currently university-based. Their six main arguments can be summarised as follows (Appendix I):

1. Due to new careers starting up and ever-shortening innovation cycles it was necessary to have lecturers who had sufficient theoretical and didactic competence to cope with this, which could only be provided by those who had university training.

2. In contrast with technological colleges, which presented knowledge already developed and embedded within clear guidelines, a university placed more emphasis on problem-solving, which would lead to individuals being able to create new courses of action, necessary for the creation of new courses for new careers.

3. The present dual system of vocational training for non-academic careers in Germany was internationally well-regarded, which was predominantly due to the scientific teacher training provided by universities.

4. The 'higher teaching certificate in vocational training' qualified teachers for employment in a range of schools providing vocational training, which was embedded in theory and practice of a university standard, which should be maintained.

5. Theoretical knowledge was required to understand both the vocational direction and the secondary subject in a sufficiently scientific manner.

6. The ‘higher teaching certificate in vocational training’ was an attractive qualification as it was embedded within the university. If this qualification were transferred to teacher training colleges, or technical colleges, it would lose its attraction and status.

The status of vocational teachers who had received a university education was high and this was something which had been fought hard to achieve, a reason why Schelten and Achtenhagen were particularly worried about it possibly being removed from the university again and ‘demoted’ to the teacher training colleges. Here we see a typical case of specific stakeholders within a commission arguing their cause (Lee, 2006, p.167).

Although their document was over two pages long we see almost nothing of it in the final report. We can identify the comment that because fewer students wished to become vocational education teachers, Baden-Württemberg would benefit from a cooperation model. The commission’s response was thus: ‘This direction is criticised by those in vocational education as it could lead to a degrading of the present courses and hence vocational training per se’ (final report, p.26).[94] I will expand on Achtenhagen and Schelten’s reaction to this in Chapter 6.

4.20 Final meeting with HEI representatives, 1-2 April 2004

Due to the momentous task of speaking to 15 HEIs over a two-day period and giving each enough time to respond to the evaluation’s proposed recommendations it was not possible that all of the commission members could attend all of the discussions. The group therefore divided into two groups and these convened in approximately the same constellations as used during the onsite visits. Report writers appear after the HEI in brackets. The only two members of the commission who were unable to take part in these meetings were Erichsen and Terhart. The division of meetings was organised thus into two groups A and B in the morning and C and D in the afternoon:

[94] Dieser Weg wird von berufspädagogischer Seite kritisch gesehen, da er zu einer Abwertung der bisherigen Ausbildungsgänge und damit der beruflichen Bildung führen kann.

University Mannheim (Mandl)	**University Heidelberg** (Tenorth)
PH Freiburg (Mandl)	**PH Schwäbisch Gmünd** (Tenorth)
Achtenhagen (chair)	Beck (chair)
Baumert	De Corte
Horstkemper	Fölling-Albers
Lübke	Fornefeld
Mandl	Gogolin
Phillips	Prenzel
Simmel-Joachim	Tenorth
Walthes	Wijnen

Table 8: Allocation of experts in HEI meeting, morning session, 1 April

University Freiburg (Horstkemper)	**University Ulm** (Gogolin)
PH Weingarten (Prenzel)	**PH Karlsruhe** (Fölling-Albers)
University Hohenheim (Prenzel/Achtenhagen)	
Achtenhagen	Baumert (chair)
Beck	De Corte
Horstkemper (chair)	Fölling-Albers
Lübke	Fornefeld
Mandl	Gogolin
Phillips	Simmel-Joachim
Prenzel	Tenorth
Walthes	Wijnen

Table 9: Allocation of experts in HEI meeting, afternoon, 1 April 2004

PH Heidelberg (Gogolin/Walthes)	**University Stuttgart** (Baumert/Schelten)
University Karlsruhe (Gogolin/Schelten)	**University Tübingen**
Beck	Achtenhagen
Fornefeld (chair)	Baumert
Gogolin	De Corte
Prenzel	Fölling-Albers
Phillips	Horstkemper
Schelten	Mandl (chair)
Tenorth	Lübke
Walthes	Simmel-Joachim
Wijnen	Schelten

Table 10: Allocation of experts in HEI meeting, morning, 2 April 2004

PH Ludwigsburg (Beck/Fornefeld)	**University Constance** (Mandl/Achtenhagen)
Beck	Achtenhagen
Fornefeld	Baumert
Gogolin (chair)	De Corte
Prenzel	Fölling-Albers
Simmel-Joachim	Horstkemper
Tenorth	Lübke
Walthes	Mandl
Wijnen	Phillips (chair)
	Schelten

Table 11: Allocation of experts in HEI meeting, noon, 2 April 2004

Following a one-hour lunch break on the Friday, the group reconvened for the rest of the afternoon to discuss the events of the preceding two days.

The lowest number of representatives from a single institution was three, but up to as many as eight people from one institution attended this meeting (e.g. PH Ludwigsburg), although this was a special case as it was represented by the university chancellor and two separate faculties, Educational Science, and Special Education, which was located at its sister HEI in Reutlingen. In the majority of cases, the HEI's chancellor was in attendance, along with Deans and Heads of Faculty. It is beyond the scope of my research to examine the reaction of each HEI to the evaluation in detail, nor is it possible to examine the impact that this evaluation achieved in the long term. However, I would like to present PH Weingarten's reaction as an example, stressing of course the uniqueness of each HEI, which means that no two HEIs will have shown an identical reaction or had similar recommendations made.

4.20.1 Discussions with representatives from PH Weingarten

Weingarten is a typical example of a small regional teacher training college serving the local community in Baden-Württemberg. When HEIs have a change of leadership (new president or rector) there can be either periods of stagnation in which the new rector waits until he/she is familiar with the existing structures before making any changes. Alternatively, a new leader can immediately engage in processes of reform if this is seen as necessary. Rector Ossner had taken up his post shortly before the evaluation took place,

which he saw as a means of helping him to introduce change into his institution where he had already identified deficits. In addition to the evaluation of Education, Ossner had asked *evalag* to conduct an evaluation of PH Weingarten in its entirety, which at that time was a unique request as no other HEI in Baden-Württemberg had ever called for something of a similar nature, although it was not unknown in other countries (Huber, 2003)[95] Ossner saw evaluation as supporting the process of change (hence the notion of evaluation as transformation; see also Rossi et al., 2004; Stockmann, 2007, p.62) and he would go on to select 36 recommendations from the final report which, when I interviewed him in 2007, he said were being consecutively implemented:

> We saw it as our job to unite the basic sciences of education, pedagogy and educational science with subject didactics, subject specific areas, and the result is our centres. We now have everything in a matrix structure. This means that every member of the teaching staff is a member of a subject and therefore a faculty. It is the faculty's job in particular to ensure the quality of teaching and improve it. But it also belongs to a centre and it is the centre's job to encourage interdisciplinary research projects and third party funding is an important issue today. (interview, 13.02.07)[96]

What follows next is a description of the exchanges between the evaluation team and representatives from Weingarten on 1 April 2004.

Prenzel opened the meeting by stating that Weingarten was a typical teacher training college, doing well in teaching, but less so in research, where it was not developing joint research programmes or gaining notable third-party funding. The support of early career academics was proving difficult, although attempts were being made to improve the situation. He asked whether more might be done to systematise development and support programmes. Ossner responded by saying that for the first time fewer projects were being applied for, as asking for more funding had sadly resulted in a complete lack of support. Nonetheless, three projects would be supported in the coming year and he acknowledged

[95] One of the experts in my case study, Phillips, was also asked to participate in this institutional evaluation.

[96] Also […] wir haben die Aufgabe gesehen, die Grundwissenschaften, die Bildungswissenschaften, Pädagogik, Erziehungswissenschaft, zusammenzubringen mit den Fachdidaktiken, den Fachwissenschaften und das Ergebnis sind diese Zentren. Jetzt haben wir das Ganze gemacht als Matrixstruktur. Das bedeutet, jeder Lehrende ist Mitglied eines Faches und damit einer Fakultät. Und die Aufgabe der Fakultäten ist insbesondere die Qualität der Lehre zu sichern und zu verbessern. Er ist aber auch Mitglied eines Zentrums und die Aufgabe der Zentren ist, interdisziplinäre Forschungsprojekte zu befördern und Drittmittel ist heute ein wichtiges Thema.

the importance of interdisciplinarity. Furthermore, Weingarten would be trying out a new model in relation to the teaching practice year. Regarding staff, Ossner said that one C4 professor[97] had been requested to conduct empirical school research, and a second chair would be requested after the evaluation was concluded. We should remember here that all requests for Educational Science chairs had been officially suspended by the Ministry until the evaluation was concluded. A third chair was being reorganised into two 2/3 posts, thereby strengthening the numbers of teaching staff. He admitted that pedagogy was suffering under staffing shortages and that Weingarten would try to develop a sensible personnel structure.

In terms of academic development and training possibilities, Ossner said that Weingarten needed to identify a niche which could be offered. He thought that media didactics might be one avenue worth pursuing and that this course was already being offered. He would wait with further decisions until the evaluation was finished. One of Ossner's colleagues stated that a number of personnel decisions would help to improve the situation in Weingarten.

The commission was then asked about the trimester structure (which Ministerpräsident Frankenberg wished to implement) and the experts were asked how Weingarten could profile itself or what it should concentrate its efforts on. Prenzel indicated that Weingarten would be making a mistake if it pursued only one direction of research. It would be better to cooperate in a number of areas. Teaching was undoubtedly the main business in Weingarten, and so research questions could arise from this area. Ossner replied that he envisaged research groups which would pursue one key question each. A number of these related to school issues and this would continue to be a focus, as Weingarten saw itself as serving the local region. Mandl stated that he believed research groups should not be

[97] When this evaluation took place, professors were salaried according to three levels: C2, C3, and C4 (the latter being the highest salary level for most senior professors. Following a reform in 2001, the salary structure was revised to W1 salaries for Junior Professors, W2 for professors mid-career, and W3 for most senior professors. See also the website of the European University Institute for further details: http://www.eui.eu/ProgrammesandFellowships/AcademicCareersObservatory/AcademicCareersbyCountry/Germany.aspx#GrossSalaries

allowed to grow too large and that workshops could serve as useful starting points before firming up research groups. Achtenhagen commented that the main job of a teacher training college was to define 'key competencies' and develop school-related and learning-related questions for research. One of Weingarten's representatives answered that this was a valid point, but that it was difficult to identify the 'organisational-sociological minimum' for qualification. The geographical location and smallness of Weingarten was forcing it into other forms of cooperation, which would clash if concentration and integration were what was being called for. Prenzel asked how many projects could be undertaken with only a certain number of fixed posts and he wondered where it might be possible to tighten up the structures and where timelines for research might be improved. This would also help the planning of publications. At this point Mandl said that of course not everyone wished to engage in research, and so funding could therefore be given to those who had already shown willingness to be so involved. However, Ossner immediately enquired where he could find funding for student support staff. There was none available. He therefore asked the commission to make a point of reporting on the chronic shortage of funding.

Phillips then asked about the cooperation with Constance. Ossner answered that this was not as comprehensive as it might be, being only an exchange of teaching staff between the two HEIs. However, there was a desire to expand this area, for example in the centre for teacher education and cooperation in media didactics/IT. There could be more engagement in the training of *Gymnasium* teachers (Constance) and cooperation might help to reduce the clear boundary between *Gymnasium* training at university on the one hand, and teacher training for the other school types (Weingarten) on the other. This cooperation had been established in 2002, but he asked the commission how they envisaged a fruitful cooperation with regard to teaching. Another colleague stressed that Constance could benefit from the subject didactics taught in Weingarten. Ossner stated that subject didactics were managed by Rottweil. Mandl therefore asked how the cooperation between the teacher training colleges and subject didactics worked. He was told that there were meetings every two years in intercultural didactics and Weingarten was trying to organise a

congress on German didactics, in cooperation with Psychology and Educational Science. Horstkemper suggested that cooperation which supported early career academics could be envisaged between the different HEIs. Ossner told her that they already had a FuN-project in Bioethics and that they were presently trying to start another project with Ludwigsburg[98]. However, his main problem was the lack of young academics, which was preventing his HEI from becoming a member of the DFG and secondly, no-one was able to continue with a doctorate upon completion as they had to work in schools for three years. This was a point which he was keen for the commission to discuss with the Ministry, to help reduce the rigidity of present legislation. One of his colleagues also commented that it was the same situation when chairs were being advertised, as interesting candidates without the obligatory three year teaching experience were not even invited for interview.

When the commission was asked which strategy would be best for Weingarten to pursue, Mandl said that it was better for the HEI to develop a profile and maintain key areas in which they had already built up expertise. Pedagogic careers should be professionalized and the HEI should decide in which areas it had competence and focus on these. Prenzel reinforced this comment by saying that there were plenty of problems that needed addressing within school education. If research in this area were strengthened, then the HEI would have a strong position in the future.

With regard to the Ministry's proposal to introduce a trimester structure Ossner thought that this would cause more problems and increase their isolation from other HEIs, reducing possibilities for cooperation rather than increasing them. He asked the commission about their views regarding the cooperation model with Constance. We should remember here that the model had been criticised in the Ulm Declaration and Constance was held to be out of line with the rest of the country. It therefore came as no surprise that Mandl stated that this model was an exception and that he believed it would not be sustainable over time.

[98] FuN projects (Forschung und Nachwuchs/Research and Young Academics) are relatively new in German HE. They were established to give the most promising young academics possibilities to become more integrated into ongoing research teams.

However, Beck argued that his own institution cooperated with Constance and that this worked well if both HEIs engaged mutually and with equal strength. He stated that the model with Weingarten was not functioning properly at present because the relationship was not balanced, although the media didactics/IT project was a good start. He suggested that Weingarten concentrate on its main business of training teachers, but that it should improve research possibilities and strengthen continuing training and academic development.

In conclusion, the commission suggested that Weingarten needed to define its core curriculum more clearly, although this was restricted in part by what the Ministry stipulated. Prenzel would be asked to contribute some thoughts on how research could be better incorporated because although Weingarten was willing to increase its research output, staff did not have the knowledge/experience of how to do this.

4.21 Editorial meeting, 29 April 2004

Following the two-day meeting with institutional representatives and having allowed time for the various report writers to revise their scripts and return them to Mannheim, a one-day editorial meeting was set up for the end of April (protocol, *Protokoll der Redaktions-sitzung vom 29. April 2004*). This meeting was not attended by any members of the sub-committee, but was managed entirely by Achtenhagen and Prenzel. Equally surprising was the fact that Röbbecke was also absent, although her place was taken by Schreier (*evalag*).

This document is particularly interesting because it details various changes to the wording of the reports (agenda point 1), some of which refer to gender, for example, turning words such as 'Professoren' (masculine plural form) into the politically more correct 'Professoren/-innen'. Furthermore, the language of imperatives was softened, so that in recommendations and evaluative statements we see the verb *müssen* (must) becoming *sollen* (should). Statements without hard indicators should be softened, and so we now see constructions such as 'It seems to the commission that…' (*Der Kommission erscheint…*).

Furthermore, the group agreed to include more detailed statistics in the final report. Tables were included to reveal the average length of Educational Science degree programmes in teacher training colleges and universities, in addition to the average length of teacher training courses in both institutional types, which would allow for more cross-institution comparison where necessary (agenda point 2b). They also agreed to add some comment in the main introductory text on the success rates of students studying to be *Gymnasium* teachers in comparison with those on other teacher training courses, which would again help to highlight the quality of teaching in each institutional type if, for example, it was obvious that some HEIs had particularly high drop-out rates.

The editorial group also decided to include more data on the amounts of funding which each institution had acquired and they asked university offices to provide specific information on exactly how much money had been awarded to each research project type (agenda point 2c). As there might be disparity between the statistics provided by the departments and the university per se, the group decided to only use university rather than departmental data, which was seen as being less reliable.

Achtenhagen offered to contribute a brief chapter together with Schelten on the co-operation that existed between the PH Ludwigsburg and the technical college (*Fachhochschule*) in Ludwigsburg with regards to trade and vocational training, which would be inserted into the main body of the introductory text, as a result of which it would be necessary to edit out evaluative statements from other onsite reports which might double up some of the new material (agenda point 3a).

Agenda point 3b recommended that the report's introduction should indicate the commission's view that they believed there should be an appropriate balance between the number of professorships and other academic staff members; in other words there should be no imbalance in favour of one group. They would speak positively of those cases where a balance had been achieved.

In terms of the future planning and organisation of the final report, they decided to electronically incorporate handwritten edits of some points which had been discussed during the meeting and send these in 'edit' mode to both Achtenhagen and Prenzel by 10 May. The same procedure would be used for any outstanding reports which the individual report writers might still be sending to *evalag*. Achtenhagen would again edit his reports on the Universities of Hohenheim, Mannheim and Constance and send them to *evalag* (agenda point 6)**.** I have decided to omit any comment on agenda point 5 in order to preserve the identities of those involved in discussions of a particular difficulty which arose with one institution.

Following inclusion at *evalag* of additional statistical information, and the drafts of onsite reports from all of the report writers, third drafts of both the evaluation criteria (*Referenzrahmen*) and the introductory texts were sent to the group in advance of the 28 May meeting, in which all of the documentation required for the final report would be examined in a day-long meeting. Individual comments were discussed, agreement on wording reached, and it was now *evalag's* task to assemble the final document, which was accordingly dispatched to the Ministry for publication in advance of its August public release.

It seems appropriate here to examine the criteria which were finally laid down by the commission and which form the backbone of what the commission was looking for and why (final report, pp.184-189).

4.22 Agreed general criteria regarding resources for research

The report states that the evaluation examined the research environment, resources, technical infrastructure, library, and financial resources of the 15 HEIs.

1. The HEIs would be assessed according to the extent of academic staff allocation, their ability to develop, encourage and maintain research projects

2. The ratio of staff to students, in particular the number of lecturers and early career academics, including students working for professors and other staff
3. Supply of technical support staff
4. Administrative support

4.22.1 Scientific quality of research

This looked at the extent to which research quality was in accordance with international quality criteria, notably in terms of originality of work, significance of advancement in the field, high methodological and technical standards, and international recognition. This last point included the extent to which academics were invited to international and national congresses, their memberships on editorial boards or as reviewers, and invitations to work as experts for the WR or internationally peer-reviewed journals.

4.22.2 Productivity and practical relevance of research

This included the publications in which the field should be disseminating research findings, the number of publications and contributions in internationally and nationally recognised congresses and conferences. The commission wished, however, to present this particular category in greater depth than merely listing which journals were being published in as they wished to gain an understanding of the distribution of publication activity among staff members. The team would assess this kind of information according to more subjective and qualitative methods which they regarded as necessary given the sub-disciplines and the nature of the different types of institution involved.[99]

In terms of assessing the worth of the practical relevance of research, the commission was also aware that their judgement would be subjective and open to debate. In other words, there had been discussion on the significance of research (impact factor), but until there were internationally accepted criteria for assessing this, the evaluators would have to use

[99] Die Kommission ist sich bewusst, dass mit der Relativierung subjektive Faktoren ins Spiel kommen. Sie hält dieses Vorgehen jedoch angesichts der institutionellen Unterschiede und der Besonderheiten der einzelnen Subdisziplinen für notwendig. (final report, p.185)

their own professional judgement. In those cases where the commission had not necessarily reached a consensus of opinion, this would be specified.

4.22.3 Research funding

The commission agreed that funding was crucial to maintaining and developing the research profile of each HEI and that it was also essential in supporting early career academics. The assessment criteria therefore included the number of projects funded by the prestigious DFG and those which received part-funding through this body. Furthermore, they would examine the number of staff who were engaged in applying for funding or had received funding.

4.23 Early career academic support

As this was something the commission laid particular emphasis on, it seems appropriate to present this section in its entirety.

> The systematic support of young academics is the most important instrument for long-term structural disciplinary development. During the qualification period, the foundations for enculturation into the international scientific community are being laid. A necessary, albeit not sufficient criterion for success is solid methodological skill. This is applicable to areas predominantly in the social sciences as in empirically operating fields. The deficits of young academics working in empirical educational research are particularly great in comparison with peers in the humanities and other social sciences.[100]

The criteria set out by the commission pointed clearly to the fact that more could and should be done to improve the situation for young academics to give them opportunities to participate in conferences, develop their methodological skills and create possibilities to co-author papers with more senior academics.

[100] Die systematische Nachwuchsförderung ist das wichtigste Instrument für die langfristige Strukturentwicklung der Disziplin. In der Qualifikationsperiode wird die Grundlage für die Enkulturation in eine internationale Wissenschaftsgemeinschaft gelegt. Eine notwendige, wenn auch nicht hinreichende Voraussetzung für Erfolg ist ein solides methodisches Können. Dies gilt sowohl für primär geisteswissenschaftlich als auch empirisch arbeitende Bereiche. Im Bereich der empirisch arbeitenden Erziehungswissenschaft sind die Defizite des Nachwuches im Vergleich zu anderen Human- und Sozialwissenschaften besonders groß. (final report, p.186).

4.24 Cooperation and international networking

Because there seemed to be deficits in the extent to which departments were able to offer Educational Science in its entirety, the commission placed emphasis on collaborations and international networks which would go some way to compensate for this deficit. Their expectations in this were high and they specified that one way of working towards this aim was to promote the idea of joint publications between members of different institutions.

4.25 Criteria for judging the structures of teaching and learning provision

As had already been specified, although this evaluation was charged with assessing both research and teaching, the group had decided that it was unable to examine the quality of individual teaching. They had therefore agreed that the evaluation could only examine the structures and provisions that were in place which included examining the levels of staff qualification, breadth of teaching courses, numbers of staff, levels of support for students, numbers of tutors and student support staff, and numbers of administrative support staff. This would therefore take into account the expectations of what a university offering Educational Science should be able to provide overall. The curriculum as between foundation and main degree courses should be clearly differentiated and organised to include modules and the assurance of particular courses at certain stages of degree programme. Methods which encourage independent and research-linked learning should be present and be sympathetic to current international research trends. Equally, there should be clear links between the aims of training programmes and the actual training courses being provided. Students should be prepared for internships, given support during these and also helped in the discussions of their teaching practice once the internships were concluded. In addition to this, there should be provision of skills training which would give students career-specific competences.

Regarding the didactics of courses, the commission specified that it was looking for innovative approaches and suitability in terms of subject matter, in particular where

lecturers were dealing with large student numbers. They encouraged the use of electronic support via email and the internet overall and wished to see employment of modern media techniques increase. Teaching programmes should be planned progressively, systematically, and complemented by outside support if possible. Student support provision should therefore be comprehensive and be internationally oriented, which would enable foreign students to participate in exchanges, and German students should equally be able to take part in foreign exchanges, too.

Internal quality assurance was regarded as a foregone conclusion and the commission expected an ongoing preoccupation with quality enhancement within departments. The methods of quality assurance would be documented by the commission, which would form the basis for their judgement. They would take the following into account: the extent to which individual student trajectories were being observed (and supported), the destinations of graduates, the methods of self-evaluation being used in order to optimise teaching quality, support of academic development within the HEI, and encouragement and awards for excellent teaching.

4.26 Summary

This chapter of my study has contextualised the different phases of the evaluation and revealed who was responsible for particular input at various times. The reactions and comments of some individuals have been recorded and we have learnt about the roles which certain members within the commission played. Table 11 overleaf summarises the key stages of the evaluation.

Sub-committee meeting with Ministry for first discussions 26.02.03

Evalag prepares 1st draft of questionnaire 10.04.03

1st plenary meeting 16.04.03

Introductions; evaluation rationale; time allocation

Questionnaire draft of 10.04.03 sent to commission. Feedback returned to sub-committee

Sub-committee works on questionnaire 16.04.03-05.05.03

Questionnaires A + B (self reports) sent to HEIs

HEIs return questionnaires in August, which are dispatched to commission

Sub-committee meeting to prepare 2nd plenary meeting 11.09.03

2nd plenary meeting 25.09.03

Discussion of onsite meetings; group allocation; questions for HEIs; evaluation criteria

Letter received from professor, sent to commission 06.10.03

JB sends 2nd draft on quality criteria to commission (long +short) 08.10.03

Onsite visits 20-24.10.03

Onsite visits 03-07.11.03

Onsite visits 16-31.11.03

JB meets Knorr, Herberger, Bolius to discuss structural matters 05.12.03

Evalag *distributes 1st drafts of onsite reports and DP's comments* 08.12.03

3rd plenary meeting 10-11.12.03

Discussion on how to compile report. JB + HK to work more on evaluation criteria/ Referenzrahmen*. JB to write recommendation texts using* evalag *template; MP to write on innovations in teaching; H-U E to write on innovations teacher training; IG to write on teacher training colleges; BF to write on Special Education*

Sub-committee meeting to prepare for 4th plenary meeting 16.01.04

Discussion of 3rd draft of evaluation criteria; funding data; cooperations

Sub-committee meeting with Minister Frankenberg 16.01.04

4th plenary meeting 27.02.04

Discussion of 2nd main draft for final report, to be revised after meeting with HEIs; onsite reports; meeting with HEIs. FA + AS to write on vocational + business education; BF to write on Special Education

2nd draft of onsite reports sent to evalag	13.03.04
*FA/AS statement sent to commission and 3rd draft of onsite repor*t	23.03.04
Plenary meeting with HEIs	01-2.04.04
Editorial meeting (FA, MP, GS)	29.04.04
Inclusion of more statistical data; 4th draft of onsite reports from experts to be received by 10.05.04.	
3rd draft of Referenzrahmen *sent to commission*	14.05.04
3rd draft of introductory texts sent to commission	19.05.04
5th plenary meeting	28.05.04
Day-long editorial discussion to finalise order/contents of final report	
4th and final draft of Referenzrahmen *and 4th draft of introductory text prepared by* evalag *for publication by Ministry*	
4h draft of final report	15.07.04
Publication of final report	August 2004
(Abbreviations used are initials of evaluators. See Appendix C for details)	

Table 12: Overview of evaluation

The entire commission came together five times for plenary meetings, with an additional two-day meeting with HEIs. The sub-committee convened five times to prepare for the plenary meetings, discuss evaluation criteria or meet with Ministerial members if this was deemed necessary. Jürgen Baumert also had an additional meeting with Ministerial members to discuss structural matters. In terms of the number of drafts that were necessary before the commission felt satisfied, we can see that four drafts were required for the *Referenzrahmen*, which progressed from a standard evaluation text provided by *evalag* to one that was specific to Educational Science. The introductory text in which the criteria and modus operandi of the commission were explained also went through four drafts. We can reduce the stages of the commission into eight key ones. I do not include here the stage in which the experts were recruited, because this was not directly in association with their

own evaluative work, although we know that some members did provide input as to who else might be included:

1. decision on what should included and excluded from evaluation
2. compilation of questionnaires for HEIs (self report)
3. review of self reports in advance of onsite visits and compilation of onsite questions for HEIs
4. onsite visits
5. compilation of evaluation criteria (*Referenzrahmen*)
6. compilation of onsite reports
7. feedback from HEIs
8. revision and compilation of final report; publication

Underlying each of these decision-making stages were discussions which involved different perceptions dependent on the values of those participating. We have already seen that values and opinions were not always the same, although efforts were made to reach consensus on most questions, even if it was acknowledged that these were not always entirely satisfactory.

From the above we see that the sub-committee played a major role in preparing and organising the evaluation. Individual members of the commission played significant roles over and above what was expected. This was partially due to their personal motivation as they volunteered more input. However, their contribution no doubt stimulated more debate and ultimately increased the quality of the evaluation itself. We might therefore say that participants who are willing to engage more than what is expected of them may play a more decisive role in steering an evaluation in a certain direction, although this does depend on how far the other members in the group accept their views, values, and criteria. Although some might argue with the 'conventional university criteria' that were applied during this evaluation, no one could argue with the extent of engagement of some of the experts, notably Baumert, who invested heavily in this evaluation and therefore left his stamp on the proceedings overall.

The documents have provided us with a detailed understanding of the actual events which the evaluators experienced as this evaluation progressed. Analysis of the archival material has highlighted much of the difficulty in establishing what was seen to be the correct evaluation criteria to be employed, and although the group believed that they had reached a sensible basis, not all were entirely satisfied with its outcome. Equally, there was some dissatisfaction at the extent of ministerial involvement, although this was tolerated and, on occasion, found to be quite useful in those cases where the experts needed clarification. We have seen that the group was fairly conventional, basing their ideas on quality on criteria which had become the norm in other disciplines, e.g. publication lists, comparing amounts of research funding, etc.

In the following chapters, which are based on interview data, I shall explore more aspects of the *Eigenwelt* of each evaluator and will reveal how these related to the *Mitwelt* of the other participants in the evaluation, the *Lebenswelt* of Educational Science and their role within it against the backdrop of German academia, the *Überwelt*.

SECTION 3: INTERVIEW ANALYSIS

Chapter 5

Context of evaluation: roles of individuals and their work ethics

The final report reveals the consensus which was reached by the evaluation commission after the results of their onsite visits. However, documentary analysis can only go so far to reveal the decision-making which took place and the influences which underpinned this. It shows us the conclusions, but not the individual voices of those concerned as work progressed (excepting, of course, those extracts from the Phillips diary). The interviews I conducted allowed me to explore the evaluation on a different level therefore, and although we need to remember that experts can develop a different stance on a completed evaluation, especially if they take part in others subsequently, which might lead to memory distortion or some confusion, because I interviewed everyone concerned and am able to compare statements I believe that I am able to reconstruct the processes of this evaluation (Patton, 1987). The following and the next chapter, too, will allow individual voices to have their say; they will also reveal something of the shifting nature of definitions of quality in this particular field and this will reveal that experts need to maintain flexible views with regards to quality standards and their perceptions of their own field per se (Stake, 1975; Guba & Lincoln, 1989).

This section of my thesis therefore investigates the data gathered from the interviews I conducted with all of the experts, members of the two ministries involved, and the employees of the *evalag* agency who were directly involved with the evaluation. It also includes data from interviews I conducted with institutional heads.

5.1 Special areas of knowledge, age of experts, and other experience

I would like to turn first to the qualifications and histories of the evaluators in this case study so that we may understand the different contexts from which they came to this evaluation. We remember from Bolius's statement in Chapter 4 that the experts were chosen to include the breadth of areas within the discipline and that many were also experienced in university administration. Ideally and in addition, evaluators 'should be familiar with the full repertoire of social research methods' (Rossi et al., 2004, p.27). Let us see then the areas they covered, as the experts were all qualified as professors in the following sub-divisions of Educational Science (Table 13):

Area of expertise/specialisation	
Comparative and International Education	Economics of Education
Empirical Educational Research	Empirical School Research
General Education	History of Education
Primary School Education	Educational Psychology
Special Needs	Teacher Training
Vocational Education	

Table 13. Area of expertise/specialisation

This list indicates part, but not entirely all, of the breadth of the subject under examination. Each of these areas has its own areas of specialist knowledge which, as already stated, may be rather unfamiliar to other members, although all of the participants are 'in Education'. Indeed, we shall see that a large part of the work of the commission would involve a learning process about the different sub-areas. Only then would it become possible to start making judgements about the quality of each.

It has not been possible for me to glean information on the age of every participant. Unless this information was publicly provided through means of online curriculum vitae (CV), I did not think it polite or appropriate to ask them in person. However, using the CVs of those available to me, and seeing that the chairperson and some of the other members were obviously above the age of sixty, it is possible to state with some certainty that the average age of the participants was late fifties and in some cases they were nearing retirement age

of 65. Most had occupied high administrative positions, be that as university presidents, deans, members of the WR, or chairs of the German Society for Educational Science. They had profiled themselves over and above their personal academic merit, which gave them knowledge on university management and academic structures. Table 16 provides a breakdown of additional positions of responsibility, as far as I was able to determine. The experts have been rendered anonymous in this table and are listed on the left as numbers 1-18.

Expert	Dean	Rector	Director of an Institute	Board of DGfErz.	DFG	Senate member	German Science Council Assoc.	Member of Int. Edit. Boards	Int. chairs & Member-ships	HRK	KMK
1			x		x						
2					x	x		x			
3		x								x	
4											
5	x			x							
6	x										x
7									x		
8											
9											
10											
11							x		x		
12				x							
13				x					x		
14		x								x	
15			x		x	x			x		
16	x							x			
17											
18				x		x		x	x		

Table 14: Positions of responsibility held by evaluators

(Abbreviations used: DFG (German Research Foundation); DGf.Erz (German Society of Educational Science); HRK (Standing Conference of German Rectors); KMK (Standing Conference of the Ministers of Education and Cultural Affairs of the *Länder* of the Federal Republic of Germany).

Most of the experts held at least one of the positions specified above, and some occupied far more. Most had an average of two. Even if individual experts did not have experience in particular areas, these were inevitably covered by other members, and so overall the breadth of experience was considerable. As can be seen from the types of positions, these were positions which gave an added value to the judgemental ability of the individual expert in question. As Lincoln & Guba (1985, p.256) remind us: 'Respondents are much more likely to be both candid and forthcoming if they respect the inquirer and believe in his or her integrity.' Experience in the roles stated above certainly increased the likelihood that the HEIs in this case study would be more inclined to trust the judgements of the experts involved.

5.2 Role of non-specialist expert

It has become standard practice in evaluations to include someone who is from outside the field to be examined. His or her participation is expected to shed a different light on a discipline and, as mentioned by Bolius, 'break a bit through subject blindness'. However, she admitted that much depended on the personality of the person concerned as to whether they were able to fulfil this criterion or not.

The non-specialist expert in this case was a professor of Law. Although coming from a different discipline, he had been President of the HRK and was therefore familiar with many of the issues involved in German Educational Science and the different types of institution in it, even if he was not an educationalist himself. He told me: 'As President I was often confronted with questions concerning the training of teachers and I always had the feeling that it wasn't particularly satisfactory.'[101] This statement reveals that he did not come to the evaluation with an entirely open mind. Indeed, he stated categorically that because the entire elite of Education, as he saw it, was represented within the commission, he thought it might be interesting to engage in debate with them, e.g. ruffling their feathers somewhat to see what changes might be made to help improve the present unsatisfactory

[101] Ich hatte schon als Präsident der HRK häufig mit Fragen der Lehrerbildung zu tun und hatte immer das Gefühl, dass es nicht gerade besonders zufrieden stellend ist, was da abläuft.

situation. This was not directly adversarial in tone, but clearly an indication that he would not let things which he felt needed addressing more directly pass by without comment. When I asked him to describe his experiences of his role within the commission he said:

> Everyone in the commission was a full-blooded Educationalist […] and sometimes people didn't make particular progress with each other. That's when I would ask questions, which sometimes helped to get things going again. I am a great fan of the inclusion of non-specialist experts, not exclusively, but in addition.[102]

This was reinforced by comments made by others, including Lübke, who described the engagement of this expert as follows:

> He's a lawyer. That's right. He contributed a good deal and got very involved in the discussions. Although he always said […] that he couldn't judge things in the way the educationalists do. […] He was completely integrated into the group and accepted by it.[103]

I have already indicated that the group's confidence in him was such to call upon him to contribute a section in the final report on teacher training.

5.3 Prior knowledge of *evalag* and recruitment to evaluation

Of the 18 evaluators involved, 12 had no previous knowledge or experience of working for the *evalag* agency. The others knew of its existence, but had not been directly involved. The only person who had worked for *evalag* before was Simmel-Joachim, who had been part of the evaluation commission to examine faculties of Social Work in Technical Colleges within Baden-Württemberg in 2002. Baumert, Erichsen and Phillips knew *evalag's* Managing Director, Dr Schreier, as they had all at various stages in the past worked with him while he was employed by the WR. Others knew of the agency as they had already been asked to participate in an evaluation of Education (Tenorth & Terhart). In her capacity as the then President of the German Society for Educational Research Gogolin had thought it important to familiarise herself with the different evaluation agencies that

[102] In der Kommission waren alles eingefuchste Erziehungswissenschaftler, […] und man kam gelegentlich miteinander nicht recht weiter. In diesen Situationen habe ich Fragen gestellt. Das hat gelegentlich dazu geführt, dass es wieder weiterging. Ich bin ein großer Anhänger der Einbeziehung fachfremder Evaluatoren, nicht nur, aber auch.

[103] Der ist Jurist, richtig. Der hat eigentlich relativ viele Beiträge geliefert. Der hat sich sehr intensiv auch beteiligt an der Diskussion, wobei […] er hat […] das dann auch immer wieder deutlich gemacht, dass er das jetzt nicht von der Seite des Erziehungswissenschaftlers beurteilen kann. […] Er [war] als Mitglied der Gruppe voll integriert und da auch akzeptiert.

existed: 'Well, I knew of *evalag* out of this context just like other similar agencies – Aquin […], but that was the first and only time I did a project with *evalag* […]'[104].

One of the first questions I asked all of the experts in interview was how they were recruited to this evaluation as I was interested to learn about the extent to which networks existed, or who may have been suggested by whom. Although I already had a fair idea on this on the basis of my documentary research, I wanted to discover if this was of any importance to them. An interesting phenomenon was that many evaluators could not clearly remember the circumstances of their recruitment, although we have documents which specify what happened. In their recall, the majority told me that they were contacted by *evalag*, which in itself does not reveal the extent of the networking which might be taking place behind the scenes. Gogolin, however, reported that she had been contacted inofficially first by Baumert. After she had indicated her willingness to participate, she was then officially contacted by *evalag*. Horstkemper put forward Beck's name as being an expert in teacher training[105]. I have already mentioned that during the first plenary meeting the group realised that there were not enough representatives from Special Education, which led Fornefeld to contact Walthes, who came to the commission after the first plenary meeting.

5.4 Familiarity with other colleagues / previous networks

Although Germany is a large country, the discipline of Educational Science remains one of the smaller academic units, even if does cover a rather large number of sub-areas. To some extent we can expect its members to be mutually aware of each other and in some cases they may have worked together, either in a research or in an advisory capacity in some other commission. This poses an interesting challenge for evaluations. On the one hand an

[104] Also, aus dem Kontext kannte ich die *evalag* wie eben auch andere entsprechende Agenturen – Aquin, aber mit der evalag ein Projekt durchgeführt habe ich da zum ersten und zum einzigen Male […].

[105] Der Professor hat einen Experten aus der Schweiz gesucht, der möglichst mit Veränderungen in der Lehrerinnen- und Lehrerbildung vertraut ist, und kannte mich von meinen Forschungsarbeiten her und hat mich daher vorgeschlagen. Und es war auch in diesem Gremium neben Professor Baumert, war Frau Horstkemper, und Frau Horstkemper kannte mich ebenfalls aus einem anderen Zusammenhang. So bin ich in diese Gruppe gekommen.

external evaluation with members drawn from outside the HEIs under inspection is claimed to be of higher legitimacy in that it provides empirical evidence which institutions can then use in order to raise their quality (Rossi et al., 2004). However, the selection of those who represent certain paradigms may serve to influence the evaluation itself. I therefore asked the experts to identify which members on the committee they had worked with previously, or at least knew of through the literature. I wondered if they might feel it a hindrance or a help to have known colleagues involved. Would this influence their decision-making? Would it have an effect on their ability to argue if they were together with acquaintances, in some cases with 'du'-friends?[106]

Most experts appreciated being able to work with colleagues they knew. These were regarded as known quantities, even if they did not always share quite the same values. Expert N described what he thought about who should make up the commission as follows:

> A determining factor is the personal constellation. That you get numerous references beforehand, see whether the individual is able to reach sound judgements, establish whether they are agreeable, or if they pursue their own agendas. […] I think those are the most important […] and that they get on with each other.[107]

Underlying this comment is that experts should be amicable, open to suggestions and not totally set in their opinions. Expert R said that it was difficult to work with argumentative people who might strike out in opposite directions. Critical argumentation was of course important, but an ability to reach compromise during such projects was necessary:[108]

> Dissent remains relatively great and the variance between people and the ability to discuss openly does not suffer, I think, from knowing each other well. But what is

[106] As in France, which distinguishes between the familiar *tu* or more formal *vous* form of address, Germany has a similar differentiation in address between *du* and *Sie*. In German academia, the *Sie* form is usual; the *du* form is only reserved for those with whom closer contact has been established (usually over years). In the commission there were members who addressed each other in the familiar *du* form, whereas others would be addressed as *Sie*.

[107] Entscheidend ist die personelle Zusammensetzung. Dass man bei der Zusammensetzung der Kommission vorher mehrere Referenzen einzieht, ob die Personen also ein klares Urteil haben, ob sie verträglich sind, ob sie nicht ihre eigene Agenda verfolgen [...] das sind, glaube ich, die wichtigsten... und dass sie zusammenpassen.

[108] Und es ist ja schwierig mit Leuten, die eher immer quer argumentieren, kritische Argumentation ist selbstverständlich wichtig, aber die dann auch bereit sind, Kompromisse einzugehen in solchen Verfahren.

> probably the case, I will admit, is that there is a systematic price, of course, in the choice of members.[109]

In terms of who knew whom, I have constructed the following matrix (see Fig. 3), which displays the number of known connections. This matrix cannot be regarded as definitive, as some statements were vague inasmuch as an expert might say that he or she knew everyone personally, which might not be corroborated by statements from others. Nevertheless, the matrix gives us some notion of the extent to which members of the commission knew each other beforehand.

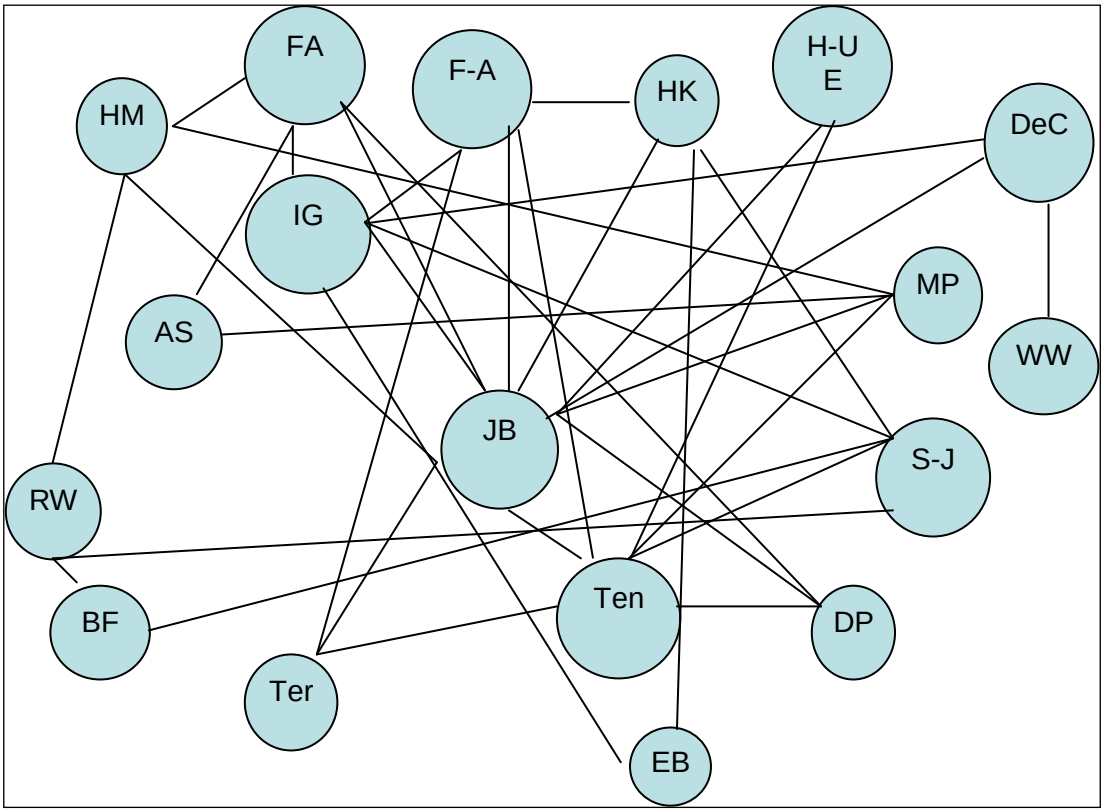


Figure 3. Matrix of evaluator contacts

(Abbreviations: AS = Schelten, BF = Fornefeld, DeC = De Corte, DP = Phillips, EB = Beck, FA = Achtenhagen, F-A = Fölling-Albers, HK = Horstkemper, HM = Mandl, JB = Baumert, IG = Gogolin, MP = Prenzel, RW = Walthes, S-J = Simmel-Joachim, Ten = Tenorth, Ter = Terhart, WW = Wijnen)

In terms of his contacts and networking the matrix reveals that the person at the heart of this evaluation was Jürgen Baumert, as he connected in some form with most members in

[109] Trotzdem ist der Dissens relativ groß und die Varianz zwischen den Leuten und die Offenheit der Diskussion leidet, nach meinem Eindruck, muss ich sagen, nicht darunter, dass man sich relativ gut kennt. Was wahrscheinlich der Fall ist, das würde ich Ihnen einräumen, bei der Auswahl der Mitglieder für diese Kommission gibt es natürlich einen systematischen Preis.

the committee. Significantly, the only members he did not form a connection with, as far as I could tell, were those engaged in more peripheral subjects, notably Walthes, Fornefeld and Schelten. This reinforces the role he occupied within the commission, as there was general agreement that the most influential person in this commission was Baumert, a point to which I will return in greater detail later.

De Corte, who we recall was originally asked to be the chairperson, was well-connected and together with Phillips held most links among the foreign participants. Perhaps the Ministry was aware of this and it may be an indication as to why he was originally earmarked for this position. Wijnen, on the other hand, was not well-known by the German educationalists, even if he had considerable evaluation experience to offer. It could be argued that those with the greatest number of links, e.g. Achtenhagen (5), Fölling-Albers (5), Gogolin (6), Simmel-Joachim (5) and Tenorth (7), were more likely to be figures of greater impact within the evaluation, according to their status of being well-known, whereas those with fewer links, e.g. Beck, Erichsen, Fornefeld, Horstkemper, Mandl, Phillips, Schelten and Walthes, might have less input/influence. On the basis of a study conducted in Sweden by the Swedish Medical Council, which looked into how women were being scored in competence, Wenneras & Wold (1997) concluded that women would have to outperform their male counterparts to be given the same score (e.g. be 2.5 times more productive than the average male applicant). They concluded that 'being of the female gender and lacking personal connections was a double handicap of such severity that it could hardly be compensated for by scientific productivity alone' (ibid., p.164). If we extend this idea to hypothesise about the role of female experts in this commission, we might therefore argue that female experts occupying circumferential positions within an evaluation team might play a lesser part in decision-making because of their lack of connectivity, which has nothing to do with their expertise in their given area. This notion could be tested in other evaluations and would let us think more carefully about whom we employ in evaluation teams in the future. I shall return to this point later when I look more closely at the decision-making structures involved in the commission.

Another aspect which might be misleading with this matrix is that the chairperson, who seems to have the least number of links and appears to be on the periphery, could be interpreted as being of less impact, which was clearly not the case. We should be careful therefore to avoid making too many assumptions about the importance of connectivity before the evaluation.

5.5 How experts viewed their selection

During the interviews I was rather surprised to learn that most of the experts, at least as far as I was able to determine, had not given much thought to the reason why they had been asked to become involved in this evaluation. As Expert F put it succinctly: 'the important thing is to *have been* chosen' (my italics)[110]. Some said that they believed they had been chosen for their specific subject knowledge (which was true, of course); others thought that it was because of their background as being from outside the state/country (obvious, in those cases where a foreign expert was asked), or because they came from a research institution that was not part of a university e.g. the Max Planck Institute. Hardly anyone stated that they thought it was directly because of their previous experience as an evaluator, which in some cases was considerable, as at least six of the experts had taken part in evaluations in other countries. I have already referred to some of the things the Ministry was looking for in their selection criteria. Frau Bolius told me more about this:

> In this particular case we thought that people with reputations in their fields would have greater acceptance. Although, as I've said, a certain mix is what's required. […] If one says, well, these are only the pinnacles of certain fields, then it can also be right to include experts of whom one can say they are totally "unburdened" and they can carry the weight. But everything can, as in medicine, have a positive and a negative effect. So by using a certain combination or design one has to try to avoid negative effects if possible.[111]

[110] Das wichtige Ereignis ist ja, dass man ausgesucht wurde.

[111] Aber in diesem Fall hier, sind wir schon davon ausgegangen, dass fachlich renommierte Leute eine höhere Akzeptanz haben. Wobei, wie gesagt, eine gewisse Mischung macht es dann. Und genau, wenn man sagt, ja, das sind ja nur jeweils die Speerspitzen bestimmter Richtungen, dann kann es ganz gut sein, wenn man da Gutachter drin hat, wo man sagt, das sind ganz unbelastete Leute in Anführungsstrichen und die tragen das mit. Also, wie gesagt, alles kann unter Umständen, wie bei Medizin, eine gute und eine schlechte Wirkung haben und man muss eben versuchen, durch eine gewisse Zusammensetzung, oder durch eine gewisse, durch ein gewisses Design von dem ganzen Verfahren, negative Effekte möglichst auszuschließen.

However, it seems that in some cases, even on closer inspection, it is not clear why selection took place. One expert said:

> Yes, I was surprised and I enquired again why I had been approached. And then they couldn't really tell me. I spoke to people, those with whom I telephoned, they didn't know either. Somehow my name had come up because they were satisfied with my work, or whatever that means.[112] (Expert L)

Others did not question their selection whatsoever, almost as if they had expected that they would be asked to participate. This might be due to their seniority and the fact that they had already occupied positions of assessor or advisor in any number of committees, and so to be asked to be an evaluator in this commission might seem normal to them. Those who worked in very specific areas imagined that it was for this expertise that they were selected, while others believed that it was because they had some previous knowledge of the state to be examined, having worked or studied there in the past. This last notion was wrong in fact, as the Ministry wished to avoid employing experts who might have worked in Baden-Württemberg because it was imagined that this could cause animosity.

One expert had deliberated for a long time before accepting the offer to participate and described what he thought of his selection as follows:

> It was about Educational Science with a focus on teacher training, but that did not include the whole of teacher education. This is critical insofar as it involves two ministries: School and Research. And that cannot happen in any *Land* without friction of some kind. Both have different viewpoints. And so they were looking for someone who was sort of trustworthy. But I don't think they were looking for someone who was politically close to them, but rather who was a force to be reckoned with and who separates his own credo from his scientific work. That was, I think, why they thought of me.[113] (Expert N)

[112] Ja, ich war überrascht und ich hatte auch noch mal nachgefragt, warum ich dazu aufgefordert werde. Und dann hat man mir irgendwie nicht genau sagen können, ich habe mit den Leuten, mit denen ich darüber telefoniert hatte, die wussten das wohl auch nicht so genau. Man sei wohl irgendwie auf mich gekommen, weil man mit meiner Arbeit zufrieden gewesen sei, oder was auch immer das bedeutet.

[113] Es ging um die Erziehungswissenschaft insgesamt mit dem Blick auf die Lehrerbildung, aber nicht die Lehrerbildung insgesamt eingeschlossen. Das ist insofern kritisch, weil zwei Wissenschafts- ... zwei Ministerien beteiligt sind: Schule und Forschung. Und das geht in keinem Bundesland völlig reibungslos. Beide haben unterschiedliche Perspektiven. Und die haben jemanden gesucht, der halbwegs vertrauenswürdig ist. Also ich glaube nicht, dass sie jemanden gesucht haben, der ihnen politisch nahe steht, aber der berechenbar ist und sein Glaubensbekenntnis trennt von der wissenschaftlichen Arbeit. Das war, glaube ich, die ... der Grund, weshalb sie an mich gedacht haben.

Following the request to participate, it was therefore the individual's choice as to whether they would take part in the evaluation or not. We know from the documentary analysis that some of the people who were initially approached were unable to take part because of conflicting engagements. What I would now like to examine are the different reasons why these particular experts chose to take part in this commission, as this will tell us something about their motivations.

5.6 Reasons for participating in evaluation

There were numerous reasons for deciding to take part in what can only be described as a lengthy and rather arduous task. I will look at each in turn.

5.6.1 Knowledge of institutions

One of the predominant reasons which motivated experts to participate is that they were interested to look at Baden-Württemberg because of its own special structure (a special *Lebenswelt* therefore), which many of them had no prior experience of and were keen to view from the inside. One expert was representative by putting it thus:

> My notion of what I could contribute to the commission was that I have experienced, or at least perceived of Baden-Württemberg as a *Land* in which Educational Science is very, very, very heterogenous. And that just fascinated me as a field of research, if you like. Well, I already knew some individual institutions through previous cooperations. But I didn't know the majority. And in particular I didn't know the teacher training colleges. And so from that point I thought, well, it will be quite entertaining to get to know all of that a bit better.[114] (Expert F)

This was an opinion which was echoed by many, as they wondered to what extent this particular model would work and what the quality of it would be like.

[114] Und meine Vorstellung von dem, was ich in die Kommission einbringen konnte, war, dass ich Baden-Württemberg als ein Land, als ein Bundesland erlebt oder zuvor wahrgenommen habe, in dem die Erziehungswissenschaft sehr, sehr, sehr heterogen etabliert ist. Und das hat mich einfach gereizt als Recherchefeld, wenn Sie so wollen. Ich kannte nun in Baden-Württemberg einzelne Hochschulen schon vorher recht gut aus Kooperationen. Aber die meisten eigentlich eher nicht. Und ich kannte insbesondere nicht die PHs. Und deswegen habe ich gedacht, ach, das wird doch mal ganz unterhaltsam sein, das man alles ein bisschen näher kennen lernt.

Another reason was that some had already participated in a similar evaluation elsewhere and were now interested to compare their experiences:

> This gives me an opportunity to learn about another *Land*, its systems and structures, which was the case with Baden-Württemberg. The other one was six, seven years ago, and it was very interesting to look at it again from the other side.[115] (Expert M)

Others regarded the evaluation as a means of establishing a kind of personal ranking: 'Otherwise, in relation to the *Länder*, it's always quite nice to see where they're at, to see where the disciplines are […]'.[116] (Expert G)

5.6.2 Curiosity because of peers

In addition to wishing to examine the particular situation in Baden-Württemberg, for some experts there was an implicit challenge in being involved in this manner with peers of their rank from all over Germany. They were therefore stimulated by the social aspect of the evaluation. As one expert put it: 'It wasn't bad getting to meet other colleagues'.[117] The non-specialist, who we already know had accumulated a great deal of knowledge about the different HEIs in Germany, felt that because there seemed to be clear deficits in the training of teachers, there was room for healthy debate, to which he could contribute his somewhat 'removed' opinion.

It was not clear from the interviews whether individual experts enquired at the recruitment stage who else might be invited into the commission. I suspect that this might have seemed an inappropriate question for them to ask. For those recruited somewhat later this information was readily available as the commission was already underway.

[115] […] das gibt mir die Möglichkeit, ein neues Bundesland in seiner Struktur und seinen Systemen kennen zu lernen und so war das auch noch mal für Baden-Württemberg. Da war ich dann schon sechs, sieben Jahre weg und jetzt von der anderen Seite noch mal darauf zu schauen, ich fand das hochinteressant.
[116] Ansonsten ist es ja auch immer ganz schön, wenn man so ein bisschen landesbezogen sieht, wo stehen die eigentlich, wo steht, wo stehen die Disziplinen […].
[117] Es war nicht schlecht, andere Kollegen kennenzulernen.

Expert M told me about being particularly interested in the criteria which would be established during the commission:

> Well, it was curiosity. What […] will the criteria be according to which comparisons are made? So there was basically a kind of tension as there are experienced people in it. How did they deal with this? I really went into it from the viewpoint that I can learn something here. Not only that, as I have a job to do, but I can also learn something.[118]

The point about the criteria is a large issue, and I will be dealing with that at a later point, as it was not addressed specifically during the recruitment stage.

5.6.2 Accommodation of evaluation into existing schedules

Considering that the experts were being asked to dedicate a large portion of their time to this evaluation over and above their normal duties as university professors, institutional directors and so forth, it was not surprising that the time aspect was raised when I asked them about their reasons for taking part. Most did not regard the time investment as being too problematic. Expert S told me: 'The onsite visits could be done in one week which meant not having to miss too much teaching, or rather it could be done for me by colleagues.'[119] If their position was such that they could organise sufficient coverage in their absence, then the expert could decide to take part. For others like Expert B, however, the fact that they were missing their own teaching meant that their attitude was more negative: 'Then we go back again [to own HEI] and have to catch up the lost day, which I could well do without.'[120] It is not possible therefore to say that the experience for all was one of being totally free from guilt about what their own HEI was missing because of their absence.

For those who were already in retirement (although retirement in academia should be put into parentheses because we frequently see retired academics even busier than when they were officially in post), work in the commission posed no particular time problems. Indeed,

[118] Also, es war Neugier. Was […] werden die Kriterien sein, wonach verglichen wird? Es war natürlich im Grunde auch so eine Art Spannung, also da sind ja erfahrene Menschen drin. Wie gingen die damit um? Ich bin wirklich unter der Perspektive rein, ich kann hier auch lernen. Also nicht nur, ich habe hier einen Job zu tun, sondern ich kann auch lernen.

[119] Die Begehungen konnten in einer Woche abgearbeitet werden, so dass dann nicht zuviel an Lehre ausfallen musste, bzw. meine Kollegen mich vertreten konnten.

[120] Dann kommt man wieder und muss den Tag wieder reinholen, also. Da konnte ich gerne drauf verzichten.

there might be a strong argument for always including at least a couple of retired academics in evaluation commissions, because firstly they are able to see more of the HEIs in question and gain a better picture of everything, and secondly, they will not be feeling guilty about what they are supposed to be doing at their own HEI. As Expert R put it:

> When you are retired, like me, then it's OK, but many of these people are still in full posts. They can't do a full three-week tour in the way we were on tour for three weeks. Yes, one week, five visits, one week back home, then another week with five visits, then another week back home, and the last week with five visits, that's pretty strenuous… Of course, it's better if you have a commission, but, yes, you won't get anyone who wants to do it.[121]

However, although the evaluation was associated with sacrifices, Expert H said: '[…] none of us has any time, but someone has to push things forward'**,** which indicates notions of personal responsibility, duty to the discipline, and a commitment to quality.[122]

5.6.3 Experience of being evaluated

I wished to discover the extent to which the experts in this commission had already been involved as evaluation experts or been asked to pass judgements on quality in other contexts (in editorial boards, expert committees for DFG, etc.). (See also Table 14 in Chapter 4)

A number had worked as evaluators before and for these members taking part in a new evaluation was either a useful means of comparison or else it was just another in a string of similar experiences. However, others were motivated to take part because they themselves had been evaluated, which some had experienced negatively. Expert O described to me the experience of being evaluated as follows:

> […] the evaluation team which came was incompetent in our opinion. […] They suddenly started discussions with us in a biased manner. […] If you have experiences of

[121] Wenn man pensioniert ist, wie ich, dann geht das schon, aber mehrere dieser Personen haben noch ein volles Amt. Die können ja nicht drei volle Wochen, denn wir sind drei Wochen auf Tour gewesen. Ja, eine Woche, fünf Besuche, eine Woche zu Hause, wiederum eine Woche mit fünf Besuchen, wiederum eine Woche zu Hause, und die letzte Woche die fünf Besuche, und das ist ein ziemlich anstrengendes... Natürlich, wenn man eine Kommission hat ist das besser, aber, ja, dann bekommt man keine Leute, die das tun möchten.
[122] […] wir haben ja alle keine Zeit und einer muss die Sachen vorantreiben.

> that kind, then you certainly become more sensitive towards those you are dealing with. […] [123]

Evaluators who are regarded as lower in status or less competent than those being evaluated will inevitably be taken less seriously and the entire evaluation will be in jeopardy (Morley, 2003). In this case, the previously negative experience had led to a kinder and more humble approach to those undergoing evaluation. This point was echoed by another expert thus: 'I think it is very important to have been evaluated enough yourself. I found that very helpful. It taught me a great deal, having been in the role of being evaluated in quite different contexts.'[124] (Expert G)

This notion of learning is quite important and I will go into greater detail on what precisely the experts learnt further down. Statements such as 'I still regard the whole thing according to the view that I can learn something from it'[125] (Expert H) were quite common, even in those with considerable evaluation experience. Indeed, it seems fair to say that none of the members regarded the evaluation as something that was self-evident and most would admit that they definitely learnt a considerable amount from taking part.

There was only one expert who had not participated in an evaluation before and had not been evaluated themselves previously. I thought that it would be interesting therefore to discover what this person made of the experience:

> For me it was very important, because if an evaluation is done here [in own HEI], then I can turn to my experience. I also know what's important about such an evaluation. And what one needs to include in the self-report. So, from that point, it was more of a personal gain. Well, a learning gain.[126] (Expert B)

---

[123] […] da war die Evaluationskommission, die gekommen ist, einfach aus unserer Sicht nicht kompetent. […] die haben also da unter einer sehr einseitigen Brille plötzlich die Diskussion etwa bei uns angefangen […] Wenn man solche Erfahrungen hat, da ist man schon sensibler, ja, gegenüber denen, wenn man denen gegenübertritt.

[124] Ich glaube, wichtig ist, einmal dass man selber genug evaluiert wurde. Das fand ich also sehr hilfreich. Da habe ich sehr viel gelernt, selbst in der Rolle gewesen zu sein, begutachtet zu werden in ganz unterschiedlichen Zusammenhängen.

[125] […] ich betrachte das Ganze auch immer noch unter dem Gesichtspunkt, was ich lernen kann.

[126] Für mich ist das insofern sehr wichtig, wenn eine Evaluation hier kommen würde, dann könnte ich auf die Erfahrung zurückgreifen. Ich weiß auch, was wichtig ist an einer solchen Evaluation. Und was man darzustellen hat in dem Bericht, also, von daher, eher ein persönlicher Gewinn. Also ein Lerngewinn.

The personal benefits accrued through participation in the evaluation were possibly tangentially greater for this member than for others.

5.7 Significance of 1st plenary meeting

In Chapter 4 I described what was discussed during the first plenary meeting in April 2003. Table 1 in Chapter 3 revealed, however, that a certain number of members were unable to participate in this first gathering in which some of the criteria and plans of action were debated. I wondered if this had any impact on the evaluation overall, or if it was of no particular relevance that some people had been absent. Depending on the viewpoint of the individual, the responses varied considerably.

One very experienced member of the commission was not overly concerned by the fact that he was unable to attend. His attitude was pragmatic inasmuch as he felt that the detailed minutes of the meeting gave him a sufficient knowledge base from which to progress into the evaluation at a later stage when he was able to attend. For others, however, being absent was more of an issue.

Firstly, the point was raised that there was a knowledge deficit to overcome, and so they felt that they were playing catch-up, trying to understand the thrust of the debate so far, which might not always be apparent from the minutes of the meeting they were provided with. One expert stated that they still had questions which they thought could no longer be answered.

Another point related to being unable to contribute to the discussion of the evaluation criteria: '[…] by being away from the first (meeting), I was naturally unable to introduce the criteria specific to my field. It would certainly have been good to discuss this beforehand.'[127] (Expert M) In those cases where subject areas are perhaps more specialised than those at the heart of a discipline and where the other evaluators are themselves less

[127] […] dass ich dann bei dem ersten nicht da war, habe ich natürlich auch da die Kriterien nicht einbringen können für dieses Feld. Das wäre sicher gut gewesen, das vorher zu besprechen.

experienced, it seems essential to have voices represented from the more marginal areas. Arguably, therefore, if this does not occur at an early stage, then the criteria which are put in place might not be entirely appropriate. As one expert who *was* present, stated:

> I believe it is very important, because we can already identify the problem that within Educational Science in Germany there are quite a number of notions present and so one cannot say from the outset that the criteria which generalists would employ are as suitable as those which are used in empirical research. And so from that point it is an important step in such a group, which consists of various sub-disciplines with partially varying methodological approaches, to reach consensus which leads to a direction where one can finally say that we have reached agreement in the criteria in spite of the variations of sub-disciplines or research approaches. I regard that as very important.[128] (Expert G)

One expert who was absent deeply regretted having missed the meeting. This person spoke about the preparation undertaken by the sub-committee in advance and indicated some of the problems that the team experienced: they had only a limited amount of time to get through the evaluation task and so compromises had to be made, to the detriment perhaps of allowing for a more differentiated discussion:

> Well, I was unlucky enough to be unable to take part in the first meeting and I discovered that I had lots of questions about it. There was no opportunity to discuss these questions. I am not sure if that was talked about, or if things had been decided in advance. The commissions work under pressure, people are doing this job alongside everything else. So they are glad if things are established somehow and then they push them through. There wasn't much enthusiasm to try to work out the questions for yourself. It was seen as more of a disturbance if questions of that kind were raised. It was regarded as more of a hindrance to the course of the work than seen as being useful.[129] (Expert L)

[128] Ich glaube, dass es sehr wichtig war, weil das Problem ja schon auch zu sehen ist, dass innerhalb der Erziehungswissenschaft in Deutschland doch eine relative Bandbreite doch von Vorstellungen existiert und man sicherlich nicht von Vorneherein sagen kann, dass die Kriterien, die allgemeine Pädagogen anlegen würden, dass die gut dem entsprechen, was man in der empirischen Forschung anlegen würde. Also von daher ist es sicher ein wichtiger Schritt in so einer Gruppe, die aus verschiedenen Teildisziplinen zum Teil mit teilweise unterschiedlichen methodologischen Ansätzen kommt, so eine Art Konsens zu erzielen, der in die Richtung geht, dass man am Ende sagt, bei aller Unterschiedlichkeit von Teildisziplinen und Forschungsperspektiven haben wir einen Konsens darüber, was die Kriterien sind, die wir gemeinsam anlegen können. Das halte ich für sehr wichtig.

[129] Nun hatte ich das Missgeschick, ich konnte an der ersten konstituierenden Sitzung nicht teilnehmen und fand dann etwas vor, wo ich ganz viele Fragen dazu hatte. Es gab keine Gelegenheit, diese Fragen zu diskutieren. Ich bin mir auch nicht sicher, ob das diskutiert wurde, sondern, ob das nicht von Vorneherein vorgegeben war. Die Kommissionen arbeiten ja dann auch unter Zeitdruck, die Leute machen das nebenbei. Da ist man froh, wenn da irgendetwas steht und das zieht man dann durch. Also, große Lust noch mal, also, die Fragestellung für sich selber mitzuklären, das war eigentlich eher eine Störung, wenn man solche Fragen formuliert hat. Also, es wurde eher als Störung des Arbeitsablaufes eher gesehen, als dass es etwas genutzt hätte.

Although this voice within the commission may have been in the minority, it raises a fundamentally important question about the quality of evaluation work: if experts are under too much time pressure, they may be more likely to cut corners or not engage as thoroughly with the subject matter than if they are given more time.

Another expert (present) came up with a useful methodological idea for those who were absent. The following statement reinforces what I have indicated above:

> If we had made audio recordings of the first meetings and later ones, then there would have been a different mode of participation. If one could present this as a sociogram you could have presented a different network than the one in the beginning. The dominance, the initial dominance, later made space for one of more equality, without reaching a completely balanced participation. However, in my opinion it certainly developed into a clearly stronger balance.[130] (Expert T)

This comment displays empathy for those who were not there; clarity about the fact that the evaluation was, as this person saw it, 'dominated', and that this stance shifted over time. What were these 'dominances', as just stated? Did they find their origin in the commissioning body, the Ministry of Science, Research and Art? Were the experts all aware of the ministerial perspective before they accepted this task? Or did these 'dominances' originate in different perceptions of their *Lebenswelt*, Educational Science? Let us turn first to the political aspects of this evaluation.

5.8 Extent of ministerial involvement

The documentary analysis in Chapter 4 has already shown that the commissioning body, the Ministry of Science, Research and Art, was particularly interested to see how well teacher training among *Gymnasium* teachers was working as deficits had been apparent for a number of years. As one expert indicated: 'The Ministry would probably not have

[130] Wenn man also, Tonbandaufnahmen von den ersten Sitzungen und von den späteren gemacht hätte, dann wäre sicherlich ein anderes, ein anderes Beteiligungsverhalten [gewesen]. Wenn man das wie ein Soziogramm darstellen würde, hätte man da ein anderes Netzwerk abbilden können, als zu Anfang. Die Dominanz, die anfängliche Dominanz ist später in Richtung mehr Egalität, ohne bei vollkommen ausgependelter Beteiligung zu sein, zu landen. Aber es ist auf jeden Fall auf eine deutlich stärkere Ausgeglichenheit hin, hat sich das nach meiner Wahrnehmung, hat sich das hin entwickelt.

commissioned the evaluation without the PISA study.'[131] (Expert H) Germany's poor results had given the entire country a shock and everyone was searching for answers as to why German schoolchildren had performed so badly. The Ministry believed that educational research might shed light on what best practice might be, and so empirical research was also in the spotlight of its attention as this might inform teaching. To some extent, of course, the fact that the PISA results were poor gave the Ministry an underlying argument to implement changes as it now had a lever with which to push reform through.

As teacher training obviously related to what went on in schools, it was necessary to involve a second Ministry, the Ministry of Culture, Youth and Sport, and so a permanent representative from this Ministry was invited to participate in the discussions alongside the experts. When I asked him if he thought his presence or that of the other ministerial representatives had any impact on the work of the commission, he told me: 'They were really very upright and independent personalities, who didn't, I think, let themselves be influenced or intimidated in any way.'[132] Bolius, too, also spoke to me about her impressions of what she thought the group felt about the Ministry being there:

> Well, I wasn't under the impression that any of the experts […] had a feeling of being prevented from saying something. My impression rather was that there was every possibility to do so, and that it was taken up. But if someone didn't do so, they did it for other reasons, because they didn't have enough knowledge on that topic, or because they didn't wish to engage themselves, or because they didn't say anything for whatever reason. But I didn't get the impression that any of the experts, be they external or foreign, wanted to say something, but preferred not to do so. I don't believe that.[133]

Let us discover if these impressions were corroborated by what the experts said in interview about the extent of ministerial involvement in the evaluation.

---

[131] Wahrscheinlich hätte das Ministerium das ohne die Pisastudie auch gar nicht in Auftrag gegeben.

[132] […] das waren wirklich sehr integere und autarke Persönlichkeiten, die sich davon auch nicht in irgendeiner Weise beeinflussen oder einschüchtern lassen, glaube ich.

[133] Also, ich hatte jetzt nicht den Eindruck, dass […] einer der Gutachter das Gefühl hatte, er fühlt sich gehindert, etwas zu sagen, sondern mein Eindruck ist eher, dass die Möglichkeit auf jeden Fall da gewesen ist, und auch genutzt wurde, aber wenn jemand das nicht gemacht hat, das aus anderen Gründen nicht getan hat, weil er nicht genug Wissen darüber hat, oder weil er sich nicht engagieren wollte, oder weil er dazu nichts sagt, aus irgendwelchen Gründen. Aber ich habe nicht den Eindruck, dass einer der Gutachter, extern oder ausländisch, sagen wir mal... gerne etwas gesagt hätte, aber es lieber nicht getan hat. Das glaube ich nicht.

The experts were not aware from the outset that their work would be 'officially' accompanied, or in the words of one person: 'That wasn't clear to me when I signed my contract.'[134] (Expert P) The foreign experts, in particular, found this concept rather alienating, notably the chairperson, whose experience of the Dutch evaluation system had been one of complete autonomy. Two of the other foreign members said that ministerial participation was 'unusual and not completely desirable', 'exciting, but critical at the same time' (Experts U and E respectively)[135]. It is arguable whether all of the experts would have accepted the task had this been known from the outset.

Equally, although the remit was an analysis of the quality of the 15 HEIs, there were still numerous aspects which were unclear when the commission began its work: 'We didn't really know what the precise aims of the Ministry were. Did they want to save money? Did they wish to simplify structures? Did they wish to create specialisations? That wasn't clear.'[136] (Expert U) One expert put it thus:

> They were present, the ministerial representatives. And that was certainly something we needed to discuss, if they should be there, because they are in part the people to whom this is addressed. On the other hand, [the evaluation] *is* aimed at *them* [my italics], and so they should hear the discussions and should contribute to some extent on the things which the commission didn't know […] [137](Expert O)

The following extracts further reveal the problem the experts faced in having a Ministry as their commissioning body. Referring to both the Ministry's position and the role of Ulrich Lübke Expert L said:

> I don't think that's good for an evaluation. Either you decide to have an independent evaluation and accept that. I mean: as a government you are free to decide how to react to it. So, you're free to take whichever political decision on the further development of

[134] Es war mir nicht so bewusst als ich den Vertrag unterschrieben habe.
[135] […] außergewöhnlich und nicht unbedingt erstrebenswert; aufregend, aber kritisch zugleich.
[136] […] und man hat ja nicht gewusst, was exakt, was präzise die Zielsetzung des Ministeriums wäre. Möchte man Geld sparen? Möchte man die Struktur vereinfachen? Möchte man Konzentrationen in Gange setzen? Das war nicht klar.
[137] Die waren ja dabei, die Ministerienvertreter. Und das ist sicher ein Punkt, wo man diskutieren muss, ob die dann dabei sind, weil es ja die Adressaten zum Teil sind. Andererseits richtet es sich auch an die, dann sollten sie die Diskussion mitkriegen, und, er, sollten auch zum gewissen Grad Dinge einbringen, die die Kommission nicht wusste.

> teachers. Yes, or about teacher training. I don't think it's good from the outset. I don't think it's basically good. That has nothing to do with individuals. He [Lübke] was a solid, good civil servant in his subject area, that's not the problem. […] But I think if you are going to have an evaluation then you should have the courage, even as a government, to say, we'll listen to what they have to tell us and then we'll have a look at it.[138]

Another expert told me how his initial hesitation about having the Ministry alongside their work transformed to one of acceptance:

> As the evaluation went on, I must say that the idea of accompaniment, one of passive accompaniment in this evaluation through representatives of the commissioning body, gained ground with me. I was sceptical at first as I feared that independence might not be guaranteed. That wasn't the case, however. Or rather, it contributed to being able to understand the remit more clearly whenever this was unclear, which I experienced in a very positive way.[139] (Expert E)

Not all members shared this view, however. Another expert was unequivocal in their antipathy towards ministerial involvement of any kind: 'It was in part difficult to tolerate, even their presence. We voiced that in different ways and had to voice in different ways that we were independent. […] I would have preferred the Ministry to show less presence'.[140] (Expert F) Overall, however, an attitude of tolerance towards the Ministry developed slowly among the commission. A cynical view might describe this, however, as resignation.

Two ways in which the commissioning Ministry made its presence especially felt throughout the evaluation was in relation to its remit on maintaining the structure of the teacher training college for all teachers beneath the level of *Gymnasium*, in addition to not wishing to close any of the HEIs subject to this evaluation. This was not unproblematic for

[138] Ich finde das für die Evaluation nicht gut. Entweder man entscheidet sich zu einer unabhängigen Evaluation und nimmt die dann zur Kenntnis. Ich meine: man ist ja immer noch frei als Regierung zu entscheiden, wie man sich verhält. Also, die politische Entscheidung kann man immer noch so treffen über die Lehrerfortbildung, wie man das möchte. Ja, oder Lehrerausbildung. Das finde ich vom Ansatz her nicht gut. Ich finde es prinzipiell nicht gut, ganz prinzipiell nicht. Hat nichts mit den Personen zu tun. Das war ein gestandener, guter Fachbeamter, überhaupt nicht das Problem. […] Ich finde, […] man sollte den Mut haben, auch als Regierung, wir hören uns jetzt mal an, was die zu sagen haben und damit setzen wir uns auseinander.

[139] Im Verlauf der Evaluation muss ich sagen, dass diese Idee der Begleitung, aber der passiven Begleitung dieser Evaluation, durch Vertreter des Auftraggebers... das hat bei mir an Boden gewonnen. Am Anfang war ich sehr skeptisch, weil ich befürchtete, dass da die Unabhängigkeit nicht gewährleistet sei. Das war aber nicht der Fall, sondern es hat eher dazu beigetragen, dass man den Auftrag immer dann, wenn er nicht so klar war, konkretisieren konnte und das habe ich sehr positiv erlebt.

[140] Das war zum Teil schwer zu ertragen, auch die Präsenz. Wir haben das auch verschiedentlich artikuliert und haben verschiedentlich artikulieren müssen, dass wir unabhängig sind. […] Ich hätte mir gewünscht, dass das Ministerium weniger Präsenz gezeigt hätte.

members of the commission who believed that they should be entirely free to make recommendations of any kind if they thought them to be appropriate. As most of them came from within the university environment which had seen an incorporation of teacher training colleges into the universities, we can assume that their overall attitude towards the Baden-Württemberg model was one of scepticism, although they were curious to see how effective the model was. As one person said: 'We knew those were the guidelines, […] but we could live with that. […] The commission could write that it didn't agree with them in principle, but that they accepted them.'[141] (Expert N) Another expert said that it was difficult to come to terms with restrictions: 'We can think about the most diverse types of cooperation model, perhaps in relation to technical colleges and teacher training colleges, but an integration model is not what they want.'[142] (Expert T) One of the foreign members, who was totally removed from internal German politics, raised the following point:

> I sensed among the professors, well, even among the German experts, an incredible caution, which can only be explained by the fact that this is a small *Land*, everyone knows everyone else and somehow you sense your colleague's problems. And then you might be evaluated yourself. And that's really bad for an evaluation.[143] (Expert E)

This displays an acute awareness of where this evaluation was restricted and shows that any recommendations could only be made in part, because the experts were thinking beyond the evaluation and considering the personal implications for certain decisions that they might make. In other words, this is where experience of university administration or policy is invaluable, although, if we reverse the coin, it might be regarded as a hindrance because it inhibits decision-makers from making objective decisions for the simple reason that the expert with experience of this kind is already aware of what will be allowed to work in policy and what will not.

---

[141] […] wir wussten, das sind die Rahmenbedingungen […], aber damit konnte man ganz gut leben. […] die Kommission konnte schreiben, dass sie die sachlich nicht richtig findet, aber dies akzeptiert hat.

[142] Da kann man nachdenken über die verschiedensten Kooperationsmodelle, vielleicht auch hinsichtlich Fachhochschulen und PHs, aber eine Integrationslösung wird nicht gewünscht.

[143] Ich spürte bei den Professoren, also, unter den deutschen Experten selbst eine unheimliche Vorsicht, die nur erklärt werden kann, weil man ist ja ein kleines Land, man kennt sich ja doch, und irgendwie spürte man da Kollegennot. Man wird ja dann vielleicht selbst auch mal evaluiert. Und das finde ich ganz schlecht für eine Evaluation.

An additional area which had not been discussed with the experts in advance of the evaluation concerned how much the Bologna Process should be implemented within the state, a decision which obviously had to come from state level and not be left to each HEI to decide for itself. The experts had not been offered clear guidelines from the Ministry on this point:

> The second advance decision where I thought that the political commissioning body could have been more open related to the question of course design, the introduction and implementation of the Bologna Process. It was certainly obvious that debates about modularisation were taking place everywhere, but insofar as this was leading towards BA or MA structures, I'd have preferred more clarity on that from the outset.[144] (Expert T)

We therefore see a dichotomy arising. The commission wished to function independently and be free to make its own suggestions to the HEIs. However, as these are financed by the state, the experts needed to know from the Ministry about certain areas of concern, which, as identified above, had not necessarily been communicated clearly to the group in advance of the evaluation. Those that were communicated restricted their ability to make recommendations.

I will now examine how the experts described their views on what they believed the HEIs made of the ministerial involvement and the fact that the Ministry was working alongside them in this evaluation.

5.9 Experts' opinions on how HEIs perceived evaluation

When this evaluation was conducted, the HEIs in Baden-Württemberg had not experienced many evaluations previously and it was the first time that Educational Science had been given such intense external scrutiny. The procedure was therefore new to the majority of HEIs involved and the experts on the commission knew this to be the case.

[144] Und die zweite Vorentscheidung, wo ich auch fand, dass die politischen Auftraggeber, hätte ich mir an der Stelle offener gewünscht, da ging es um die Frage der Studiengangsgestaltung, also, die Einführung, die Umsetzung des Bologna-Prozesses. Da hätten wir uns eigentlich auch eine etwas stärkere Offenheit gewünscht. Das, was auf jeden Fall klar war, dass überall Modularisierungsüberlegungen schon in Gang waren, aber inwieweit das denn nun in Richtung BA/MA-Strukturen geht, also, da hätte ich mir von Vorneherein mehr Offenheit gewünscht.

When I asked them to tell me what they thought about how the HEIs regarded them as a group and the work they were conducting I received many conflicting answers, which surprised me because I had imagined that there would be more consensus of opinion. The majority (mostly men) held that the HEIs had no problem with the commission and that the HEIs generally regarded the evaluation as an opportunity to engage in reflection and see where improvements could be made (Goetze, 2005). One female expert spoke about the expectations that she felt the teacher training colleges placed in the evaluation:

> In the teacher training colleges in particular I had the impression that they pinned their hopes more on getting an objectivising outside view, that we should perhaps gather information on particular problems, workloads, overloads and lacks in resources and make strong recommendations. Those were the expectations that more or less came to the surface.[145] (Expert T)

However, the fact that the Ministry had commissioned the evaluation meant that some experts thought that the HEIs developed a particular stance towards the experts:

> But I think that for those being evaluated the fact that the Ministry was so heavily involved certainly had a threatening character. […] That they brought the Ministry in to this extent I can only explain as follows, I tried to understand why they did it like that. It certainly has to do with the fact that teacher training is an affair of the state. Teachers are civil servants. So, state inspection made its presence felt in this way.[146] (Expert L)

Expert F used the word *Angst* to describe the HEIs' reaction to the evaluation commission, which they viewed as rather unnecessary because they did not think of themselves as being 'scary':

> We didn't set forth with the intention of punishing anyone. So, it's a kind of feeling that you are being misinterpreted, and secondly, there's a strong feeling of having to defend yourself against things or traditions which seem to have worked well in the past, but

[145] Bei den Pädagogischen Hochschulen hatte ich schon den Eindruck, dass da stärker die Hoffnung dran geknüpft war, dass jetzt ein objektivierender Blick von außen, vielleicht auch noch mal die besonderen Probleme, Belastungen, Überlastungen und Ausstattungsmängel zur Kenntnis nehmen und mit Nachdruck auch formulieren könnte, das waren Erwartungen, die da mehr oder weniger deutlich formuliert zu Tage traten.

146 Ich glaube aber, dass das für die zu Evaluierenden, der Umstand, dass das Ministerium da so dick drin hängt, durchaus eine bedrohliche Seite hat. […] Das man hier jetzt das Ministerium so weit reingeholt hat... Ich erklärte mir das so, ich habe auch versucht zu verstehen, warum machen die das jetzt. Das hat sicher damit zu tun, dass die Lehrerausbildung eben noch Staatssache ist. Lehrer sind ja Beamte und so. Also, dass da die Aufsicht des Staates sich darin auch ausgedrückt hat.

> where people on the inside should have basically noticed that they weren't really working.[147]

There was also a fear of the unknown apparent in the HEIs, as there was no background of evaluation experience: 'They don't really know what it's about. They are all really scared that they may suffer cuts, that jobs might be taken away. And this fear is quite justified.'[148] (Expert F)

Although it might seem like typecasting, I did get a sense that the male members in the commission had a more pragmatic notion of how they were perceived as experts, and how the HEIs regarded their work overall, whereas the female members were more willing to speak about the emotions of fear and trepidation that they believed the HEIs held when the expert committees arrived. There was also a difference in the manner in which the teacher training colleges regarded the commission in contrast to how the universities regarded them, possibly as a result of the experts all coming from universities and one technical college: 'In the teacher training colleges I had the impression that we engaged in conversation with them the least amount as I sensed that there was an atmosphere of animosity towards the commission.'[149] (Expert F)

In terms of who was on the commission, Expert G told me:

> Well, the HEI gets an impression of the different people so to speak, yes, they can live with this commission quite well as they can see so and so, and so and so, and so and so, who we all know and that's OK. Well, I see this as a process which is helpful for the acceptance of the commission.[150]

147 Wir waren ja gar nicht so mit dem Interesse, jemanden abzustrafen, da losgezogen. Also, das ist so irgendwie das Gefühl, man fühlt sich falsch eingeschätzt und zum Anderen, eben mit einer ganz starken Verteidigungshaltung gegenüber Dingen oder gegenüber Traditionen, die sich angeblich bewährt hatten, wo man aber faktisch eigentlich auch von innen heraus schon hätte merken müssen, dass sie sich nicht wirklich bewährt haben.

148 Die wissen eigentlich gar nicht, worauf es dabei ankommt. Die haben alle fürchterliche Angst, dass ihnen etwas gestrichen wird, dass ihnen Stellen gestrichen werden. Das haben sie auch zu Recht, diese Angst.

149 Also in den Pädagogischen Hochschulen hatte ich am wenigsten das Gefühl, dass man wirklich ins Gespräch gekommen ist, weil ich das Klima dort als mehr oder weniger feindselig gegenüber dieser Kommission erlebt habe.

150 Da hat eine Einrichtung den Eindruck über die verschiedenen Personen, sozusagen, ja, eigentlich kann man mit der Kommission schon ganz gut leben, da gibt es den und den und den und den und die kennen wir alle und das ist okay. Also, das sehe ich als wirklich einen Prozess, der für die Akzeptanz auch wieder hilfreich ist.

We know that the HEIs certainly respected the 'weight' of the commission members. This leads me into a discussion of the role of the foreign members in this group and the impact they made.

5.10 Specific roles within evaluation commission

5.10.1 Role of foreign experts

It is widely regarded as best practice to include foreign experts in evaluation commissions (Vroeijenstijn, 1992). If we examine the other evaluations which *evalag* had conducted before this case study, then we discover that without exception foreigners had been included. These usually came from neighbouring countries such as Austria and Switzerland, where the experts also had German language skills. However, others came from France, Italy, and in this case from Great Britain, Belgium, Switzerland and The Netherlands. Having four foreigners on a commission was quite unusual, as the previous *evalag* evaluations had only included one, or a maximum of two.

Let us examine the role of the foreign experts from two angles: firstly, their own view of their position within the commission, and secondly, how the German experts regarded their role and their contribution.

Expert Q, one of the foreigners, summed up his role thus:

> I think that was very important to have some perspectives from outside as I would call it. That was very very important because you look at it more from a distance. I won't say the others are not objective, but you have more distance to look at it which is important, in a way which I will not say to correct the views from inside, but it has an added value, I think.

The idea of 'added value' is interesting because although the very presence of external experts certainly raises the kudos of an evaluation, it does not necessarily mean that the outside contribution will be extensive. Expert E told me about his position as he saw it:

> I was the foreign observer. As a [citing nationality], I could occasionally say things which the Germans possibly didn't dare to say. This is the big advantage of foreign experts. […] The foreigners can say how they feel about things, which is very good. So it is essential for foreigners to take part. However, we were quite aware that our voice wasn't of any great concern. Our universities are somewhat different… […][151]

This remark shows that although he felt that he had a contribution to make, he believed that his position was a weaker one than that of a German expert. Coming from a different country, the likelihood of being able to contribute fully to the discussion from the outset was therefore less. This opinion is reinforced by the following statement which he made later on in the interview:

> But it was clear that you as British or I as a [citing nationality] had a reserved caution at first and at some point we entered into the proceedings. And then I or other foreign experts were invited to participate more. It doesn't have to be equal, certainly not.[152]

Expert U told me that he felt the evaluation did not 'capitalise on external experiences'. However, during the evaluation he had felt reluctant to refer to comparisons in his own country as this might sound prescriptive in a negative sense. This presents a dilemma therefore, as information which might be of use in the debate is at hand, but there may be a social inhibition to use it where it might prove helpful. This expert indicated that he believed more rigour could be established to make more of the foreign contribution. However, as the discussions were led by the chair, much depended on the manner in which the respective chairperson involved the external opinion. I believe that this is an important point to consider in other evaluations where foreigners are invited to participate. They should not just merely be involved as tokens, so to speak, but proactively encouraged to make comparisons with their own countries and not reserve these opinions to when they are directly asked. Let us now examine the second angle and see how the German experts viewed their foreign counterparts.

---

[151] Ich war ja ausländischer Beobachter. Als (citing nationality). Also ich konnte ja auch mal Sachen sagen, die die Deutschen vielleicht nicht wagten zu sagen. Das ist der große Vorteil der ausländischen Experten. […] die Ausländer, die dürfen sagen, was sie jetzt da empfinden, das ist sehr gut. Also darum sollten auch unbedingt ausländische Experten dabei sein. Aber, wir merkten ganz genau, wir hatten da nichts zu vertreten. Unsere Universitäten, die sind ein bisschen anders […].

[152] Aber es ist natürlich schon klar, sowohl ihr als Engländer, wie ich als […], wir hatten zunächst auch eine schüchterne Vorsicht und irgendwann kamen wir dann mal. Und da würde ich, ich würde einige ausländische Experten mehr einladen. Es […] muss nicht paritätisch sein. Das sicher nicht.

Overall, the German experts had a very positive opinion of their peers from other countries and used expressions such as these to describe their work: 'enlightening', 'broadened horizons', 'cast a new light on things'. They also appreciated their perceived political neutrality:

> I really found the work of foreign colleagues helpful again, as their positions are above suspicion. As colleagues from the German scene you are always somewhat involved in this competition-based, ranking-driven climate, you yourself are partially concerned and so the work of our foreign colleagues is very helpful once again.[153] (Expert T)

However, they were specific about exactly what they wished their foreign peers to achieve. They should not simply come with systems from other countries, expecting them to be adopted into German structures as this might cause resistance[154]. They should have empathy for the systems in place and make suggestions where other notions could be helpful. One expert highlighted the origin of the home country of each expert, saying that the combination of countries was something to be considered before appointing externals to a commission. This expert suggested that representatives from the Nordic countries would also have been useful in this evaluation, although they felt that the four countries represented provided a good balance of opinions. The foreigners in this constellation were appreciated as being courteous and able to diffuse conflicts. Expert K corroborated the foreign expert's statement about being reserved:

> They were mainly reserved, so to speak, which was perhaps due to courtesy or whatever, but they could and did articulate whatever they spotted. That much was clear. They've got the outside view, which makes one stop and think sometimes.[155]

---

[153] Da fand ich zum Teil auch wirklich die Mitarbeit auswärtiger Kollegen noch einmal sehr hilfreich. Weil die selber interessenunverdächtig sind. Als Kollege aus der deutschen Szene ist man ja immer ein Stück weit in diesem konkurrenzorientierten, durch Ranking-Vergleiche angefeuertem Klima, ein Stück weit mit betroffen und da ist die Mitarbeit von auswärtigen Kollegen noch einmal sehr, sehr hilfreich.

[154] Selbst würde ich trotzdem erwarten, von denjenigen, die man in diese Evaluationskommission beruft, dass jemand doch bereit ist, sich auch auf die spezifische Struktur und Geschichte des zu evaluierenden Landes oder Bundeslandes einzulassen, denn das halte ich doch schon für wichtig, dass man nicht britische oder holländische Strukturen einfach den deutschen Universitäten überstülpt. Das kann nur Widerstand hervorrufen und führt zu nichts Gutem.

[155] Die waren zwar insgesamt eher zurückhaltend, sozusagen, das ist auch vielleicht ein Gebot der Höflichkeit, oder wie auch immer, aber […] sie konnten und haben alles artikuliert, was ihnen auffiel. Das ist schon klar. Sie haben eben den Blick von außen. Das […] stimmt einen dann schon manchmal nachdenklich.

We can conclude from these statements that the foreigners were regarded as an essential component of the evaluation. However, their origin and the combination of countries present needed to be thought over carefully before appointment. Furthermore, as we can assume that the externals would feel especially vulnerable and in a position of courtesy towards their host country, more might be done to tell them from the outset what was expected of them and give them clear guidelines as to their engagement. Similarly, more could be done before the evaluation gets underway to inform the foreign members on the intricate nature of the internal politics that exist in the country under inspection.

Let us now turn to the role of one particular foreign expert who also occupied a crucial position within the commission: the chairperson.

5.10.2 Role of chairperson

We know from the archives that De Corte had initially been asked to take over the chair of the commission, but that he declined because he did not think his German language skills were sufficient. Fortunately, the Ministry found an equally experienced person to take over this position: Wynand Wijnen from The Netherlands, who had worked in many previous evaluations and had extensive knowledge of the German educational system. He was the only member of the evaluation commission who, together with Martina Röbbecke from *evalag*, had visited each HEI in preparation for the work involved. This was very useful in terms of having an overview of all the HEIs, their specific requirements, structures, etc. How did the German experts regard their chairperson who was a foreigner, albeit someone with immense evaluation experience?

Without exception his input was viewed positively. He was mainly perceived as occupying a moderatory role, rather than being someone who moved the commission forwards in terms of contextual issues, which were mainly driven by the other two members of the sub-

committee[156]. He also managed to keep the German experts on track if their discussions became axiomatic:

> As soon as the discussions turned to fundamental questions, which the Germans tended towards, he very quickly steered things away from that. […] That was how he managed to get the pinnacle of German Educational Science to reach a result.[157] (Expert D)

As Expert E told me: '[He] was the evaluation specialist, and as he was someone with an outside view he questioned absolutely everything. I found that fascinating.'[158] Wijnen was admired for being able to ask the HEIs basic questions about the necessity of particular aspects of their management, structures, etc., asking if they could imagine things being done differently: 'And that is something which I think evaluation commissions of this type need: a fundamental willingness for developments… […].'[159] (Expert E) A component of fearlessness was therefore required in this special position and as a foreigner he was in an ideal position to challenge the German status quo. He also had the calibre and the standing to gain the immediate respect of the HEIs. However, Expert U told me that he felt the chair could have brought his considerable evaluative experience into the proceedings more in terms of comparing the German HEIs with those in his native country or those in other evaluations he had been involved in. How did Wijnen himself therefore see his own position within the commission?

Firstly, he saw his role as being responsible for ensuring the smooth running of the evaluation: he was 'more of a technical chairperson, less so one on content.'[160] He also tried to ensure that everyone within the commission was given enough space in which to articulate their opinions. Expert K told me that he was surprised when the chair used a

[156] Der Herr Wijnen war ja Vorsitzender und hat sich, wenn ich mich erinnere, in den Diskussionen, in den aktuellen Diskussionen, eher zurückgehalten, hat eher zusammengefasst, gebündelt, strukturiert, versucht die Sachen, auch wieder in eine Richtung zu bringen. Er hatte also eher nach meiner Erinnerung die Moderatorenrolle eingenommen.

[157] Sofort, wenn es da um fundamentale Fragen ging, zu denen die Deutschen neigen, hat er sehr schnell wieder ausgeblendet. Insofern schaffte er […], dass die mit dem Spitz der deutschen Erziehungswissenschaft besetzte Kommission schließlich zu einem Ergebnis kam.

[158] [Er] war der Spezialist für Evaluation und auch derjenige, der so als Außenbetrachter wirklich alles grundsätzlich in Frage gestellt hat. Das fand ich faszinierend.

[159] Und das ist auch etwas, was es braucht, bei solchen Evaluationskommissionen. Eine grundsätzliche Offenheit für Entwicklungen... […].

[160] […] ein technischer Vorsitzender und nicht so sehr ein inhaltlicher.

specific tactic to ensure this. At the end of each round of discussions Wijnen would ask each individual in person whether they had any further contribution to make. I was told that this method was unusual in German academic contexts, but that the group quickly became used to this modus operandi. Wijnen was particularly concerned about situations in which tempers might start to fray if disagreements were apparent, stating: 'So for me it was very good to 'sedate' everything and to prevent wrong things from encroaching.'[161] This ability was certainly required in those onsite visits where internal disputes almost prohibited sensible and constructive discussion. Let us now examine the sub-committee which worked alongside Wijnen and formed a specific *Mitwelt* with him.

5.10.3 Role of sub-committee

The sub-committee brought together three experts with considerable evaluative experience and knowledge of Educational Science within Germany as a whole. This was slightly unusual, as it is more common to find one chairperson and vice chairperson in these committees. As Jürgen Baumert was Director of the Max Planck Institute for Human Development in Berlin it was unclear from the outset whether he would have the time to participate in every meeting. This may have been the reason for employing a second member of the sub-committee to support him.

When I asked the other experts to describe their view of the work of the sub-committee, nearly all told me that they found the group easy to get along with and they believed that each member provided knowledge which complemented that of the others: Horstkemper provided expert knowledge in terms of teacher training and general educational research, Wijnen had international evaluation experience, and Baumert provided his experience of other evaluations, the establishment of evaluation criteria from these, knowledge of empirical educational research and his own experience of institutional leadership. This is how one of the foreign experts described the sub-committee:

[161] Denn für mich war es ja ganz wichtig, dass wir keinen Streit bekommen. So, für mich war es sehr gut, um alles zu sedieren und zu sehen, ob nicht irgendwo falsche Sachen hineinkommen.

> They agreed with each other. And so it was of course clear and that was the interesting thing, although Wijnen was the chair […] I think Baumert was the man who had of course a lot more information from the inside, […] and Jürgen Baumert is a leading figure and he is recognised as one of *the* people in Germany in the educational field, and so he was like a natural leader from the German side. (Expert Q)

This was supported by what the other experts said and they accepted that the chair was responsible for the organisational aspects, whereas Baumert and Horstkemper helped more in terms of contextual matters. In terms of decision-making, the entire group accepted that the evaluation was steered largely by the sub-committee, which was not seen in a negative light, although when I asked about the impact factor I was told that the evaluation might have been different had Baumert not been in the sub-committee. Let us examine therefore how this particular expert impacted more specifically on the commission.

5.10.4 Role of Baumert

Because of his experience of working in The Netherlands in other evaluations he was seen as a natural leader as the Dutch system was viewed as the method of choice:

> We always had this Dutch model in the background, which Baumert in previous years had said would be good for German Educational Science, if we had such a catalogue or such criteria as those in The Netherlands. And indeed, that's how they [the criteria] developed and we *did* take direction from them, modified of course according to the German situation.[162] (Expert O)

Expert S was more specific about the precise nature of these criteria, to which I shall return in Chapter 6:

> At least insofar as research went, and it's my impression that they weren't transposed directly, but that we were influenced by the Dutch evaluation in terms of internationalisation, third-party funding, ranking, because we were influenced by that and we were told that we could not and should not avoid the internationalisation process and that we must at some point start to adopt these standards in our own HEIs.[163]

[162] […] als Hintergrund [war] immer dieses holländische Modell, den Baumert damals schon wo wir zusammen waren, die Jahre da davor, [sagte] das würde den Deutschen in der deutschen Erziehungswissenschaft auch gut tun, wenn man so einen Katalog hätte oder solche Kriterien [hätte], ja, wie die dort in den Niederlanden die Erziehungswissenschaften eingeschätzt haben. Und in der Tat [haben wir uns] an denen orientiert, natürlich modifiziert für deutsche Verhältnisse ist das dann damit entstanden.

[163] Also, zumindest was die Forschung anging [ist] mein Eindruck schon, dass zwar nicht eins zu eins umgesetzt wurde, aber doch Maßstäbe dieser Evaluation in den Niederlanden was Forschung angeht, was Internationalität angeht, was Drittmitteleinwerbung angeht, was Ranking angeht, dass wir davon auch mit

We already know that Baumert worked hard to assimilate the reports which had been written by the other group members after their onsite visits. Expert D described this aspect of his work:

> As far as I can recall, Herr Baumert designed the draft of the report, or certainly worked on it significantly… and discussions always led to modifications or insertions. For me this work was characterised by a high degree of non-prejudice.[164]

So, although we can say that he had a major impact, the commission was nevertheless able to work creatively under his leadership as he was willing to take new ideas on board and adapt established criteria to the specific environment in Baden-Württemberg. However, the following comment is extremely important in our understanding of the work of this evaluation overall: 'Well, it isn't the case that this was an objective, completely objective process. It was rather influenced. The people who take part do that.'[165] (Expert E) The expert went on as follows: 'But this also makes it exciting. On the one hand it is shaped by such a person [meaning Baumert]; on the other it is this which gives it profile. You can engage with it and that's how we get a new structure for the future.'[166] Baumert had been given the power to lead the commission and his input was respected and accepted (see also Fetterman, 1995; Fetterman, Kaftarian & Wandersman (eds.) 1991).

5.11 Summary

We have seen how networks among evaluators can play a role in how they work. Also, we have seen that evaluators had many reasons for wishing to engage in the evaluation, of which curiosity, a desire to engage in academic discussions with equal sparring partners, and a wish to see what was happening in other HEIs were some of the main reasons.

---

beeinflusst worden sind, weil es dann hieß, wir können und wollen diesen ganzen Internationalisierungsprozess auch nicht entgehen und wir müssen uns, müssen irgendwann mal anfangen, diese Standards auch auf unsere Hochschulen anzulegen.

[164] Herr Baumert hat ja, wenn ich das richtig in Erinnerung habe, den Entwurf des Berichts gemacht ... oder jedenfalls entscheidend daran gearbeitet... und die Diskussion hat immer wieder auch zu Modifikationen und Ergänzungen geführt. Für mich war das ein in einem hohen Maß von Unvoreingenommenheit gekennzeichnetes Arbeiten.

[165] Also, es ist überhaupt nicht so, dass das eine irgendwie objektive, völlig objektives Verfahren wäre, sondern das wird geprägt. Es gibt Profil von den Leuten, die das machen.

[166] Aber es ist auch spannend. Da ist zum einen (…) eine Prägung durch eine solche Person, zum anderen gibt es natürlich auch Profil. Man kann sich daran reiben und dann gibt es eine neue Form für die Zukunft.

Furthermore, we have established that there were differences in attitude among those evaluators with more experience in comparison with those with less, and the impact that this may have had on their respective inputs in discussions and decision-making. I discussed the role of the foreign members in the commission, their input, and some of the limitations they may have experienced as foreigners that were not due to language problems, but rather cultural issues. Specifically, we have seen the crucial influence of particular members within the group, notably the sub-committee, and the restrictions placed upon the group by various ministerial decisions which pre-dated the evaluation.

This leads me now into a description of the next phase of work which the evaluation commission had to deal with: the establishment of criteria to be used in their assessment of quality in the 15 HEIs under investigation. Let us then examine how these were created and what attitudes and values of each individual member came into play as these were set up.

Chapter 6

Establishing evaluation criteria: reaching consensus on the notion of quality in German Educational Science

This chapter explores some of the problems which the experts in the commission encountered as they tried to establish working criteria for assessing the quality of the 15 HEIs they had been asked to examine. It does not describe each and every criterion, many of which can be identified from the questionnaires at the back of this study (where it is implicit where quality lies and can be identified). However, the following will reveal the particularly difficult areas that the experts had to address, and which, in some cases, were not completely resolved when the evaluation was concluded.

6.1 Working towards evaluation criteria

As already mentioned, the various members of the commission originated from the main of areas of specialism to be found in Educational Science in Germany. They were university-based and therefore largely influenced by the criteria under which they were accustomed to working in that particular environment. As Expert P stated: 'The classical educationalists were very strong. They led discussions and the others were somehow integrated.'[167] We also know that the remit of the teacher training colleges was quite different from that found in the universities, where the research tradition was much stronger. The archives record that there was a 'discussion about evaluation criteria', but are not very specific about the nature of this, although we know that it was this very area which posed perhaps the greatest challenge to the commission, as they were conscious that they were establishing the criteria according to which their discipline would be examined not only this time, but also in the future. As Expert Q told me:

> Some of the discussions in the beginning were quite difficult, […] because the group was looking to find the right criteria, what is the right framework and so on and of

[167] Ganz stark waren die klassischen EWler. Sie haben Diskussionen geleitet und die anderen wurden irgendwie integriert.

> course you have different opinions, but to them it was a slow process to make decisions. […] But the process was […] not very straightforward. It was going this way and that. And that is something that I have not experienced on other occasions because the criteria and the procedures were much clearer. And so it was a long process sometimes with a lot of discussions having different perspectives on the table and then you have to decide in the end.

Expert L highlighted how difficult it was to find quality criteria in arts subjects overall:

> And there's the problem in German arts subjects […] that they haven't developed any quality criteria for themselves. How they put their own work on the measuring bench. They do not have a relationship, no really good relationship towards their own teaching. They haven't got a relationship towards what they should be doing with publications. They don't give much consideration to their students' perspectives. So, all of these things are the depths into which one must not sink. One must ideally produce a host of exegetic papers, then you are given recognition in the social sciences. And that is particularly the case in pedagogy. The exegesis of something which was completed at some point is regarded as the non plus ultra of academic achievement. And I have my doubts whether that makes sense.[168]

So here we can identify not only an attempt at establishing quality criteria for one discipline in particular (Education), but extending beyond the limits of this discipline into the overarching field above it (*Überwelt*), so to speak. The last interview extract also raises the question of whether research in Education should be applied or not, which echoes back to some of the debate at the School of Education at the University of Chicago in the first few decades of the 20th century which discontinued Dewey's notion of using research to solve problems, rather than simply to be there for blue skies purposes (Pring, 2000). We see from what was said above that much research in German Educational Science in more recent times was critically interpretative, rather than conducted on the basis of empirical evidence. McCulloch & Richardson (2000) have postulated phases in the international evolution of educational research in which we can identify ascending and descending emphases (see also Denzin & Lincoln, 1994). In Germany at present the empiricists, who experienced a decline during the late 1970s, are now ascending to prominence once again,

[168] Und es gibt […] ein Problem in den deutschen Geisteswissenschaften – […] dass die überhaupt keine Kriterien für sich entwickelt haben. Wie sie ihre eigene Arbeit auf den Prüfstand stellen. Die haben kein Verhältnis, kein wirklich gutes Verhältnis zu ihrer Lehre. Sie haben kein Verhältnis dazu, wie das mit den Publikationen läuft. Sie machen sich wenig Gedanken über die Perspektiven ihrer Studenten. Also, all diese Dinge sind Niederungen, in die man sich nicht herein begibt. Man muss möglich viele exegetische Schriften erzeugen, dann ist man angesehen in den Geisteswissenschaften. Und das gilt in der Pädagogik ganz besonders. Da ist die Exegese von irgendetwas, was mal gemacht worden ist, immer noch das Non-plus-ultra der akademischen Leistung. Und da habe ich so meine Zweifel, ob das sinnvoll ist.

having occupied a prolific position in the last decades of the 19th century and the first three of the 20th.

As the remit of this commission was to examine two different institutional types and also the quality of research, teaching, learning and resources within them, the spectrum of the evaluation was very broad. Evaluations usually tend to focus on one or the other area (research or teaching), but seldom both together. As the commissioning ministry also had its specific interest in the empirical research being conducted within Educational Science, the experts needed to clarify their precise remit from the outset, which they attempted to do by restricting the extent to which they could judge the quality of teaching, as they identified that this was well beyond what was possible in the given time. It is a common phenomenon that clients, in this case the Ministry, have ambitious and sometimes over-enthusiastically long wish-lists of 'to-dos', which an evaluation commission cannot achieve, and so the list must be reduced to a manageable size (Owen & Rogers, 1999). Frau Bolius admitted that the evaluation did 'shoot in all directions' (*Rundumschlag*), but that the objective underlying it was to encourage the HEIs to examine their own quality in comparison to the status quo elsewhere in order to make decisions about which directions to pursue in the future.

When I spoke to one of the institutional heads, he described the situation within this field, which further illustrates some of the problems the commission encountered:

> Interdisciplinarity is the main commandment in teacher training. And so people have to work together. […] There is a deep chasm between pedagogy and Educational Science on the one hand, and subject didactics and the subject experts on the other. The reason for this is that their backgrounds are entirely different. Those doing subject didactics define themselves according to their subject, not according to the pedagogic-didactic part. Didactics is what many first learn when they are in their profession.[169]

[169] In der Lehrerbildung ist Interdisziplinarität ein oberstes Gebot. Also, man muss zusammen arbeiten. […] Es gibt einen ganz tiefen Graben, zwischen Pädagogik, Erziehungswissenschaft auf der einen Seite und den Fachdidaktiken und Fachwissenschaften auf der anderen. Der Grund dafür ist, dass die Herkunftswelten der Personen völlig verschieden sind. Die Fachdidaktiker definieren sich über ihr Fach, nicht über den pädagogisch-didaktischen Anteil. Die Didaktik ist dasjenige, was viele selber erst in ihrer Profession mit gelernt haben.

These comments give us a notion of different self-perceptions and understandings of quality and show that for some, at least, perceptions of quality were based on what they were doing in practice rather than anything dependent on published work or research.

Before we examine more carefully what criteria were adopted by the commission, we should briefly reconsider the framework within which quality in higher education has now come to be usually measured.

6.1.1 Frameworks for quality

One of the most visible signs of quality of research in an academic department is the number and quality of publications being produced in peer-reviewed journals as this reveals much about the extent of engagement of an individual or department. In the UK, for example, departments are often asked to present the four best publications per member of staff over a period of a few years. These should ideally have appeared in internationally recognised and peer-reviewed journals. Some disciplines (notably the natural sciences) expand their examination of publications to include their impact factor, as a result of which citation indices are examined. Furthermore, it is usual to identify how much third-party funding has been acquired by individual members of staff or departments as a whole, and in Germany those projects which are financed by the DFG are regarded as being of especially high quality. Third, staff-student ratios are examined as this can have a major impact on the quality of teaching. German HEIs generally suffer from staff shortages, with large student numbers progressing through their studies with little direct contact with staff members. Indeed, many students only have immediate contact with professors once they have progressed beyond the first stage of their degree programmes, e.g. past the *Grundstudium* and into the *Hauptstudium,* where more specialist knowledge is provided. Fourth, it is usual to examine the resources available; therefore library collections and numbers of books available to students will be scrutinised. Fifth, research projects are examined for relevance and impact. Sixth, interdisciplinarity is regarded as increasingly desirable, which we see predominantly in the natural sciences, where colleagues from

different disciplines frequently form collaborations. It is now expected that faculties and departments should reach out beyond their own borders to seek collaboration with other HEIs or individuals elsewhere, although it is not necessarily the tradition of the social sciences to do this (Henkel, 2000). Seventh, quality review examines how young academics are being supported, especially those taking doctorates and the extent of their engagement in research which is conducted within the department. Eighth, statistics on pass rates, numbers of students, male/female ratios, staff-student ratios etc. are gathered to present an overview of the institution and its capabilities. This is how one of the foreign experts described what they thought the commission should be looking for:

> And for the research of course it's also the criteria of what is the input in terms of quantity, but especially what is the quality of the work done here, the quality in terms of the criteria that you usually use to evaluate research reports, and also is there a kind of research programme in this institution or is everybody just doing his own thing? And so is there a kind of system here? Or are decisions made on how the research programme is developed and is executed and is the input, if you look at the output, in what journals is this published? What is the value of this work, and so on? (Expert Q)

We know that many members of our commission were already experienced evaluators who were familiar with many of the above criteria to assess HEI quality. Now, however, the situation in Baden-Württemberg required them to establish a framework that would be appropriate to the special situation in that state. The following comment describes the situation at the outset when the commission members became aware that their evaluation criteria needed to be suitable to both the universities and the teacher training colleges:

> For example, the question which I now in hindsight regard as a very important question was: can we put all of the things which we apply in the year 2003 as a necessary and important standard for HEIs and universities, internationalisation, peer-review, all these things, can we apply these to institutions which have so far had a different self-concept? By this I mean: can we measure them according to something against which they weren't measuring themselves, as we had adjusted them to different assumptions and given them a different job? Isn't that unfair?[170] (Expert S)

[170] Die Frage war zum Beispiel kann man das, was jetzt vor allem aus meiner rückblickenden Sicht eine wichtige Frage war, kann man das, was man jetzt im Jahre 2003 als notwendigen und wichtigen Standard an Hochschulen und Universitäten anlegt, eben Internationalität, Peer Review, all diese Dinge, kann man das an Einrichtungen anlegen, die in ihrer Selbstverständnis bisher anders waren. Das heißt, darf man sie an etwas bewerten, woran sie eigentlich nicht zu bewerten sind, weil man sie unter anderen Voraussetzungen eingestellt hat und ihnen auch einen anderen Auftrag erteilt hat. Ist das nicht unfair?

The notion of fairness underpinned much of their thinking as they did not wish to 'punish' the teacher training colleges for having a different self-concept (less research-based), although the commission held firm that more might be done to encourage them to work in a similar manner to their university colleagues. The next comment also shows the dilemma of applying well-established evaluation criteria to a discipline which possibly evades the norm:

> What is it like if, in this evaluation of Educational Science, we apply the typical criteria which are valid in other evaluations in a similar way? Which roles do different audiences play for publications? How far can we apply particular typical standards, or do we perhaps need to expand them, or replace them? If we look at the different aspects, how do we take account of the peculiarity of supporting young academics in a field where people as their teachers have a different career from those who stepped in via a diploma course?[171] (Expert G)

It was a major challenge therefore to try to establish criteria and apply these to 15 different HEIs. Expert K made the point that ultimately it would be impossible to make a general statement, as there was so much diversity in what was being presented:

> The situation in Baden-Württemberg is so difficult that we cannot formulate a recommendation which is basically relevant to all, the big message, because it is a different situation everywhere. And so we need an incredibly differentiated picture and need to reflect that according to each individual location.[172]

We should remember here that the evaluation criteria were established *after* the HEIs had already been asked to write their self reports. Expert F spoke to me about their feelings regarding this somewhat unusual state of affairs:

> Something which worked very well is that the criteria which we used were strengthened as the process continued and converged, which gave the process more clarity. I think that

[171] Wie ist es überhaupt, wenn man die Evaluation im Bereich Erziehungswissenschaften anlegt, sind typische Kriterien, die bei anderen Evaluationen gelten, gleichermaßen anzuwenden? Welche Rollen spielen unter-schiedliche Zielgruppen für Publikationen? Inwieweit kann man bestimmte typische Standards hier anwenden, oder man muss sie vielleicht ergänzen, oder ersetzen? Wie sieht es mit, wenn man die verschiedenen Aspekte durchgeht, der Besonderheit Rechnung zu tragen, Nachwuchsförderung in einem Feld, wo die Leute als Lehrkräfte eine andere akademische Karriere haben, als Leute, die über Diplom-studium eingetreten sind?

[172] […] die Verhältnisse in Baden-Württemberg sind so schwierig, dass wir eigentlich eine Empfehlung, die auf die, sozusagen grundlegend für alle gilt, sozusagen, die große Botschaft, die können wir gar nicht formulieren, weil es überall anders ist. Wir brauchen deshalb ein unheimlich differenziertes Bild und müssen das Standort für Standort getrennt ausweisen.

> worked very well. Although, and I'll reiterate it here, I didn't agree with everything. But I think that we managed the final result quite well. To begin with I would have preferred, for example, that the documents we received from the HEIs, that when we asked for these documents that we had been able to provide them with the criteria. That would have been a fairer process. […] it's clear that the criteria change throughout the process because they have to suit what is actually taking place. But with regard to the way it worked, as I saw it I don't think that it was clear enough from the outset.[173]

This comment is corroborated by what one of the institutional heads I interviewed told me; he described the problems his institution had with the process:

> Perhaps what the evaluation commission expected of us just wasn't clear enough from the start. While we were working on the self report we often thought, well, what do they actually want? What we need to discuss in this self report is so unclear to us. There was possibly a real deficit in the precise operationalisation of the concept of the evaluation commission. So we sometimes really asked ourselves: what do they actually want in this self report? What should we take into account and discuss?[174]

Essentially, therefore, although consensus may be reached in some areas, evaluation criteria need constant reworking and experts need an acute flexibility and broadness of vision to allow for idiosyncracies and variances in what they are evaluating, especially in this case where it was apparent that the 15 HEIs were, in some cases, markedly different (see also Stake, 1975, Guba & Lincoln, 1989).

In terms of establishing a good code of evaluation practice themselves, the commission decided to follow guidelines already set down by the WR, which provided the basis of what HEIs should report on. When the experts had compiled their own reports on the HEIs these were returned for review and editing. No HEI could later claim that they had been

[173] Was dann wiederum sehr gut gegangen ist, ist, dass wir die Kriterien, die wir angewendet haben, im Verlaufe des Prozesses ausgeschärft haben und aufeinander zubewegt haben und im Verlaufe des Prozesses dann noch Transparenz hergestellt haben. Das ging dann, glaube ich, ganz gut. Wenngleich - das sag ich nochmal - ich nicht mit allem einverstanden war. Aber von der Endsituation her haben wir das, glaube ich, ganz gut hingekriegt. Am Anfang hätte ich mir gewünscht, man hätte zum Beispiel die Dokumente, die man aus den Hochschulen erwartet hat, man hätte gleich mit der Anforderung dieser Dokumente die Kriterien bekannt geben können. Das wäre das fairere Verfahren gewesen. […] es ist schon klar, dass sich im Laufe eines solchen Prozesses die Kriterien immer noch mal ändern, weil die sich an der Praxis erst mal bewähren müssen. Aber so wie es gelaufen ist, war am Anfang aus meiner Sicht die Transparenz nicht hoch genug.

[174] Vielleicht waren einfach dann die Vorgaben auch seitens der Evaluationskommission nicht deutlich genug. Wir haben schon im Zuge des Selbstberichtes oft überlegt, ja, was wollen die jetzt eigentlich? Die ist so unklar jetzt für uns, was wir, worauf wir jetzt Stellung nehmen sollen in diesem Selbstreport, da hat es vielleicht auch an der wirklich genauen Operationalisierung der Vorstellungen der Evaluationskommission gefehlt. Und wir haben manchmal also wirklich, was wollen die jetzt eigentlich im Selbstbericht, wozu sollen wir jetzt Stellung nehmen?

falsely portrayed, or, in the words of one Expert N, cause a situation in which the commission 'ended up in trouble'.

6.1.2 Considerations for 'first-time' commissions

The following expert described to me one of the fundamental problems that the commission had with the time-frame in which the evaluation was conducted:

> What I under all circumstances would do differently in the future, and is something I would have liked to have seen being done differently at the time, was that we might have had more preparation time. As a commission. A phase in which we as a commission discussed the criteria. Everything in this process started at the same time, namely getting information from the HEIs, setting up the commission, the whole organisation of members amongst each other and so on. I didn't think that was very successful, simply because it didn't leave enough space for particular discussions. And if we had had more preparation time, then I think it would have been clearer what we all expected of each other.[175] (Expert F)

Although one might argue that each evaluation is fundamentally a unique undertaking, this seems particularly the case where so-called 'first time' evaluations are conducted. Where an evaluation examines a discipline in its entirety for the first time, it should be a pre-requisite of the commissioning body to allow the members of the evaluation enough time and space in which to establish clearly the exact evaluation criteria that are to be applied *before* engaging the HEIs in any preparation of self reports. Although we know that during this commission the criteria were refined and adjusted as work progressed and therefore made more suitable to the particular context within Baden-Württemberg, it was perhaps somewhat unfortunate for the HEIs involved that these criteria were not available from the outset.

[175] Was ich auf jeden Fall anders machen würde zukünftig und was ich damals auch gerne anders gesehen hätte, wäre, dass wir eine Vorbereitungsphase gehabt hätten. Also als Kommission. Eine Phase, in der wir als Kommission uns verständigt hätten über Kriterien. Die Gleichzeitigkeit, mit der dieser Prozess in Gang kam, nämlich die Information von den Hochschulen einholen, die Kommission etablieren, die ganze Organisation der Mitglieder untereinander und so weiter, die find ich nicht glücklich. Weil sie einfach zu bestimmten, zu bestimmten Diskursen nicht Raum lässt. Und wenn wir etwas Vorbereitungszeit gehabt hätten, dann wäre, hätte das, glaube ich, klarer sein können, was wir voneinander erwartet haben.

6.1.3 Camps and hierarchies: consequences for establishing evaluation criteria

Apart from a couple of HEIs where there already was a concentration on empirical research, the majority of HEIs (especially the teacher training colleges) were more hermeneutically-oriented, which was generally regarded as problematic. As Expert B put it: 'All institutions in which empirical research was less apparent came under critical fire.'[176] Reflecting the status quo of Educational Science overall, there were two main 'camps' within the commission membership: the empiricists and hermeneuticists (cf. Denzin & Lincoln, 1994), and these ranged across the different subject areas, which reminds us of the tribalism within academia to which Becher & Trowler **(**2001) have referred.

To some extent the hermeneuticists in the commission were now cast into an inferior position, as they knew the Ministry was especially interested in empirical research, and so they had to work particularly hard to uphold their own interests. Expert K gave me some of the background on these two directions and their influence within this discipline in Germany:

> The empirical preoccupation with education migrated into medicine to pediatrics or into criminology, criminal youth, etc. But in Germany it was clear that pedagogy can only work philosophically, only then is it science. The rest is empiricism, positivism, which we don't need. And that so to speak is how the Arts direction crept in. That was intended from the start, and the universities also said that they will accept pedagogy as science only if it is philosophical.[177]

This reflects the strong tradition of philosophy within German Educational Science, occupying for so long a prestigious status, not only locally, but internationally. Hermeneutics was therefore originally the dominant force within Educational Science, which did not need to be based on 'hard' evidence, as it were, as it was more conjectural

[176] […] alle Institutionen bei denen empirische Forschung weniger ausgeprägt ist standen im Feuer der Kritik.
[177] Und die empirische Beschäftigung mit Erziehung, die wanderte aus in die Medizin zur Pädiatrie oder in die Kriminologie, also verbrecherische Jugendliche usw. [A]ber die deutsche, das war klar, die Pädagogik darf nur philosophierend sein, nur dann ist sie Wissenschaft. Das andere ist Empirie, Positivismus, brauchen wir nicht. Und so kam sozusagen die geisteswissenschaftliche Orientierung da rein. [D]as war von Anfang an so gewollt, die Universitäten haben auch gesagt, wir nehmen Pädagogik als Wissenschaft nur als philosophische.

and philosophical. However, as the PISA study had shown such deficits within German school education, the discipline as a whole could no longer continue in this vein, as hard evidence was sought for what was going on in the classroom and elsewhere. The empiricists now came to the fore and research of this type was regarded as what was needed. This new emphasis was suitable to some members of the commission who were already working empirically, but for those who were not, it was more problematic to see the emphasis shift away from what they had hitherto regarded as the acceptable form of research, and hence of quality. As Expert P said when referring to the new direction: 'I am an Arts person through and through, and so I've got my problems with that.'[178] The reverse was true from those in the empirical camp:

> Psychologists who were working empirically, empirically working Educational Scientists very probably felt much more at ease than those who were historically-oriented, which also includes the special education and vocational educationalists, who also had their problems. They simply had a status problem, as their areas are so small […]. But one certainly sensed it and this research school of thought. For some disciplines it isn't half as important as for others, such as those Educational Science areas which work empirically, that's clear. That's my school of thought and so I felt very comfortable with that. But I can imagine that a historically-oriented educationalist may occasionally have had his problems. Those discourses were tangible.[179] (Expert E)

It was this dichotomy which made me ask the experts whether they believed that everyone's voice was on equal footing within the commission, or was there a risk that in this emphasis on empirical work other voices might be lost or weakened?

The overall response from those in a position of strength (e.g. from within the empirical camp and from within the more mainstream areas) was that the experts did have an equal voice within the commission. Expert R disagreed, however, telling me that they thought the real decision-makers were limited to the voices of five people only:

---

[178] Ich bin durch und durch Geisteswissenschaftler, darum habe ich meine Schwierigkeiten damit.

[179] Empirisch arbeitende Psychologen, empirische arbeitende Erziehungswissenschaftler haben sich da sehr wahrscheinlich wohler gefühlt, als historisch orientierte, sicher auch als Sonderpädagogen und die Berufspädagogen, die hatten da auch mal ein Problem. Die hatten einfach auch ein Statusproblem, weil sie so klein sind […]. Aber das spürte man sicher, auch diese Forschungsorientierung. Das ist ja für gewisse Disziplinen lange nicht so wichtig, wie für eine empirisch arbeitende Erziehungswissenschaft, das ist ganz klar. Das ist meine Richtung, ich fühlte mich da wunderbar wohl. Aber ich kann mir vorstellen, dass ein historisch orientierter Erziehungswissenschaftler oder Pädagoge da manchmal seine Probleme hat. Diese Diskurse waren ja dann auch spürbar.

> These are the five people who determined more or less the climate and the direction. And then the others modified it or some such. But these did, I think, generally determine the path which the commission took, more or less.[180]

Those who were in less strong positions confirmed this opinion. One person stated that they felt they had been brought in as part of the display (of experts), but that they were there only to 'look good', but not serve a real purpose. They were there to 'window-dress': 'My objections were greater or less, which sort of belonged to the game. But they weren't to be taken seriously.'[181] (Expert L) There was a difference therefore in the role of the individual. Those whose knowledge was very specific (e.g. special education, vocational education) would, to some extent, be less likely to have an impact on the decision-making processes than those whose knowledge was more mainstream. These sub-groups faced an additional danger when the evaluation criteria were being established:

> And in terms of group dynamics you know the phenomenon, too. Whoever has started off not speaking in plenary meetings will only get back into the game with great difficulty. If someone has long remained silent during a meeting and was only nice during the breaks they can't expect to say in the third meeting: I don't like any of this. That isn't OK. So these are observable group dynamics, occurrences you can see and processes which come into practice, roles, to which you are allocated.[182] (Expert J)

There is a danger, then, that unless the experts are aware of these unwritten rules, their voices will be drowned out at a crucially important phase of an evaluation in which the decisions on criteria, for example, are being made. This also reminds me of one of the problems that foreign experts have when working in commissions. Due to courtesy towards their host nation, they are less likely to be vociferous in the early stages of an evaluation. In this example, therefore, it seems that the empiricists from within the larger

[180] Das sind so fünf Personen, die mehr oder weniger das Klima und die Tendenz mitbestimmt haben. Und dann die anderen, die haben da etwas modifiziert oder so etwas. Aber diese haben doch, glaube ich, im Großen und Ganzen […] den Weg der Kommission im Großen und Ganzen bestimmt.

[181] Und auch meine Einwände waren dann mehr oder weniger. Das gehörte sozusagen zum Spiel. Aber nicht ernsthaft in Erwägung zu ziehen

[182] Und gruppendynamisch gesehen kennen Sie das Phänomen ja auch, wenn man erst einmal angefangen hat in Plenarsitzungen in einer Gruppe nichts zu sagen, dann kommt man ganz schwer wieder ins Spiel. Wenn man eine Sitzung lang geschwiegen hat und nur in der Pause nett war, kann man nicht in der dritten Sitzung auf einmal sagen, das gefällt mir hier aber alles nicht. Das geht nicht. […] insofern sind das auch gruppendynamische Ereignisse, die man da beobachten kann und Prozesse, die sich einspielen, Rollen, die einem zugeschrieben werden.

areas of German Educational Science were more likely to be the main decision-makers with regard to establishing evaluation criteria than those on the periphery of the field.

In her observation of the group dynamics of this evaluation Bolius stated that there was a general reluctance among professors to speak about areas they were not totally familiar with, identifying a cultural difference that might be specific to German professors only:

> The more competent they are in their subjects, the more afraid many professors are to discuss anything which is even only a tiny amount beyond the borders of their area. […] In the US, I think, it is far more common that professors will give their opinion on anything and everything.[183]

This is a very interesting point in reference to the ability of an evaluator of higher education from within a tradition of non-expression beyond his/her area. Evaluation requires voice, and it requires opinion-making and surely polemical debate to some extent. However, as Bolius mentioned further, there was a general reluctance to polemicise discussion:

> I'd say that everything is very careful. There's no-one who says that it's all a waste of time, or something like that. It's rather a bit like: "Yes, well…" and "Very good, but… perhaps here it could…" So, I think there's rather a certain fear of being polemical."[184]

Let us now turn from the stance of individuals within the commission to examine how the Dutch evaluation system influenced their thinking.

6.1.4 Influences from The Netherlands

If evaluations are supported by quantitative objective data such as publications in peer-reviewed journals, then they are usually regarded as robust and rigorous. When the

[183] Viele Professoren […] haben irgendwie je kompetenter sie in ihrem Fach sind, desto mehr Angst haben sie, sich eigentlich aber auch nur ein Stück außerhalb von ihrem Fach zu äußern. […] In Amerika, denke ich, ist es viel mehr üblich, dass sich Professoren zu allem und jedem äußern.

[184] Ich würde eher sagen, dass alles immer sehr vorsichtig ist. Es gibt eigentlich eher niemand, der sagt, das ist doch hier alles absolut nicht zu gebrauchen, oder so. Sondern es ist alles eher so ein bisschen „Ja, aber", und „Ganz gut, aber… Vielleicht hier könnte…". Also es ist eher, denke ich, so eine gewisse Scheu zu polarisieren da."

evaluation commission started work to find their evaluation criteria, they turned to the Dutch system of evaluation. Drawing on Vroeijenstijn's (1994) list, Van Berkel & Wolfhagen (2002, p.341) summarise some of the areas that should be assessed in terms of teaching and learning as follows:

1. Study program: aims, characteristics, contents

Aims and objectives

Program characteristics

Level of examinations and methods of student assessment

Students' skills

Level of final thesis/final projects

2. The student and his/her education

Student numbers, enrolment and total numbers

Dropouts and completion rates

Average duration of study

Study load

Organisation of a program

Student counselling

Facilities and infrastructure

3. The graduates

Level of the graduates' knowledge and skills

Labor market perspectives

4. Academic staff

Staffing

Qualification of academic staff

Students/staff ratio

Graduate/staff ratio

Personnel management

5. Internationalization

Participation in ERASMUS, LEONARDO DA VINCI, etc.

International contacts

6. Internal quality assurance

Self-assessment

What was done with the outcomes of a previous assessment?

Evaluation process

Connections with alumni

While most of these seemed straightforward and easy to determine, others were not quite so simple. Kieser (1998) had already identified a number of deficits with the Dutch system. For example, he argued that defining research output according to the criteria of quality, productivity, relevance and viability, as suggested, was problematic. Relevance, to quote just one criterion, could be interpreted in many different ways and it would depend on the viewpoint of the expert in question. However, although the commission may have been aware of Kieser's criticisms, the Dutch system remained the predominant point of reference, certainly in terms of its overall process. Nonetheless, the Dutch criteria still needed re-thinking in the specific context of Education as this comment illustrates:

> How far can one simply transpose this Dutch method? Where does it fit? Where should it be expanded? In which manner? And how can one try to present the whole thing? How does one deal onsite with such categories, how do we judge them together when onsite and so forth? Well, there were many many points which really needed clearing up in the meetings before our onsite visits.[185] (Expert G)

Expert N tells us about what was required in terms of expanding the Dutch system to suit the present context:

> Practice was evaluated somewhat more. And the teaching part was new. We had a model from The Netherlands, but it wasn't so well suited. [Name of expert] played a relatively major part in developing the teaching part, the criteria for that. And they stood the test quite well. I was quite surprised how well it worked.[186]

The group did not have any difficulty implementing the Dutch model in terms of evaluating research. However, they developed an idea that individual criteria could be classified into a kind of rating scale, a notion alien to most of the experts, some of whom were therefore reluctant to use this method. In the end, though, the idea was adopted. This comment tells us why this was the case: 'They adopted it under the notion that this [the

[185] Aber inwieweit [..] kann man dieses niederländische Verfahren einfach so übertragen? An welchen Stellen passt es? An welchen Stellen muss man ausweiten? In welcher Art und Weise versucht man, das Ganze abzubilden? Wie geht man vor Ort mit solchen Kategorien um, bewertet man die jetzt gemeinsam gleich vor Ort und so weiter. Na ja, viele, viele Punkte, die in diesen Vorfeldsitzungen eigentlich auch zu klären waren.
[186] Die Praxis wurde ein bisschen mehr begutachtet. Und neu war der Lehrteil. Den haben wir ... da gab es ja auch […] Vorbild aus den Niederlanden, aber das war nicht so gut geeignet. Und da hat […] eine relativ große Rolle mitgespielt bei der Entwicklung des Lehrteils, also der Kriterien. Die haben sich eigentlich gut bewährt. Also ich war ganz verblüfft, wie gut [es] funktioniert hat.

evaluation] was a developmental process and that we tell the HEIs: 'We know that you weren't employed to work according to these criteria, but we also know that this is the direction in which the system will be moving.'[187] (Expert N) However, those with reservations felt that a rank did not necessarily reflect the quality of the so-called product. This expert thought that a more individual approach would do greater justice to the area being evaluated and that a broad comparative sweep should be avoided: 'What I think is much better is to do a strengths/weaknesses analysis and, working with the university, consider what can be done to improve the strengths and limit the weaknesses.'[188] (Expert H) We therefore see an evaluation commission which was challenged to use quantitative and so-called 'objective' criteria, but which was restricted in its possibilities by the very discipline under inspection and the lack of homogeneity among the institutions present.

A further idea which was overturned was that it was impossible for an individual professor to maintain high quality without being part of a larger team or collaboration with members of an HEI elsewhere, e.g. there should always be collaborations or inter-disciplinary connections, or at very least research groups within a department. One of the more hermeneutically-minded experts told me that he was strictly against the research group model, which was more commonly found in external research institutions. He said that within Educational Science this notion of quality was less suitable: 'We got into a lot of trouble in the discussions and I made a conscious effort to raise this issue and I thought it was important to do so.'[189] (Expert J) This again reflects on the notion of the strength of individual voices within the commission, which the expert in this particular example had the courage to express.

[187] Und zwar mit der Perspektive, dass dies ein Entwicklungsprozess ist, dass man denen sagt, den Hochschulen sagt, "wir wissen, dass ihr nicht unter diesen Bedingungen berufen worden seid, aber wir wissen auch, dass das die Entwicklungsrichtung des Systems sein wird.

[188] Was ich viel besser finde, ist, dass man eine Stärken-/ Schwächenanalyse macht und dass man, und das würde ich immer erwarten von so einem Auftraggeber, dass man überlegt und auch von der Universität, was man tun kann, um die Stärken zu verbessern und die Schwächen zu relativieren.

[189] Und da haben wir viel Ärger drüber gehabt in den Debatten und die habe ich auch bewusst angezettelt und da habe ich auch Wert drauf gelegt.

6.1.5 Publications and impact factor

The commission's overall outlook was dominated by a perspective on quality in higher education that was predominantly university-based, a slightly different culture therefore to that which was found in the teacher training colleges. It was also apparent that some criteria that were frequently used in other disciplines were highly unsuitable in this context. The impact factor was one of these, as some of the research being conducted was specific to very small groups and not applicable in broader educational contexts: 'Quality, teaching quality, but also research quality, can't always be determined by the impact factor. That's one difficulty.'[190] (Expert M) Another expert told me: 'We didn't have an impact [factor]. Well, citations weren't possible, because Site Citation only covers two or three German journals and so that doesn't make sense.'[191] (Expert N) The commission discussed which journals should be included on their list of those to which the discipline should be aspiring. We remember that Jürgen Baumert put considerable effort into the compilation of this list, which, as far as they were aware, had hitherto not existed for the discipline in Germany. The commission also discussed the notion of publishing in English-speaking journals; however, this was not without its problems as much educational research in Germany is specific to local interests only, and so authors might not strive to publish their findings internationally. One of the foreign experts suggested that the HEIs should use a system similar to that in his home country where a few key publications should be presented. However, as the following comment illustrates, in Educational Science it was difficult to establish which publications should be included and the criteria according to which they should be selected:

> Well, I'll give you an example of something which I still find implausible, namely that we're still using the same criteria to measure publications in Educational Science as those we use in psychology. To my mind that isn't appropriate for Educational Science. From my perspective it's inappropriate to have the notion that we should publish lots in international journals, namely in English. Firstly, I find that German as a scientific language suits me, and secondly, it has a particular level of argumentative depth. My English is fairly good and I can communicate it well, but there is a subtle depth of

[190] Qualität lässt sich ja, Lehrqualität, aber auch Forschungsqualität nicht immer bestimmen, an dem Impact Faktor. Das ist schon die Schwierigkeit.
[191] Impact hatten wir nicht. Also, Zitationen waren nicht möglich, weil der Site Zitation erfasst nur zwei oder drei deutsche Zeitschriften. Das macht keinen Sinn.

> argumentation which I cannot achieve in English. And so not only would I still like to publish in German and be well respected, but that the same should apply to young academics. We didn't reach agreement on this point.[192] (Expert F)

To return to Kieser's points (1999b) about those international publications which should be aimed for, he tells us that the majority are edited by Americans. These may, for example, have a different outlook on a discipline than that held in continental Europe. High quality publications in German Educational Science are therefore not necessarily suitable for placement in US journals. We can empathise with the above expert who wished their publications to be accepted as being of high quality, even if they appeared within the German market only.

Although Educational Science as a discipline finds its roots in philosophy and psychology which have their own clear and established references for good quality research, much of the research that is done today in Education cannot necessarily be categorised as neatly as in some of the more clear-cut disciplines where it is possibly easier to establish guidelines for good quality. This again impacts on the kind of publication in which the research is disseminated. The next comment illustrates the problem further: 'To what extent should we regard special contributions, which were written for teacher journals, as moving science forwards or not? Are they full of quality or not?'[193] (Expert G) This next comment is even more explicit about changing assumptions on quality within the group:

> What does it mean when a [school] subject specialist writes a book? Is this an accomplishment, or is it an escape from work? To begin with we were all sceptical until a subject specialist within the commission told us on the first day when the problem was raised how difficult it is to write a textbook for schools, how creative you have to be, how

[192] Also, ich sage Ihnen mal ein Beispiel, was für mich nach wie vor nicht plausibel ist, ist, dass wir bei der Beurteilung von Publikationen innerhalb der Disziplin Erziehungswissenschaft die gleichen Maßstäbe anwenden wie in der Psychologie. Das ist, finde ich, für die Erziehungswissenschaften nicht angemessen. Es gibt... aus meiner Sicht hängt [...] die Unangemessenheit hängt insbesondere an der Vorstellung, dass man möglichst viel in internationalen Zeitschriften publizieren muss, sprich auf Englisch publizieren muss. Ich finde, das Deutsche als Wissenschaftssprache, erstens liegt mir was daran, zweitens gibt es ganz bestimmte... eine Argumentationstiefe. Ich kann ganz gut englisch und ich verständige mich auch gut, aber es gibt eine Argumentationstiefe in der Subtilität, die kann ich im Englischen nicht erreichen. Ich möchte also weiterhin ... nicht nur *ich* möchte, dass *ich* auf Deutsch publizieren kann und das trotzdem angesehen ist, sondern ich möchte zum Beispiel auch, dass das für einen Nachwuchswissenschaftler gilt. An der Stelle sind wir eben nicht im Konsens geblieben, gewesen.

[193] Inwieweit betrachtet man bestimmte Beiträge, die verfasst werden für Lehrerzeitschriften, als wissenschaftlich weiterführend oder nicht weiterführend, oder als qualitäts- oder als nicht qualitätsvoll?

> you need to break it down into learning levels, how you need to create the course concept which should be maintained for the duration of an entire school life. And this isn't research in the normal sense of the word, but it is an incredible accomplishment, relating to the construction of knowledge in applied contexts. And we all grew more understanding at that point. […] And so in this case we sometimes started asking: so why aren't you writing any textbooks?[194] (Expert J)

This is where we need to ask the question as to the relevance of 'normal' publication criteria, as we have just seen from the above example that some of the experts in the commission, at least, needed to revise their understanding of what quality within Educational Science was and that this did not necessarily appear in the usual context within peer-reviewed journals, but that publications of high quality and impact could be appearing in entirely different formats that should be judged accordingly, but were difficult to incorporate into standardised ranked lists, for example.

However, as one of the ministerial members indicated, by establishing quality criteria it can become the case that a discipline moves towards a given standard, even if it takes time to do so, but this can also have negative consequences:

> If I establish a criterion, for example, 'Publications in Internationally Peer-Reviewed Journals' and I uphold this exclusively and for a number of years, then the whole scientific community will adjust itself to that accordingly, which means they will try to *only* publish in these journals, will stop writing any books, I exaggerate here, will stop doing projects which aren't suitable. They will adjust their behaviour and that can again lead to a restriction, a negative effect. And so, […] although this remains an accepted criterion, it always bears a certain inherent danger.[195]

[194] Was bedeutet es, wenn ein Fachdidaktiker ein Schulbuch schreibt? Ist das eine Leistung, oder ist das Flucht vor Arbeit? Am Anfang haben wir alle skeptisch geguckt, bis uns dann die Fachdidaktiker in der Kommission, am ersten Abend, als das Problem auftauchte und das war gleich am ersten Tag, dann mal erzählten, was das für ein Mühsal ist, ein Schulbuch zu schreiben. Was man an Aufgabenerfindung haben muss, wie man auf Lernniveaus runter brechen können muss, wie der einen Lehrgang konzipieren muss, und da über ein ganzes Schulleben hinhält. Und dass das vielleicht zwar nicht Forschung in dem Sinne ist, aber eine ungeheure Leistung, bezogen auf Konstruktion von Wissen im Verwendungskontexten. Und da sind wir auch alle verständiger geworden dabei. […] Und wir haben dann hier eher manchmal gefragt, warum schreibt ihr keine Schulbücher?

[195] Und wenn ich jetzt ein Kriterium aufstelle und es ist zum Beispiel `Publikationen in einem international referierten Fachjournal' und dieses Kriterium halte ich quasi ausschließlich und für lange Jahre aufrecht, dann wird sich die ganze Wissenschaftslandschaft danach ausrichten. Das heißt, man wird versuchen, nur noch in diesen Journalen zu publizieren, keine- also, ich übertreibe jetzt- keine Bücher mehr zu schreiben, keine Projekte, die sich nicht eignen, oder sonst irgendetwas zu machen, der wird sein Verhalten darauf ausrichten und das kann auch wieder zu einer Begrenzung, oder zu einem negativen Effekt führen. Also, […] obwohl das natürlich immer noch ein anerkanntes Kriterium ist, aber es hat auch immer eine gewisse Gefahr in sich.

One expert described the process of readjusting expected standards of quality thus:

> We all agreed that we have to try to develop quality criteria which the discipline must learn to accept, […] but I am not convinced that we have already developed these criteria. Quite a lot more still needs doing. […] So, for myself, I decided to say, OK, this is a working situation for the time being from which we can move on. But it isn't anything with which I am conclusively happy.[196] (Expert F)

6.1.6 Research funding and notions of quality

In most disciplines in Germany one of the main quality criteria is to see how much DFG funding has been applied for and received. However, it initially seemed irrelevant to some of the experts in the commission to apply this criterion to some of the specialist areas within Educational Science where there was no tradition of acquiring funding from the DFG whatsoever. Some experts, therefore, made a point of ensuring that this criterion was not superimposed on to their areas as they thought it would make little sense in the present situation. Expert L expressed their concerns about the criteria that were established:

> I had the feeling that this commission worked according to, how should I say, very conventional university criteria. Very conventional criteria. And that means number of publications, of course only important publications […], the acquisition of third-party funding, research money, research projects. And in a subject such as pedagogy it seems to me, at least insofar as in relation to practice-oriented questions such as school scientific inquiries, the improvement of teacher training […], then the criteria seem to me to be unable to describe the worth or quality of that kind of training.[197]

In spite of reservations, the commission did ask for third-party funding to be listed, which was the first time that the 15 HEIs in Baden-Württemberg in the present study had been asked to list the source of funding for research projects and for many this was an entirely new and unexpected exercise:

---

[196] Also von daher war bei dieser Prämisse, waren wir uns alle einig, wir müssen versuchen Qualitätsmaßstäbe zu entwickeln, die die Disziplin sich dann auch gefallen lassen muss. Muss sich an diesen Maßstäben dann auch messen lassen, aber ich bin nicht davon überzeugt, dass wir diese Maßstäbe auch schon entwickelt haben. Da muss noch einiges geschehen. […] Also ich habe mich dann für mich selber so entschlossen, dass ich sage, okay, das ist jetzt mal ein Arbeitszustand, von dem aus man jetzt weitergehen muss. Aber da ist für mich nichts womit ich am Ende zufrieden bin.

[197] Ich hatte das Gefühl, dass diese Kommission sehr stark an sag ich mal konventionellen universitären Kriterien gearbeitet hat. Sehr konventionellen Kriterien. Und das heißt, Anzahl der Publikationen, natürlich nur der wichtigen Publikationen, die jetzt irgendwie, also, die begutachtet sind. Einwerben von Drittmitteln, Forschungsgeldern, Forschungsprojekte. Und bei einem Fachbereich wie Pädagogik scheint mir das, zumal wenn er sich auf praxisorientierte Fragestellungen bezieht, wie schulwissenschaftliche Untersuchungen [und] Verbesserung von Lehrerausbildung […] scheinen mir das Kriterien zu sein, die nicht der Güte, ja, nicht die Qualität des Ausbildungsganges beschreiben.

> For the HEIs it was a real challenge that they had to deliver third-party funding data per person, who had applied for it, how long the projects were, and that they had to list the funding bodies from which they had received money […]: it was surprising – they didn't really manage it. […] It wasn't a prejudice, but a pervasive and protectionist argument that they said the DFG doesn't want our research, and so we don't fit in and so we won't apply for any.[198] (Expert N)

By signalling that funding through the DFG was indeed one of the criteria that the commission was seeking and held to be an indicator of good quality, the experts were implicitly indicating to the HEIs that this was what they should be aspiring to in the future and that there were no excuses for not applying.

We know that some departments were engaged in research activities of a predominantly applied nature and very specific to local interests. They received small pockets of funding which belied the fact that their research may have had quite an impact locally. Referring to one institution which had produced some useful applied research, Expert L continued as follows:

> It's an achievement of this subject area undoubtedly. You can see that this subject area has managed to shape and change practice with staff members who are more or less well trained. I think that is a major accomplishment. And what can we find to praise in this subject area as it doesn't fit the other classic scientific criteria of determination of worth? There wasn't any DFG funding or some such. There weren't any refereed publications, none. If we only adhere to refereed publications, then we are in the classical academic mainstream. Scientists determine what science is and they determine it mutually.[199]

This last comment is reminiscent of Kieser's conclusion (1998, p.411): 'Evaluations create the reality they profess to measure. The evaluands create a system, which corresponds with

[198] Für die waren das schon eine richtige Herausforderung für die Hochschulen, also dass die Drittmittel abliefern mussten pro Person, wer die eingeworben hat, Laufdauer der Projekte, dass sie die Zuwendungsgeber angeben mussten, von wem sie das Geld hatten; […] das war schon überraschend - die haben das nicht hingekriegt. Es ist ein weiterverbreitetes, aber nicht ein Vorurteil, aber eine protektive Argumentation, die sagen, die DFG will unsere Forschung nicht, wir passen innerlich nicht rein, deshalb beantragen wir gar nichts.

[199] Das ist eine Leistung des Fachbereichs, das muss man ohne Zweifel sehen, dass es diesem Fachbereich gelungen ist, mit seinen Leuten, die auch mehr oder weniger gut ausgebildet sind, die Praxis zu gestalten und zu verändern. […] Das finde ich eine große Leistung. [Und] was finden wir denn an diesem Fachbereich, was wir irgendwie loben können, weil er ja alles, was die anderen wissenschaftlichen klassischen Kriterien der Statusbemessung angeht, nicht ausmacht? Also, das waren keine, die Deutsche Forschungsgemeinschafts-gelder oder sonstige, ja. Und das waren, die Publikationen waren nicht referiert, die ganzen Publikationen. Wenn man jetzt nur an referierten Publikationen festhält, dann ist man im klassischen akademischen Mainstream. Also, Wissenschaftler definieren, was Wissenschaft ist und sie definieren es gegenseitig.

the ideal picture of the evaluation's constructors'.[200] In this case, how does an evaluation commission give credence to something which it recognises as having 'good quality' (e.g. a well-written school textbook with a high impact factor), although it is placed on the lowest rank of what academics usually perceive good quality to be? This illustrates the problem of wishing to determine in set rules what a discipline should be trying to achieve. Quality assurance can be inclined to too much rigidity in its approach and disallow credit for good work that is happening outside of what is regarded as the norm. As Expert K said: 'Third-party funding for me isn't such a determining factor. The use of money isn't an indicator, or not in itself an indicator for good science.'[201] Money allocated for a research project might not necessarily lead to excellent research or high volume and quality output in respected journals (although it is to be expected). In direct contrast Expert J disagreed, saying that the only really hard indicator was the statistic on third-party funding, as all the other indicators which one can quantify are 'softer'.[202] However, he continued as follows:

> Third-party funding, which is strongly competitive and issued very selectively and with high quality criteria, such as those of the DFG, can be differentiated from others. And you can do that in a more reliable and simpler manner (than with publications). […] But the statistics don't reveal as much as one might think to begin with.[203]

While the statistics provided by the HEI might give a commission some insight into its quality, the commission should remain flexible in its interpretation of the figures provided. The above comments reveal that although the experts may have reached agreement in the final report, the debate on the issue was far from over and might continue into future evaluations.

Although the criteria which were finally agreed upon might not be regarded as definitive for the discipline, they nonetheless represented a step forwards for German Educational

[200] Evaluationen schaffen die Wirklichkeit, die zu messen sie vorgeben. Die Evaluierten schaffen ein System, das dem Idealbild der Konstrukteure der Evaluation entspricht.
[201] Drittmittel ist für mich nicht, ist auch ein Faktor, aber ist nicht so entscheidend. Der Verbrauch von Geld ist nicht ein Indikator, ist nicht allein ein Indikator für gute Wissenschaft.
[202] Der harte Indikator ist wirklich die Drittmittelstatistik, der ist wirklich hart. Während alle anderen Indikatoren, die man quantifiziert hat, eher weich sind.
[203] […] Drittmittel, die stark wettbewerbsorientiert sind und sehr selektiv vergeben werden und mit hohen Gütekriterien wie bei der DFG, von anderen unterscheiden kann. Und das kann man auch verlässlicher und einfacher tun. […] Aber die Statistiken geben nicht so viel her, wie man anfangs meint.

Science in general. Expert P told me: 'The evaluation was innovative, but they needed to take ideas on board.'[204] Referring to the criteria another said: 'Educational Science in Germany didn't have that up to now. We now have explicit criteria, or explicit dimensions.'[205] (Expert O) What the evaluation commission had achieved was to establish a framework of criteria which they felt the discipline should at least be aware of in the future, and start working towards if possible, even if this meant that the teacher training colleges in particular would have to redefine themselves to some extent:

> It was a tough framework for our universities and particularly for the teacher training colleges, as they had not defined their institution up to now according to international standards, according to publications in peer-reviewed journals, according to the quality and extent of third-party funding. Professors in the teacher training colleges haven't usually even done the *Habilitation*, which means that after their doctoral thesis […] they aren't required to do any more major scientific work. Well, they have obviously published, otherwise they wouldn't have become professors, but in the last 20 years, the older generation, which is currently leaving our institutions, wasn't very highly qualified. Instead, they defined themselves as teacher trainers with a strong relationship to practice. That is the job we now give academic assistants, the so-called *Mittelbau*. As I see it, this is the role which professors back then occupied for many decades.[206] (Expert L)

6.2 Summary

In this chapter we have seen the evaluator's journey towards the creation of evaluation criteria with which they felt they could judge both the teacher training colleges and the universities fairly and appropriately. We saw that this process was modified as work progressed, echoing Guba & Lincoln's (1981) notion of evaluation as a flexible instrument that has to change according to the needs of the stakeholders involved. In addition, we identified the different paradigms underlying Educational Science and the debate within the

[204] Die Evaluation war zur Innovation, aber die sollten Ideen mit an Bord nehmen.

[205] Aber es war überhaupt für die Erziehungswissenschaft ist es ein enormer Fortschritt. Es hat die Erziehungswissenschaft in Deutschland überhaupt nicht gehabt bis jetzt. Jetzt haben wir das: explizite Kriterien, oder explizite Dimensionen daraus gearbeitet.

[206] Das war dann schon für die hiesigen Universitäten oder gerade auch pädagogischen Hochschulen ein hartes Raster, weil sie unter diesen Gesichtspunkten bisher nie sozusagen ihre Hochschule definiert hatten nach internationalen Standards, nach Veröffentlichungen in peer-reviewed Zeitschriften, nach der Qualität und dem Ausmaß von Drittmitteleinwerbung. An den PHs sind die Professoren in der Regel nicht mal habilitiert, das heißt, sie haben nach der Doktorarbeit, das muss man sich wirklich vor Augen führen, keine größere wissenschaftliche Arbeit machen müssen. Gut, sie haben danach natürlich publiziert, sonst wären sie keine Professoren geworden, aber in den letzten 20 Jahren, diese ältere Generation, die dann jetzt gerade so die Hochschulen verlässt, die waren oft wissenschaftlich an den PHs nicht besonders hoch qualifiziert, sondern die haben sich als Lehrerbildner mit dem ganz wichtigen Bezug zur Praxis definiert. Das war die Aufgabe, die wir heute den wissenschaftlichen Mitarbeitern, sozusagen dem Mittelbau zuweisen. Diese Rolle haben praktisch die Professoren damals, so sehe ich das, über viele Jahrzehnte eingenommen.

commission to accept the dominance of empiricism, which, in turn, influenced the evaluation standards that were adopted.

Let us now turn to some of the organisational aspects of evaluation work, examining some of the problems that the commission faced when they went onsite, and the manner in which they resolved these.

Chapter 7

Organisational issues and problems: self-concept of evaluators and learning processes

Establishing evaluation criteria which would be appropriate for HEIs with different self-concepts was one of the main hurdles the commission had to overcome. However, they were faced with numerous other difficulties such as the sheer volume of material with which each expert was confronted and expected to engage with. One of the most obvious of these was the self reports provided by the HEIs, to which I shall turn first.

7.1 Techniques for working with large amounts of information: the self reports

Whenever *evalag* received a self report from an institution, this was forwarded to the 18 experts, or in the words of Expert O: 'the material arrived in pretty big boxes'.[207] The group did not have much time to read through the material before going onsite. As Expert U noted in his diary at the time:

> A big problem for us is that we have far too much material. No one of us can master all this documentation, and we have no time even to read the material just sent to us in time for our series of meetings, which begin in Mannheim tomorrow.

At this point the experts had long been allocated their HEIs, and so they knew which five they would be visiting. Would this then have an impact on the manner in which they processed the literature? Would they only examine the literature on their assigned HEIs and leave the rest to the others? As Evans (1999, p.180) has stated, 'to call for evidence and not examine it is a fault'. She also states that 'when a decision-maker has to take on board a good deal of evidence a problem arises about reading it all.' (ibid., p.179). How much information would the experts in this evaluation be prepared to reveal to me on this subject, considering that if some only read five reports this might reflect poorly on the quality of their own work or their engagement as an evaluator in general?

[207] Zuerst sind die Materialien dann gekommen und zwar […] in große Schachteln.

The majority of experts told me that they read with care only those HEI reports which they knew they would be visiting, concentrating in particular on those for which they already knew they would be compiling the evaluation reports: 'The net result was to look more carefully at those which one was going to visit. The others were sort of scanned.'[208] (Expert O) They used colourful language to describe their feelings towards the literature and the word which was repeatedly used was that it was overwhelming, '*erschlagend'*. Which particular strategies then did the experts develop in their work on the self reports? Expert G described his methodology as follows:

> I don't read it from front to back, but try to approach it rather with questions to get an impression of the site. That starts of course by my trying to remember which colleagues work there: who do I know, what do I know? So, I certainly look on the internet to make sure of the institutional structures in its self-presentation, to get some idea of the place.[209]

Expert E said that in any case it was impossible to gain an overview of all 15 HEIs:

> You can't process it all and of course you wish to enter into discussions with sound knowledge. So for those HEIs where I was going to chair the discussions […] I prepared myself far far more intensively than for the other HEIs, those I couldn't visit. […] I kept myself far more in the background in those cases, and also in the plenary meetings. [210]

Here we see a highly conscientious expert keen to do justice to the HEIs he is visiting; however, the limit of his knowledge on the other HEIs will prohibit him from taking part in discussions on them during plenary meetings. This point was emphasised by a number of other experts and they said that they were unhappy with this situation. However, due to the size of the commission and the time it took to see all 15 HEIs they accepted that there was no alternative under the circumstances. The number of sites meant that it was impossible

[208] Im Endeffekt hat man die, die man dann besucht hat, genauer angeschaut. Die anderen waren so etwas mehr im Überblick.

[209] Ich lese es nicht von vorne nach hinten, sondern versuche eben, ein bisschen mit Fragestellungen daran zu gehen und mir einen Eindruck von dem Ort zu verschaffen. Das fängt natürlich auch damit an, dass ich mir noch mal versuche ins Bewusstsein zu rufen, welche Kollegen dort sind, wer mir bekannt ist, was ich überhaupt weiß, dass ich sicherlich auch noch mal ins Internet gucke, um mich da noch mal ein bisschen über die institutionellen Strukturen der Selbstdarstellung zu vergewissern. Also, so eine Idee von der Einrichtung zu kriegen.

[210] Das kann man nicht verarbeiten und man möchte natürlich schon mit fundierten Kenntnissen an diese Gespräche herangehen. Also ich habe dann in den Einrichtungen, in denen ich quasi der Sprecher war, also für die Diskussion, da habe ich mich viel viel intensiver vorbereitet, als in den anderen beispielsweise, die ich nicht besuchen konnte. Da war […] meine Zurückhaltung viel größer, also auch in den Gesamtsitzungen.

for the group to visit each, and so they had to develop a basis of trust that their colleagues would be able to provide a comprehensive report on those they were unable to see. This leads me to examine the general attitude experts displayed towards the documentation with which they were provided.

7.1.1 Self reports and the art of seeing through façades

Although most of the experts engaged extensively with the self reports, most were clear that these could only present part of the story about the HEIs: 'My main reservation about the (self) reports is that they are restricted to certain lengths, and secondly, the way they are written can't always convey the situation as it really is.'[211] Expert D continued as follows:

> I have problems with the written reports. I know from almost all my experiences which I have made in connection with the [DFG's] Excellence Initiative that paper documents do not represent reality. And so during the discussions on the HEIs which I did not visit I kept myself in the background as I kept telling myself "I'm only getting half of the truth."[212]

This remark was corroborated by many of the others who concluded that self reports form a basis which can provide an initial idea of the HEI under investigation. However, only the onsite visits will enable a more differentiated understanding of the institution as people can be approached directly and have to answer questions immediately: 'Then they have to say something and one knows a lot about things which can be critical in an institution. You can see that very quickly.'[213] (Expert E) This point was also supported during a recent seminar by Sir David Watson at the Learning Institute at Oxford, 28 May 2009, who said that in general it is possible to gain a good understanding of the quality of an institution in less than 48 hours when onsite. Only one of the experts in the case study

[211] Nein, mein Vorbehalt gegenüber den Sachberichten ist, dass sie einmal auf bestimmte Umfänge begrenzt sind, und zum Zweiten ist die Formulierungskunst auch gelegentlich nicht so, dass die Verfasser der Berichte wirklich das abbilden, was die Realität ist.

[212] Ich habe einfach Probleme mit den schriftlichen Berichten. Ich weiß aus fast allen Erfahrungen, die ich jetzt auch wieder im Zusammenhang mit der Exzellenzinitiative gemacht habe, dass auf dem Papier einfach nicht das hinüberkommt, was Realität ist. Und daher ... […] aber in den Diskussionen über die Hochschulen, an denen ich kein Site visit mitgemacht habe, habe ich mich schon zurückgehalten, weil ich mir immer gesagt habe, "das ist nur die halbe Wahrheit, die ich kenne."

[213] Da müssen, da müssen sie es beantworten. Da müssen sie irgendetwas sagen können und man weiß sehr vieles, was kritisch sein kann an einer Institution, das sieht man sehr schnell.

(Expert N) stated that he believed the onsite visits were actually unnecessary and that it would instead be sufficient for representatives from institutions to come to a meeting at the evaluation agency. However, we could counter this argument when we consider that our evaluation experts had meetings with large groups of students who would have been unable to attend such a meeting. Furthermore, by visiting an institution experts are able to see for themselves the state of resources and the overall ambience of the site, which cannot be discerned from paper alone, but which make up part of what needs to be taken into account when assessing the quality of an HEI.

Returning to the self reports, without exception the experts engaged with them as far as they were capable. Arguably, however, their knowledge was restricted to those they would be visiting. Another point of criticism was that the self reports should have been more specific and that the criteria provided by the evaluation commission were not clear enough from the outset: 'Well, I thought we could have asked them more intelligent questions. Then they wouldn't have produced as much paper which was unnecessary in the end.'[214] (Expert F) This opinion was supported by this comment from one of the foreign experts:

> I think they need to be given quite clear instructions and in my opinion this wasn't sufficiently the case, on what should be in the report and to do it in the appropriate way, because I have seen reports, where people are, for example, losing themselves in all kinds of bureaucratic and administrative information that is relatively irrelevant. I know you need to have an idea of how the thing is organised and constructed, but sometimes they are telling their whole history, which is just a waste of time and paper. And so if they have clear instructions of what should be in it, then I think you have a better guarantee of steering them to reflect on what is important. (Expert Q)

I also asked about the extent to which the experts thought that the self reports accurately portrayed what was going on within an institution. Generally, their response was that the self reports were fairly good, although they witnessed an inevitable amount of window-dressing, which they described to me as '*Potemkinsche Dörfer*' or façades. This phenomenon has been frequently reported in the literature (cf. Kromrey, 2006) and this evaluation proved no exception.

[214] Also, ich fand, das habe ich vorhin schon mal gesagt, ich fand, wir hätten denen besser früher klügere Fragen gestellt, dann hätten sie nicht so viel Papier produziert, welches dann am Ende völlig unnötig war.

This was an aspect of the evaluation which was confirmed to me from a different quarter. In advance of the evaluation Martina Röbbecke had visited each HEI for two reasons: firstly to ensure that the logistics of the onsite visits would be in order and secondly, to inform and prepare the HEIs in advance of these. She told me that she had learnt a great deal during these visits and in some cases had been given insider information:

> You learn certain things, about internal conflicts, you find out about criticisms they might have with the university rectors and so on. Of course, they won't mention that when the commission gets there. That's clear. They will present a world that's more or less intact. The rector will be described as being successful, capable and so forth. You basically sit there as someone with prior information and watch how courageously and valiantly the commission battles its way through and tries to see through the façades [*Potemkinsche Dörfer*] which have possibly been created.[215]

In her role as the evaluation's organiser Röbbecke was of course unable to offer evaluative statements which the others might like to consider, although we know from her own background that she was trained in evaluation methodology and was highly capable of making such observations. Furthermore, in terms of the breadth of her knowledge of the 15 HEIs in the investigation, she probably had more insight into them than many of the other evaluators. She described her position as being akin to doing the splits or walking a tightrope as she tried to make subtle suggestions, without directly saying to the evaluators that they might like to focus more closely on one particular phenomenon.[216]

In terms of the quality of the self reports, these were rather diverse and some experts criticised the fact that they felt the HEIs did not take more of an opportunity to engage more seriously with considerations about the quality of their institution: 'There was so little perspective on change. What I had hoped for was that they would use this to reorganise themselves, to say this is an occasion to drive forward internal reforms. I was

[215] Sie erfahren meinetwegen irgendwelche Dinge, über interne Konflikte. Sie kriegen etwas von der Kritik am Rektorat mit und, und, und. Das sagen die Leute natürlich nicht, wenn die Kommission kommt. Das ist völlig klar. Das stellen sie alle mehr oder weniger heile Welt dar. Das Rektorat wird auch als erfolgreich, leistungsfähig und so weiter und so fort zu verkaufen. Und im Grunde sitzen Sie dann als Vorinformierte dabei und gucken, wie tapfer und wacker sich die Kommission dabei schlägt und versucht, diese Potemkinsche Dörfer, die da zum Teil aufgebaut werden, zu durchschauen.
[216] Der schwierige Spagat und Gratwanderung, in die man da rein gerät.

surprised that it was taken up so little.'[217] (Expert M) Admittedly, part of the problem lay in the fact that many of the 15 HEIs, and departments of Education in particular, were unused to compiling self reports, as quality assurance was still nascent in German higher education. Hornbostel & Simon (eds., 2006, p.20) have described how one of the main aims of external evaluation is to encourage self-reflection; however, it seems from what the experts told me that the HEIs took up this opportunity to greater and lesser degrees, which was partly reflected in the quality of their self reports. The commission was surprised by the poor quality of some of these and rather depressed that the academics in HEIs were not more responsive.

However, we might like to remind ourselves here that because the evaluation was commissioned by the Ministry there was an underlying sense of mistrust, which Röbbecke worked especially hard before the actual evaluation started to counter. She visited each HEI and spoke to numerous stakeholders who would be involved:

> I talked until I was blue in the face, to put it frankly. To weaken preconceptions, to motivate people and build them up a bit, to create trust. I think that that's a very very important prerequisite before arriving with the evaluators onsite, to have created a certain acceptance onsite.[218]

7.2 Attitudes to gaps in statistics

In order to gain an understanding of an institution, especially in terms of staff-student ratios, levels of doctoral support, staff numbers and publications, for example, it is necessary to examine a large amount of university data. This is frequent practice in many countries where quality assurance in HE has already become commonplace (Lohnert & Rolfes, 1997; Huber, 2003). However, when this evaluation was conducted, it was still unusual for German departments of any kind to keep accurate and up-to-date facts and

[217] […] so wenig Veränderungsperspektive deutlich wurde. Also, das, was ich gehofft hatte, nämlich dass sie das nutzen, um sich anders aufzustellen, um zu sagen, das ist jetzt ein Anlass dafür, innere Reformen voranzutreiben. Das hat mich schon überrascht. Dass das so wenig genutzt wurde.

[218] Ich hatte mir zum Teil den Mund fusselig geredet, um es deutlich zu sagen. Um da Vorbehalte so ein bisschen zu entkräften, die Leute zu motivieren, so ein bisschen aufzubauen, Vertrauen zu schaffen. Ich finde, das ist eine ganz, ganz wichtige Voraussetzung, um nachher mit den Gutachtern vor Ort zu kommen und so eine gewisse Akzeptanz geschaffen haben vor Ort.

figures and even the HEIs in their entirety did not hold comprehensive and complete datasets. One of the associated problems in this evaluation then was that although the self reports contained a wealth of statistical information, the material provided was not necessarily complete. It was on occasion difficult for the experts to gain an accurate picture of the HEIs under examination: 'We were given data which were totally wrong. You could tell that when you looked through it.'[219] (Expert F) Statistical information is frequently undervalued by those providing it (Maats, 2007). Expert H commented ironically: 'Well, that's a big phenomenon anyway, that this whole educational data is totally incorrect.'[220] Indeed, one of Röbbecke's main tasks and challenges as the evaluation's coordinator was to ensure that these statistical gaps were filled; a considerable amount of her time was therefore spent contacting HEIs to chase up missing data and to ensure that what she was given was actually correct and complete. Her efforts were recognised by all of the experts who knew how essential it was to be able to report accurately on figures which were true and comparable. If an institution discovers that an evaluative judgement on their HEI was based on incorrect data, then this reflects badly on the experts:

> You immediately lose credibility as a commission if they can prove that you don't know about the figures you are talking about, where they came from, and there will be insecurities on the quality of the data. That's lethal.'[221] (Expert T)

The experts criticised the fact that the data were incomplete: 'You really should have recourse to documentation systems which are maintained on an ongoing basis. That will happen at some point, but at this time it wasn't the case, or just beginning to get there.'[222] (Expert F) Due to issues relating to data protection in German HEIs in general, it seems to be the case that there is no agreement overall on what statistics should be kept, nor are these readily available. Other countries, notably Great Britain, maintain student statistics in

[219] Wir haben Daten gekriegt, die haben hinten und vorne nicht gestimmt. Das konnte man dann merken, wenn man mal durchguckt.
[220] Aber das ist sowieso ein großes Phänomen, das diese Bildungsstatistiken alle von vorn bis hinten nicht stimmen.
[221] [M]an verliert […] bei den jeweiligen Institutionen sofort an Glaubwürdigkeit als Kommission, wenn die einem nachweisen, die wissen gar nicht, über welche Zahlen sie reden, was die Quellen sind und da sind Unsicherheiten über die Güte der Daten. Das ist also tödlich.
[222] Eigentlich müsste man auf Dokumentationssysteme zurückgreifen können, die kontinuierlich geführt werden. Es wird auch irgendwann mal so sein, aber zu dem Zeitpunkt war es eben immer noch nicht, oder noch lange nicht so.

large databanks (cf. HEFCE) which allows for cross-HEI comparison and long-term analyses of standards. We can determine, for example, precise numbers of students in HE and their final qualifications on departure. Although the German Federal Statistics Office (*Bundesamt für Statistik*) maintains data on a number of areas, it seems that a national data base for German higher education would be useful for quality assurance work in general.

It seems useful at this juncture to briefly re-examine the kinds of information which the experts required and which the HEIs had hitherto been unused to keeping information on, as this indicates again where their emphases on quality in HE lay. These two extracts reveal the reactions and problems associated with requesting data from HEIs in Germany at this time:

> For many [HEIs] it was disconcerting. Many institutional heads didn't have personnel statistics readily available and they don't know what was going on, that we wanted [information] on short-term contracts, that we wanted to know what support there was for teachers, that was unclear to most. That we wanted to know if there was a monitoring of the academic careers of graduates was unfamiliar to most, if not to all.[223]

> We wanted data on all members of staff, also data on length of contracts, publication data. And that was where there was a disagreement with the main staff union, which didn't agree with our request. And the Ministry decided to go ahead with it anyway, but allow the research staff to submit their data voluntarily. [224] (Expert N)

The last extract reveals the extent of pressure from above to comply with the commission's request, but it also shows that the Ministry had a 'soft' approach, giving the implicit message to staff that while being free to decide for themselves, it was still expected that they would produce the data accordingly. Although the data provided were not entirely complete, the commission was finally able to fill most of the gaps.

[223] Das hat ... das war für viele beunruhigend. Viele Hochschulleitungen hatten die Personalstatistik nicht sofort, die wissen auch nicht was passiert; dass wir befristete Stellen haben wollten, dass wir wissen wollten, wann die Stellen frei werden - das war vielen unklar; dass wir wissen wollten, was für die Lehrer ... Angebot für Lehrer speziell angeboten wird, war für viele unklar. Dass wir wissen wollten, ob es ein Monitoring der Academic Careers, der Absolventen gibt, war für die ganz ungewöhnlich, für viele, nicht für alle.

[224] Wir wollten ja die Daten aller Personen haben, also auch die Befristungsdaten, die Publikationsdaten. Und dort hat es einen Widerspruch des Hauptpersonalrates gegeben. Der Hauptpersonalrat hat dem nicht zugestimmt. Und das Ministerium hat entschieden, dies trotzdem zu machen, aber die Ablieferung der Daten bei den wissenschaftlichen Mitarbeitern auf freiwilliger Basis laufen zu lassen.

7.3 Numbers of students at onsite meetings

The onsite visits and individual groups of experts who were allocated to each HEI were organised by *evalag*. One aspect which the HEIs were allowed to organise for themselves were the meetings with students, although Röbbecke did ask for a certain number of participants (20) who were usually recruited via the student representative body, Asta (the student union). They could determine how many participants, and from which degree courses/terms of study these came. It seems that no clear guidelines were issued, and so this was something I wished to explore with the experts as I thought that there might be problems arising from this. If, for example, the HEI handpicked its students from those groups closely affiliated with a particularly successful professor, for example, then their opinions might be biased and not representative of the wider student body. Alternatively, if the experts only met extremely small groups, they, too, might not be representative and lead them to a false impression. I therefore asked each expert what their experiences of the student discussions had been like.

Overall, little thought seems to have been given to the composition of the student groups, although when I enquired further, they admitted that there might have been methodological concerns with this. Nonetheless, all emphasised the importance of being able to meet students. This comment is representative:

> The discussions with students were always very informative and important. The number isn't that relevant provided that you have a couple chosen randomly from the different degrees and as we always held these talks before our final discussion with the professors and institutions we got some instructive findings.[225] (Expert J)

Speaking to the students before the professors and members of the teaching staff allowed some of their comments to be fed directly back. In one case, for example, the professors were utterly dismayed and shocked by the criticisms their students had voiced. Röbbecke

[225] [D]ie Gespräche mit den Studenten waren von daher auch immer höchst aufschlussreich und wichtig. Da ist die Zahl gar nicht so wichtig vorausgesetzt, man hat ein paar, die einfach nur so aus den unterschiedlichen Studiengängen dabei waren und die waren ja auch in der Abfolge der Befragungen immer vor der Schlussbesprechung mit den Professoren und den Instituten und es gab immer sehr aufschlussreiche Befunde.

commented on this as follows: 'They [the professors] came to the meeting with a healthy sense of self-confidence and I buy their shock. That wasn't put on.'[226]

The experts encountered a variety of different situations in the HEIs they visited, as some presented small numbers of students, whereas in others entire lecture theatres were filled, much to the surprise and initial consternation of the evaluation team. The following diary extract shows that the commission had to be prepared for any eventuality: 'We walk to another building, where we are expecting about 15 or so students from various subjects – and we find that over 200 have assembled. This throws us a bit off balance…' The logistics of moderating a discussion group that contains 200 rather than 20 students are quite different, and in this case the evaluation spokesperson decided to act as compère while her colleagues fielded questions.

Expert E told me that an ideal size for a student group discussion was between 20 to 50, but he reiterated what his colleague above had said about ensuring that representatives from different courses should be present, including the most important ones:

> I don't think that HEIs should be allowed to organise [an onsite visit] for themselves. The HEIs get into an odd situation if they create the impression that they didn't advertise it properly, then you only get too few students, or only those who were handpicked to ensure the right arguments are made. Both are insufficient. And so you should have a template which says you imagine meeting between twenty and fifty students in this discussion including different courses, such as diploma, and then more bachelors, masters to let you react. Otherwise someone important is left out and then you only get a one-sided comment from the students.[227]

[226] Sie sind selber mit einem gesunden Selbstbewusstsein in das Gespräch reingekommen und ich nehme ihnen diese Erschütterung sogar ab. Das war nicht gespielt.

[227] Ich glaube nicht, dass man das den Institutionen überlassen darf. Die Institutionen kommen damit in eine komische Situation, wenn sie dann den Eindruck erweckt, man hat es einfach nicht groß angeschlagen, darum kamen nur wenige Studierende, oder man hat es irgendwie ganz besonders handverlesen, damit da die guten Argumente kommen. Beides ist unzureichend, also sie müssten eine Vorlage haben, aha, so stellen sie sich das vor, also wir wollen zwischen zwanzig und fünfzig Studierenden begegnen können in diesem Gespräch und das sollten eben unterschiedliche Studiengänge so wie Diplomstudiengang und dann mehr Bachelor, Master sollen schon vertreten sein, damit man auch reagieren kann. Sonst fehlt da jemand wichtiger und dann hört man eine ziemlich einseitige Stellungnahme der Studierenden.

7.4 Compiling reports

Each group that went onsite had a report writer who would compile the group's findings post-visit and then feed them back to the wider commission after the round of visits was concluded. I was interested to learn how such a large group organised this task, as I imagined that each group might lay emphases on different areas and that there might be quite large variances in style, depending on the sub-discipline each person came from. The report writers had been given a template to use, which this expert commented on as follows:

> It's sometimes rather mechanical. You've got the document, have to fill it in and you can't really write properly, […] but in the commission you have to simply do things like a mechanic, otherwise it won't work. If we are all geniuses, then we couldn't work together.[228] (Expert K)

However, this pragmatic view which described the job as almost robotic was not the majority opinion, as most of the others were quite emotive in their response to my question on the writing of reports. They related how different subject areas did indeed impinge on their ability to write in a standardised manner: 'Put frankly, if an educational psychologist writes a report of this kind, it will be different from the manner in which an historically-oriented educationalist writes it. There will be considerable differences.'[229] (Expert E) This was described to me in greater detail by Expert G thus:

> But the question is: how comparable were things among the other groups? One got the impression, well, that was my personal impression, that the approaches were somewhat different according to whom we had in this […] commission and how they tackled things and what the peculiarities, preferences of individuals were. And so it's always a little bit the question of whether the criteria were always alike and that was one aspect, above all in the editing stage, in which we invested a great deal of time and also in plenary sessions that dealt with the onsite reports, to try to find a comparable style and equivalent criteria. For example, the necessary restraint in making suggestions, appropriateness in the phrasing of criticism and perhaps paying some attention to the weaknesses insofar as one

[228] Das hat zwar manchmal etwas Mechanisches sozusagen, die Datei ist schon da, man muss nur einfüllen und man kann gar nicht richtig schreiben und das, also, aber in der Kommission muss man dann einfach wie ein Mechaniker diese Dinge machen sonst klappt das alles nicht. Wenn wir alle Genies sind, dann können wir nicht zusammenarbeiten.

[229] Also, etwas extrem gesagt, wenn ein pädagogischer Psychologe solch einen Bericht schreibt, kommt es anders heraus als wenn ein historischer erziehungswissenschaftlich orientierter Pädagoge das macht. Und das ist erheblich unterschiedlich.

> thinks that certain publication situations or research projects need to be developed or not; well, there was a certain amount of exchange and a process of reaching agreement.[230]

Being able to do justice to the HEIs they had visited by writing a report that bore correct testimony was something which therefore worried a number of the experts and was time-consuming. While they acknowledged that their own evaluative judgements might not always be in conformity, much effort went into formulating their findings in a manner which would be homogenous, although we know from various comments that some report writers invested more, or less, time in their reports, and as a consequence what was offered to the group in plenary sessions was not always of equal quality from the outset. The chairman told me that much of the work of creating a similar style was done by Röbbecke in conjunction with Baumert and Tenorth who both invested heavily in the editing work: 'What you can see in this report is pretty homogenous, uniform […]. It was important that we agreed on the same structure.'[231] We should recall that Baumert and Tenorth represented two different approaches to Educational Science, and so the fact that both of them worked on the editing ensured that the final result was balanced between the two main paradigms. Although there may occasionally have been situations when opinions were particularly polemic, it was seen as necessary to avoid this in the presentation of the final report. This is in contrast to more recent trends in evaluation which allow and indeed encourage evaluators to take a polemical stance (Lee, 2006).

Regarding those who had been asked to compile the individual onsite reports (see Tables 7-9, Chapter 4), these experts perhaps inevitably exercised more influence on the

[230] Aber die Frage ist, wie vergleichbar war das jetzt über andere Gruppen und da ist auch so ein bisschen der Eindruck entstanden, also mein persönlicher Eindruck entstanden, dass die Zugänge ein bisschen unterschiedlich waren, je nachdem wer halt in dieser […] Kommission war und wie man das angegangen ist und was auch so die Besonderheiten, Vorlieben der einzelnen Leute waren. Das ist natürlich auch immer ein bisschen die Frage, ob die Kriterien überall ähnlich waren und das war schon ein Punkt, wo wir vor allem bei der redaktionellen Arbeit, bei dem Bericht über die Standorte auch im Plenum immer wieder doch immer auch eine Menge Zeit rein gesteckt haben, um dann eben auch zu versuchen, einen vergleichbaren Stil zu finden, auch vergleichbare Kriterien. Eben auch die gebotene Zurückhaltung was Vorschläge anbemisst, die Angemessenheit der Formulierung von Kritik und eben vielleicht auch ein bisschen das Eingehen auf die Schwächen, wieweit man bestimmte Publikationslagen oder bestimmte Forschungsprogramme als entwicklungsbedürftig betrachtet oder nicht, also, da gab es schon einiges an Austausch und Abstimmungsprozessen.

[231] Denn was man sieht in diesem Bericht ist ziemlich homogen, einheitlich, und ich glaube, dass hat sie getan. Wichtig ist, dass wir ja eine gleiche Struktur verabredet haben.

commission than those who did not do any writing. This mirrors the situation already mentioned of those who do not have as much influence as they chose to be less vocal in plenary meetings because they believed they did not have enough expertise to contribute meaningfully to the discussion:

> Whoever stays in the background has less chance of representing their opinion and it is a sad fact that [writing reports] does take up a lot of time, which you have to take into account. And so not every commission member participated equally.[232] (Expert T)

There is a danger then that reticence is equal to less influence and might lead to bias in what a commission can achieve.

7.5 Discrepancies between discussions and final report

In Chapter 4 I provided details of the statement Achtenhagen and Schelten prepared in which they voiced their reservations about the proposed cooperation model of technical colleges and teacher training colleges in the training of vocational teachers. These had hitherto been located within universities and were of high status. They felt that any such cooperation would lead to a devaluing of the status of this area in the discipline. Knowing that almost all of what they had written and recommended had been omitted from the final report, I was keen to discover how they felt about this and I also wished to learn more about the background of this particularly difficult area where it seemed that the commission had elected to move away from those members who were the vocational experts. I begin with a statement from the Ministry, which told me why this cooperation model was part of the discussion:

> Years ago […] we developed a new model to train vocational teachers in cooperation with technical colleges and teacher training colleges, because we weren't getting enough vocational teachers out of the universities. And this parallel model was rejected strongly by those who represent vocational training in other *Länder*. We noticed that very clearly. They had their own specific viewpoints and own interests at stake along the lines of

[232] Wer sich da sehr stark zurückhält, hat weniger Chancen seine Position zu vertreten und die traurige Tatsache, dass das eben leider auch viel Arbeit macht, muss man dann in Kauf nehmen. Und da hat sich natürlich auch nicht jedes Mitglied in gleicher Weise engagiert.

> "You've got a rival model? We don't want that. We reject that. That isn't our idea of quality."[233]

We see here that the idea was not to downgrade the training of vocational teachers to technical colleges and teacher training colleges per se, it was rather to have this occurring alongside the existing university structure, in particular where there were not enough chairholders to ensure sufficient coverage of the subject.

Another member of the commission told me that in their opinion the vocational experts were engaging in an ongoing 'battle' for status: 'they fought for recognition with the other men, which was really obvious. Well, it was always about acceptance.'[234] (Expert L) This was indicated in the following statement from one of the other men in the commission:

> Well, the statement with which we [names of two experts] always managed to cause the most displeasure among the vocational experts was: you could basically train your people in technical colleges. That is a statement which, as long as you have vocational experts in a commission, must never be written anywhere. To some extent it is written here, because we explicitly praise cooperation models between teacher training colleges and technical colleges.[235] (Expert J)

Although the evaluation of Education was concluded several years ago, the omission of the vocational expert's recommendations was still a sore point; they described this as being 'outrageous' (*empörend*). One of the other experts told me why this might be: 'As is often the case of those coming from the world of practice who seek to gain a place in the heights of science, they have to hold up the flag of science even higher than science itself does'[236]

---

[233] […] wir haben vor einigen Jahren […] ein neues Modell entwickelt, indem wir Berufsschullehrer in Kooperation mit Fachhochschulen und Pädagogischen Hochschulen ausbilden, weil wir über die eigentliche Berufsschullehrerausbildung an den Universitäten nicht genügend Nachwuchs bekommen. Und dieses Parallelmodell wurde von denen, die die Berufsschullehrerausbildung an Universitäten in anderen Ländern vertreten, zum Teil deutlich abgelehnt. Da haben wir sehr deutlich gemerkt, da waren auch sehr eigene Sichtweisen, und sehr eigene Interessen mit im Spiel. Ich sag jetzt mal „Konkurrenzmodell? Das wollen wir nicht, das lehnen wir ab. Das entspricht nicht unserer Vorstellung von der Qualität."

[234] […] die haben gekämpft mit den anderen Männern um die Anerkennung, das war sehr deutlich. Also, es ging immer um Anerkennung.

[235] Also, der Satz, mit dem wir immer […] großen Unmut bei den Berufspädagogen erzeugt haben, indem wir gesagt haben, an sich könntet ihr eure Leute auch an der Fachhochschule ausbilden. Das ist ein Satz, der darf, solange Berufspädagogen in einer Kommission sind, nicht, nirgendwo geschrieben werden darf. Der steht hier trotzdem in gewisser Weise drin, weil wir ein Modell der Kooperation von Pädagogischen Hochschulen und Fachhochschulen ja ausdrücklich loben.

[236] Und wie das immer mit Aufsteigern aus der Praxis in die Höhen der Wissenschaft ist, die müssen ungeheuerlich die wissenschaftliche Fahne hochhalten, mehr als die Wissenschaft selber.

(Expert L), an indication that the members of the commission, despite knowing each other well, did not always occupy the same status in academia even if they held many credentials and were well-known and highly respected. The majority of the commission therefore adhered to their view that some of the sites in Baden-Württemberg should seek cooperation with technical colleges (*Fachhochschulen*) in particular, as some of these were operating on a level which was on a par with that offered by the universities, a fact that 'could not simply be swept aside.'[237] (Expert N)

If we expand the notion of discrepancy between the opinions presented during plenary meetings to those published in the final report we discover that there was a tendency to be cosmetic and risk-averse. To some extent, this echoes Power's (1994, p.xvii) notion that evaluation can obscure insight into the underlying complexity of what is being examined. The evaluators themselves are stakeholders and even in their role as outside reviewers they are inextricably bound into disciplinary structures and social 'borders' which make it difficult to step considerably beyond the borders of the norm. Although there was agreement that the chair and sub-committee in particular were able to achieve consensus within the commission, it could be argued that this decision was made at a cost of presenting more radical ideas within Educational Science, which might possibly have given the discipline more impetus. Although this notion is hypothetical, it is supported by various statements in interview, of which this is an example:

> You can sense that here and there in the solutions, well, they were far more radical during the evaluation process than in the final report. I sense a clear discrepancy between the process and the course of writing up interim reports in comparison with the final report, which is presented in a more moderate form than the manner of the process in the middle or even towards the end.[238] (Expert E)

Röbbecke also confirmed that within the report 'there are many careful formulations, weighings up. I think you can hear clearly in what kind of an area of conflict we were

[237] Das kann man nicht so einfach abbügeln.
[238] Und das spürt man dann hier und da bei den Lösungen, also, die waren sehr, viel radikaler im Verlauf des Evaluationsprozesses, als im Schlussbericht. Ich spüre eine deutliche Diskrepanz zwischen dem Prozess, oder dem Verlauf zu den Teilberichten, zum Schlussbericht, der in einer moderateren Form daher kommt, als der Prozess so in der Mitte, oder auch noch gegen Ende sich gestaltet hat.

moving'.[239] It might therefore be argued that compromise, while desirable, itself compromised the final result, and as a consequence the evaluation may have been less effective in its outcome. This again raises the issue of utility. Evaluation results are not arbitrary, they are intended for use (cf. Vedung, 2004, p.132), and so if they are more controversial in their outcome, this might affect a greater reaction than if everything is toned down too greatly, with the intention to avoid harm. This situates evaluation clearly in a field of conflict between social and political interests (cf. Stockmann, 2006b, p.262). In an evaluation which attempts to take all aspects of a discipline into account in a range of different institutions, this task becomes almost impossible, or even absurd in its undertaking as it will be unable to gain that degree of interpretation and understanding on all levels, in spite of the best efforts of the evaluators to the contrary. Bolius had mentioned that she wished the evaluation to be forthright; however, the very fact that there was an influential Ministry in the background meant that any forthrightness that might have existed during the plenary discussions had been consciously toned down in the final report. There was no lack of discussion during the meetings. Indeed, Röbbecke told me that a number of the German evaluators were quite belligerent, in contrast to the foreign experts who were far more conciliatory.

7.6 How the experts thought they were perceived by the HEIs

Considering that 15 HEIs were examined and that each expert was only able to visit five HEIs in total, I did not expect to be given consistent answers as to how they thought they were regarded by the HEIs, and how their task and their suggestions were accepted. However, I was not prepared for the variety of responses which they gave me. This ranged from experts who felt that the HEIs 'received [us] unconditionally and in an open manner. They were willing to co-operate'[240] (Expert D) to those 'who appeared to be very defensive and showed so little perspective for change'[241] (Expert M) to a general 'they were

[239] Es gibg ganz viele vorsichtige Formulierungen, Abwägungen. Da hört man glaub ich deutlich in was für ein Spannungsfeld wir uns da bewegt haben.
[240] Die Aufnahme in den einzelnen Institutionen war aus meiner Sicht vorbehaltlos und offen. Die Bereitschaft zu kooperieren war gut.
[241] […] etliche (waren) sehr defensiv eingestellt und ...[hatten] so wenig Veränderungsperspektive.

scared'[242] (Expert P), although this last comment was made in reference to the teacher training colleges rather than the universities. I was told that the universities seemed generally less concerned about this evaluation, reflecting perhaps the low status of Educational Science: 'Well, they put the significance of an evaluation commission into perspective'[243] (Expert B), or the fact that the universities were somewhat more used to being evaluated than the teacher training colleges had been thus far. Röbbecke confirmed, however, that the HEIs regarded the evaluators generally as being the elite of their discipline: 'You have really gathered together everyone with reputation and standing in Educational Science.'[244]

I interviewed three of the 15 HEI rectors (20% of the total number), two from teacher-training colleges and one from a university. Would the expert's idea of how they were seen by the HEIs be corroborated by these institutional heads? One of the rectors from a teacher training college immediately stressed that the experts were all from universities:

> This was a problem for us because they kept asking: where are your research projects, where is your third-party funding and so on, because they had a university-based outlook. That was like a *leitmotif* in all of the evaluations we had.[245]

This comment suggests that a fundamental methodological problem existed in association with the teacher training colleges. As the experts were all selected from outside Baden-Württemberg, no one had any direct experience of this institutional type or knew how these operated. Although it could be argued then that the experts would not impose any insider knowledge upon the evaluation, it could equally and perhaps more strongly be argued that this was a deficit, as they lacked a deep understanding of the culture of the teacher training colleges.

The rector continued his portrayal of the evaluation commission as follows:

---

[242] […] sie hatten Angst.
[243] Also, die relativieren die Bedeutung einer Evaluationskommission.
[244] Sie haben hier wirklich alles zusammen gesucht, was Rang und Namen in der Erziehungswissenschaft hat.
[245] Das war für uns immer ein gewisses Problem, weil die immer fragten, wo sind eure Forschungsprojekte, wo sind eure Drittmittel und dergleichen. Weil sie aus ihrem universitären Verständnis kamen. Das zog sich aber als roter Faden durch sämtliche Evaluationen, die wir hatten, durch.

> There was an article entitled “Here come the evaluators” about this evaluation in a scientific journal. And that pretty much hits the nail on the head. OK, we’ll go along and see if you are doing a good job and we’ll stay for a day and then we’ll be able to tell you precisely what you could be doing better. This was really pronounced in this evaluation of Education, although one should realise that in this field we are to some extent competing with the universities. Universities have Educational Scientists, too, as do teacher training colleges. And we were somehow given the impression that *real* Educational Science [my emphasis] was being done at universities. The Educational Science you’re doing in teacher training colleges can’t really be the thing. Otherwise the colleagues would have become university professors.[246]

From this comment it would appear then that it is essential for evaluators to have considerable knowledge of the object of evaluation and its culture. Although the experts in this case were highly knowledgeable in their subject areas, they were unfamiliar with the culture of the PHs, which could be regarded as a deficit. This leads me on to a discussion of how the experts regarded their work and the qualities that they believed they should bring to this.

7.7 Self-concept of evaluators

When experts are selected to work in an evaluation some will have more expertise (or a better overview of the discipline) than others, and some will have greater evaluation knowledge than others. I was particularly interested therefore to learn how they felt about themselves and which qualities they thought they should bring to the evaluation. What follows is a list in the order of priority according to which the experts named them, with the most-frequently named first.

7.7.1 Ability to reach compromise

[246] In einer Fachzeitschrift stand damals ein Artikel über diese Evaluation mit der Überschrift "*Die Evaluatoren reiten ein*". Und das charakterisiert eigentlich relativ gut, so, jetzt kommen wir dahin und jetzt gucken wir mal, ob ihr eure Sache auch richtig macht und wir gucken einen Tag und dann können wir euch genau sagen, was ihr besser machen könntet. Das war auch sehr ausgeprägt in dieser Evaluation Erziehungswissenschaften, wobei wir in diesem Bereich, das muss man vielleicht auch noch sehen, in gewisser Weise mit Universitäten konkurrieren. Auch die Universitäten haben Erziehungswissenschaftler. Die Pädagogischen Hochschulen auch. Und es wurde uns so ein klein bisschen der Eindruck vermittelt, die richtige Erziehungswissenschaft gibt es an den Universitäten. Die Erziehungswissenschaft, die ihr an den PHs macht, das kann ja gar nicht so richtig etwas sein. Sonst wären die Kollegen ja Universitätsprofessoren geworden.

In Chapter 4 I indicated that the commission was keen to reach consensus of opinion and compromise and in so doing avoid direct confrontation with their peers, as they viewed their task overall as one of support, improvement and transformation. They perceived of themselves as being 'agents of change' through support (Hanft, 2003, p.15). Although compromise was the most cited quality, what this did not mean was that it disallowed for debate or the fact that individual voices were completely lost:

> It's certainly desirable to reach general recommendations through intensive discussions, without it being diluted into a compromise that no one is really happy with. And I thought that our commission did that rather well.[247] (Expert T)

> [There was] an ability to reach compromise while maintaining the core essentials, which one wished to keep.[248] (Expert M)

This opinion was supported by the ministerial observer who watched the group while they were discussing the different aspects of their work:

> From time to time I saw things being discussed controversially, but they, and I ascribe this to the group's leaders, always managed to get the group to a point where they said, OK, we can agree with this, we can give our approval.[249]

This underlines the notion that the chair and sub-committee were highly influential in enabling this to occur.

7.6.2 Knowledge of field and publishing experience

We have already examined the criteria for evaluator selection which were used by the Ministry and *evalag* in advance of this evaluation (see Chapter 4). Each and every expert had been chosen for their specific subject knowledge and in some cases considerable university administrative experience, which added a further dimension to the way they could assess the Baden-Württemberg discipline of Education within the overall context of Education in Germany. If experts who were appointed were less able than those they were

[247] Wünschenswert ist sicherlich, dass man über intensive Diskussion sich eher auf gemeinsame Empfehlungen einigen kann. Ohne dass es vermischt wird zu einem Kompromiss, mit dem eigentlich keiner richtig zufrieden ist. Und das fand ich in unserer Kommission recht geglückt.
[248] Kompromissfähigkeit bei gleichzeitiger doch Beibehaltung der Kernessentials, die man vertreten möchte.
[249] Ich habe eigentlich, zwar erlebt, dass man Dinge auch gelegentlich kontrovers diskutiert hat, dass man es aber immer geschafft hat, und das lag sicherlich an der Leitung der Gruppe, dass man es immer geschafft hat, die Gruppe zu einem Punkt zu bringen, wo man gesagt hat, gut, so können wir einverstanden sein, so können wir zustimmen

assessing this would have a negative impact on the evaluation (Morley, 2003). In the words of one expert: 'You can't get [HEIs] to list their publications if others [referring to experts on commissions] haven't published. [The experts] have to be above average, otherwise it just won't work.[250]' (Expert N) Establishing publication criteria that involved inclusion, if possible, of international journals, many of which were written in English, was rather difficult for the group, considering that most of them were not used to doing this, although they were putting this forward in their final report as one notion of excellence in Educational Science: 'For them it was frightening that English publications were now being included. But they went along with it in the end'.[251] (Expert N) We have an opportunity to find out from an observer what he thought about those who were selected, given that he was present during their meetings and was able to reach an understanding of their levels of ability:

> If I don't know the significance of good research and good publishing of research results, then I can't say if it is good here or bad elsewhere. And in this case you could observe that they [the experts] had a very good overview and on occasions very profound knowledge.[252]

Although this member was no educational scientist, we should assume that he was sufficiently able to judge the level of depth which the experts brought to the debate and that he certainly believed that they covered the breadth of knowledge necessary to be able to assess the 15 HEIs accurately.

If we now turn to the experts, the interviews were consistent in that they felt it was necessary to have not only breadth of knowledge, including if possible international experience, but also subject expertise. Furthermore, if they had experience of different institutions this was seen as an advantage. Personal knowledge of the field should also include an ability to see where connections might be made: 'Yes, I think especially where

[250] Man kann nicht Publikationen auszählen lassen, wenn die Personen selber nichts publizieren. Die müssen schon im oberen Mittelfeld oder darüber liegen, sonst funktioniert das einfach nicht.
[251] Und für die war das schon unheimlich, dass plötzlich die englischsprachigen Publikationen gezählt wurden. Aber letztlich haben dann doch alle mitgespielt.
[252] Wenn ich nicht weiß, was gute Forschung und gute Publizierung von Forschungsergebnissen eigentlich heißen kann, kann ich nicht sagen, dort ist es gut und dort ist es schlecht. Und das hat man hier gemerkt, die haben einen sehr guten Überblick gehabt und zum Teil ein sehr profundes Wissen.

things are becoming more globally oriented, not just networks, but possibilities for cooperation: where, for example, is a colleague in Germany with whom I could perhaps cooperate in order to push research forwards?'[253] (Expert M) The following extract comes from an interview with one of the foreign experts:

> The competencies that are needed to do this? I think it depends a little bit from what perspective you are participating in it. For instance, I think that depends, when I speak especially about when I look at a teaching evaluation, I look at it from my expertise as an educational scientist, and so I think it is important to have quite a bit of experience with looking at what's going on in educational situations, but you also have to have a framework that allows you to evaluate, because you are having to base yourself mainly on a kind of self study report, and you have to look at this from a certain educational framework in my opinion, and so you need to have quite a bit of experience as an educator. On the other hand of course if it is an evaluation like I was involved in the [name of discipline] faculties, then you need to have someone who is an expert in the domain. Different members of the committee therefore need to have different kinds of expertise. (Expert Q)

7.6.3 Empathy with academic staff

In this evaluation the experts were very aware of their role as someone who was 'passing judgement' and everyone expressed the view that although they would say what they believed was true, they would attempt to do so in a manner which would not incur pain. There was repeated use of the terms 'injury' (*Verletzung*), or 'hurt' (*weh tun*). This point was supported by the ministerial observer, who nonetheless stressed that this did not mean that recommendations were suppressed to avoid inflicting 'pain'. The groups consciously tried to create a good working atmosphere with the HEIs they visited, to relax the mood, and thereby enable a constructive discussion of what was working, and what might be improved: 'It's more about a collegiate discussion, which takes place there, rather than a top-down one.'[254] (Expert O) As they were dealing with colleagues who in turn might later come to evaluate members of the commission, the tone of the written report needed to be empathetic without skirting the issue of stating perceived realities.

> I think it is important to be able when you take part in the visits to have a kind of empathy for the situation because every institution has its own culture. You need to have

[253] Ja, man muss auch da, ich denke, häufig speziell da, wo alles ein bisschen globaler ausgerichtet wird, nicht die Netzwerke an sich, aber Kooperationsmöglichkeiten, wo ist zum Beispiel ein Kollege irgendwo in Deutschland, mit dem ich vielleicht kooperieren könnte, um eventuell die Forschung voranzubringen.
[254] Es geht eher um die kollegiale Besprechung, die dort stattfindet und nicht irgendwo von oben herab.

> a kind of feeling for what the situation is like here. You need to be capable of asking the right questions that really go to the essence of what the exercise is about. (Expert Q)

However, this was linked to their own self-concept in terms of the question: How do experts dare to give themselves the 'power' to pass judgement over their peers?

> I arrive and I have to let them ask me the question how I dare judge them? What gives me the right to do this? I think that is a very justifiable opinion, which I also have towards myself, to ask myself, why do I see myself entitled to say anything. And I think I have a view from the outside, I can ask questions on the basis of my own experience and the outside perspective and that colleagues can find this helpful to some extent. So: first of all I'm not coming to do the dirty on someone or to cut them off or something like that, or to suggest to him how he can do things better, because I know it all, that can't be it. And so, to impart the attitude that through discussions or questions you can help him see his own case from that point of view, to stand next to himself and to see what he's doing. And that he doesn't have to defend himself the whole time. And it is really difficult to get them out of this attitude of having to defend themselves all the time for what they are doing… […][255] (Expert L)

7.7 Gender issues

Although this study does not concern itself with feminist issues, the topic did arise during one interview with regard to two aspects. After receiving the self reports from the HEIs, the statistics had not been presented with any gender differentiation, as this had not been requested in the initial brief to the HEIs. One member of the commission had therefore asked for a more detailed breakdown of the gender of students overall and those completing doctorates and *Habilitationen*. This might be regarded as an initial point of criticism to be levied against the evaluation commission, as these kinds of gender-differentiated statistics are now commonplace in HEI analyses. The final report does, however, provide this data as *evalag* went back to the HEIs and the breakdown was given.

[255] […] Weil ich jetzt komme und ich muss mich fragen lassen, was maßt Du Dir hier eigentlich an, uns zu bewerten? Woher nimmst Du das Recht, das zu tun? Das finde ich erst einmal eine sehr berechtigte Haltung, die ich auch mir gegenüber habe, mich frage, warum sehe ich mich eigentlich berechtigt, was zu sagen. Und ich denke, ich habe einen Blick von außen. Ich kann Fragen stellen auf dem Hintergrund meiner eigenen Erfahrungen und mit dem Blick von außen und das kann in einer gewissen Weise auch für die Kollegen hilfreich sein. Insoweit, erstens komme ich nicht, um jemanden in die Pfanne zu hauen oder abzuwürgen oder so etwas, oder ihm vorzuschlagen, was sie besser machen müssen, weil ich es so gut weiß, das kann es eigentlich nicht sein. Also, diese Haltung zu vermitteln, dass man ihnen, sozusagen, im Gespräch oder mit Fragen darauf zu verhelfen kann, ihre eigene Sache von der Seite zu sehen. Also, sich mal neben sich zu stellen und zu gucken, was man tut. Und sich nicht nur zu verteidigen. Und das ist ganz schwierig, dass man aus dieser Haltung rauskommt, dass sie sich immer nur verteidigen für das, was sie machen, was sie noch alle Tolles tun und überhaupt und so.

One expert told me that there was a complete lack of gender awareness among the male members of the commission: 'That really surprised me. I hadn't expected it to be like that anymore. […] Although I should have thought as much, it still took me by surprise.'[256] (Expert L) I then enquired about the overall position of female experts within the commission:

> The women were certainly accepted, I think. The female colleagues who were there were accepted as long as they didn't get going on any women's stuff. […] They are all very experienced, stand their ground, they know exactly how they need to represent themselves in the HE arena in order to be given recognition. And as soon as they start up with any feminist issues, they no longer get recognition. You are then pushed into a corner and you are cut off, so to speak. [257] (Expert L)

Although I cannot prove this statement, similar observations on the status of women in academia have been made, as already referred to in Chapter 5. If we return to the matrix of connectivity among the experts in this case study, we might surmise that if an expert was both female *and* less well connected within the group her impact would be less great that those whose connectivity was greater. The matrix I constructed in conjunction with comments from other experts seems to support this argument. The females on the outside of the matrix were less well connected than their male colleagues, although we should remember that their expertise was from very specialised areas of Educational Science which in itself would make their connectivity within the discipline as a whole less strong. Arguably, however, if these areas had been represented by better connected male colleagues, then the input from these areas might have been greater.

We should be aware that what this expert said may be completely unrepresentative of how the other female experts saw their own position within the group and there is a strong case that the others were more pragmatic and neutral in their attitudes:

---

[256] Also, überrascht hat mich, obwohl ich es eigentlich hätte wissen müssen, noch mal diese völlige Abwesen-heit von Genderbewusstsein bei den Männern. Das hat mich sehr überrascht. Das hätte ich so nicht mehr erwartet. […] Obwohl ich es eigentlich hätte wissen müssen, es hat mich trotzdem überrascht.

[257] Die Frauen waren durchaus akzeptiert, also, ich denke, die Kolleginnen, die dort waren, die wurden durchaus auch akzeptiert, solange sie nicht mit dem Weiberkram gekommen sind. Und die haben auch, die haben ja lange Erfahrung, sind ja alles gestandene Kolleginnen, die wissen sehr genau, wie sie sich in der Hochschulszene inszenieren müssen, damit sie auch anerkannt werden. Und sobald sie anfangen mit diesen Frauengeschichten, wird man nicht mehr anerkannt. Dann ist man abgeschoben in eine Ecke und dann ist man, sozusagen, abgesondert.

> I'm always asked if that [the evaluation] is a female or male-specific thing. I don't see it like that. Perhaps the others see if differently that you are in a different role according to if you are a woman or a man. Both in terms of their numbers and their verbal acumen the men had more talking time than the women, I think. But I didn't get the impression that female or male aspects played a part. But maybe I'm too uncritical. I can't rule that out.[258] (Expert S)

However, we could suggest that the male members of an evaluation commission need to be aware of what their female colleagues might *not* wish to say, as they may not want to compromise their own positions of authority and respect which they may have worked especially hard to achieve. It also reveals that among this group at least, gender issues were not high on the agenda and that in itself is indicative of the status of this within the discipline. Had this one expert not raised the criticism about there being no gender-differentiated statistics, then the final report might have lacked this information.

7.8 Difficulties with HEIs

Although most of the onsite visits were conducted in an amicable atmosphere which we know the experts were keen to establish and maintain, they sometimes encountered unforeseen difficulties during onsite visits. Some were characterised by situations which were beyond their control:

> [Name of HEI] is pretty well equipped in staffing terms, but they are all at loggerheads internally. Attack each other. It's incurable. Even when we did our onsite visit they laid out all of their internal arguments in front of us. […] I haven't seen anything like it before. It was grotesque and totally inefficient. And incurable. All you can really do is to close it [the department] down. [259] (Expert N)

Another expert present supported this: 'It was dreadful. There was something absurd about it. A public denunciation.' (Expert P) The person chairing this meeting said his role was 'very difficult. It was like being in an, how shall I call it, open psychiatric

[258] Ich werde immer wieder mal gefragt, ob das eine frauen- oder männerspezifische Sache ist. Ich sehe es nicht so. Vielleicht sehen es andere anders, dass man da als Frau oder als Mann in einer anderen Rolle ist. Sicher haben nicht nur von der Anzahl der Personen her, sondern von der Wortmächtigkeit mancher Personen her die Männer mehr geredet, als glaube ich, die Frauen geredet haben. Aber ich hatte zumindest nicht den Eindruck, dass dort frauen- oder männerspezifische Aspekte eine Rolle gespielt haben. Aber vielleicht bin ich da zu unkritisch. Das will ich nicht ausschließen.

[259] [name of HEI] ist relativ gut ausgestattet vom Personal her, völlig zerstritten intern. Fallen über sich her, nicht heilbar. Selbst bei der Begehung haben die ihre ganzen internen Streitigkeiten vor der Kommission ausgebreitet. Das war schon sehr witzig (lächelt), so was hab ich noch nie erlebt, das war schon richtig grotesk und völlig ineffizient. Und auch nicht heilbar. Im Grunde kann man es nur schließen.

institution and that was catastrophic.'[260] Expert K told me more about his role in this situation:

> I had to do social management, so to speak, to stop things from getting out of hand, and we cut it short [the meeting] as much as we could […]. Afterwards I sat there bathed in sweat and we looked at each other and said that it's a complete breakdown. And then we said that every decision [from now on] has to be made externally and that only externally-appointed commissions should work there and that this subject group should be given no responsibility for itself. This was by far the most demanding thing I have ever encountered in such a situation.[261]

Closure was not an option the experts could recommend, however, as the Ministry had stated from the outset of the evaluation that it would not close down any HEIs, although it was well aware of the situation within this particular one. The chairperson told me that he discussed the subject with a member from the Ministry afterwards who told him: 'I'm sorry that you had to witness this. We have known for years that it's like that, but there's nothing we can do. I said, yes, with our recommendation you'll be able to do something'[262], referring to the fact that the immediate thought was one of closure. We should recall that university chair holders in Germany are civil servants and those in tenured positions cannot be removed from their posts, hence the inability of the Ministry to do anything about the situation. The only way in which a professor could be removed from one HEI would be if he were called to take up a chair elsewhere, which had not happened in this example.

When it came to writing up the onsite report Expert B, who was responsible for this task said:

> I can remember being criticised for toning down the report on [name of HEI]. […] I'm not sure if any great changes were made to the way I put it, but I'm not under the impression that my report was changed very much in the final version. But there were

[260] […] das war sehr schwierig. […] das war wie so eine, wie soll man da sagen, wie eine offene Psychiatrie sozusagen […], das war katastrophal!

[261] Ich musste sozusagen dieses soziale Management machen, damit das da nicht wegrutscht, er, und wir haben das dann auch möglichst verkürzt, er, weil das, hinterher habe ich da schweißüberströmt gesessen und wir haben uns angeguckt und haben gesagt, also das ist ja, da geht ja überhaupt nichts mehr, ne, und dann haben wir gesagt, dass alle Entscheidungen von Externen gefällt werden sollen und dass nur externe Berufungskommissionen da arbeiten sollen dort und dass man da die, er, also diese Fachgruppe keine Verantwortung mehr für sich übernehmen darf. Das war mit das Anstrengendste, was ich bei so was erlebt hab.

[262] […] es tut mir leid, dass Sie das sehen mussten hier. Wir wissen, dass das seit Jahren so ist, aber wir können nichts machen. Ich sagte, ja mit unserer Empfehlung werden Sie was machen können.

> certainly those who had been to [name of HEI] who said that they had a different recall of the visit than the manner in which it was written up.[263]

If we examine the wording of the final report it does indeed mask something of the crisis:

> […] As the onsite discussions show, the institution is characterised by internal differences among colleagues. […] In the self report on [name of HEI] the significant differences within the faculty were not flagged. These differences must have contributed to the fact that the scientific personnel of this faculty have been unable to create a recognisable, specific development profile in research and teaching. From the present personnel we cannot expect any improvements in this regard. A generational change should be used as an opportunity to create new areas of focus. The present incumbent should only be involved in new appointments as far as is set down by the law of the [institutional type].[264]

Time, therefore, was the only means of introducing a change or improvement to this faculty, which must have been a depressing situation for the experts to witness, keen as they were to offer support and advice. Here they were bound by the law and the matter was beyond their control. This might have been the reason for the report writer to use milder tones in his description, as there really was nothing to be gained from stronger language.

Another difficult occasion which a different onsite team had to manage was the complete antipathy displayed by staff in one HEI. Expert J described this to me as follows:

> The people came with an aversion: you haven't got the right criteria. You don't have any idea what this is about. You're assessing us wrongly. You're measuring us according to false standards. That kept coming through as a latent defensive argument and in [name of HEI] it was clear at the end of the final meeting [1-2 April 2003 meeting with Rectors and institutional representatives in Mannheim]. I don't know if you've heard about this scene. There was a real éclat. [Name of expert] threw them out. And we thought, stone the crows […].[265]

---

[263] […] da kann ich mich schon daran erinnern, dass mir da vorgeworfen wurde, dass ich da denn Bericht in [name of HEI], dass ich den zu sehr geglättet hatte. […] Weiß ich nicht, ob es da noch Formulierungsänderungen großartig gab, aber ich habe nicht den Eindruck, dass mein Bericht da stark verändert worden ist in der Endfassung. Aber sehr wohl gab es dann da von den teilnehmenden Besuchern in [name of HEI], die gesagt haben, das hab ich aber anders in Erinnerung, wie es dort jetzt steht.

[264] [wie] die Gespräche vor Ort gezeigt haben, durch Differenzen der Kollegen untereinander gekennzeichnet. […] Im Selbstreport der [name of HEI] wurden die erheblichen Differenzen zwischen den Mitgliedern der Fakultät nicht angesprochen. Diese Differenzen dürften wesentlich dazu beigetragen haben, dass es dem wissenschaftlichen Personal dieser Fakultät nicht gelungen ist, ein erkennbares, spezifishes Entwicklungsprofil in Forschung und Lehre aufzubauen. Mit dem derzeitigen Personal sind keine diesbezüglichen Verbesserungen zu erwarten. Deshalb sollte der Generationenwechsel genutzt werden, neue Schwerpunkte zu setzen. An den Neubesetzungen der Stellen sollten derzeitige Stelleninhaber nur in dem Umfang beteiligt werden, als es das [name of institutional law] als unabdingbar vorschreibt.

[265] Dass die Leute im Grunde mit einer Aversion kamen, ihr habt nicht die richtigen Kriterien. Ihr wisst gar nicht, worum es hier geht. Ihr betrachtet uns falsch. Ihr messt uns an falschen Maßstäben. Das kam immer so

The entire commission was horrified by this experience and agreed that it was unprecedented. The chair had been publicly insulted and the member from the Ministry of Culture told the representatives from this HEI that it would take years for them to make good their behaviour.[266] (Expert J) The final report described two possible scenarios for this HEI: either closure (which the team wanted, but knew that it would not happen), or the development of a concept to ensure maintenance of minimum standards for a scientific teacher training, which was not the case at the time of the evaluation. The reaction of this HEI can, of course, be regarded as an outlier as the experts were generally treated with respect. However, this example reveals how embittered some situations can become and that under such circumstances rational discourse becomes impossible. A concept of quality management that did not appear in the final report is the notion of self-damage in academia, which was a point raised by one expert in the aftermath of the above scandal.

> Generally, we learnt at the same time that in this institution there is no safeguard against disgracing oneself. There's no institutionally-anchored discussion about criteria, and so it was all related to individuals, to the working unit, to the subject, but also the institution.[267] (Expert J)

This point does not seem to be addressed by the quality debate. There would seem to be a strong argument for introducing some mechanisms within university departments to discourage internal discrimination and ensure that communication channels on quality debates always remain open. Although the academics in this one HEI were 'naming and shaming', they were disgracing themselves in the process. Their HEI had no means of intervening and nor did the Ministry. If the replacement of existing chair holders through time is the only mechanism whereby a move towards improving quality can be achieved, then the system overall would surely seem to be failing.

---

als Abwehrargument latent durch und bei Ulm ja ganz manifest dann am Ende in der Schlussbesprechung. Ich weiß nicht, ob Sie diese Szene schon gehört haben. Das hat ja dann einen richtigen Eklat gegeben, der [Name des Experten] hat sie ja dann vor die Tür geschmissen, und wir haben gedacht, uns laust der Affe […].

[266] Und der Beamte, der aus dem Kultusministerium mit dabei war, hat denen dann auch gleich gesagt: Das dauert Jahre bis ihr das wieder gut machen könnt.

[267] Man hat in der Regel auch gleichzeitig gelernt, dass es in der Institution überhaupt keine institutionelle Vorkehr gegen Selbstblamage gibt. Keine institutionell abgesicherte Diskussion über Kriterien und so war das also auf die Personen bezogen, auf die Arbeitseinheit bezogen, auf das Fach bezogen, aber auch auf die Institution bezogen.

7.9 What the experts learnt from the evaluation

Although there had been some negative experiences, the experts were unanimous in that they felt that their own knowledge of Educational Science in Germany had been enriched by the experience of working in this commission. One of the main reasons specified was that they had engaged in a mutual process of discovering more about their discipline and the situation regarding this discipline beyond the walls of their own HEI. Their horizons were clearly expanded by finding out about the *Lebenswelt* of others. Expert D described this as being something which could not be easily transmitted unless through personal involvement:

> What you can learn from such a process is about dealing with people. You can use that in practice in all contexts. That is something which you can't learn anywhere or teach, that comes from experience and a building block of this experience is being part of such a commission.[268]

Expert M may have experienced the greatest learning curve: 'I found it a new challenge to come into a completely new group and to find out about this kind of evaluation, institutional evaluation. That was new territory for me. I hadn't done something like that before.'[269]

Once inside the commission, the experts discovered to which 'camp' they belonged (predominantly either empiricists or hermeneuticists) and this influenced how the debates on criteria went. They therefore had to address their own values and the *Mitwelten* of those working directly alongside them. Within this they could witness power struggles as the different camps fought for supremacy, although we already know that at this moment in time the empiricists had the upper hand, as this was the emphasis of the Ministry. An institutional head, who had also chaired a number of expert commissions, confirmed that the within-group struggling for supremacy was a frequent phenomenon within evaluation

[268] Was man natürlich aus so einem Verfahren lernen kann, ist etwas über den Umgang mit Menschen. Das kann man an allen Zusammenhängen praktisch anwenden. Das ist etwas, das kann man auch nicht irgendwo lernen oder lehren, das kommt aus der Erfahrung und ein Baustein dieser Erfahrung ist das Mitarbeiten in einer solchen Kommission.

[269] Ich fand das aber eine Herausforderung. In eine ganz neue Gruppe zu kommen, und diese Art von Evaluation, institutioneller Evaluation kennen zu lernen. Das war für mich auch Neuland. Ich hatte so etwas vorher nicht gemacht.

commissions: 'the quality of experts and the personal autonomy and independence is the one situation. The politicising of experts by group dynamicists in the expert commission is something altogether different. It does happen.'[270]

One of the key struggles, of course, was establishing the criteria according to which they would assess the HEIs in Baden-Württemberg. Because this was the first time that such criteria were being established for Educational Science within both teacher training colleges and universities, there was much to be addressed, and much to learn about differing values and opinions on certain aspects. As Expert J told me:

> We discussed really intensively and explicitly which criteria to use and that had to be, in order to progress under the banner of scientific research, evaluation research and to see where open questions remain, because if you take another look at the criteria, we still haven't solved the question of which indicators to use in order to measure things easily.[271]

As the above comment shows, much remained unresolved. This notion is supported by many others in the commission, who acknowledged that their work could not be regarded as conclusive:

> The criteria which we used are to some extent a vision for the future, a developmental perspective. And from that point of view it's OK to say we have established a status quo here, we have been able to make a point. However, we would like a development in this or that direction. The criteria were fair in that sense and I believe that in that sense we tried to formulate the report in the end like that and not in the sense of saying that in the past you failed to do something, but you couldn't have known about that. So, from that aspect we were fair, I think, as fair as we could be. But it wouldn't be fair to measure past work of Educational faculties or educational research according to these criteria as I know of no single educational department which could have stood up to these criteria at that moment in time. An institution such as the Max-Planck-Institute could have stood up to these criteria, but they work under totally different requirements than we do in the universities with a mass subject. […] I think it's good to have something that gives you traction, in your inner discipline and so you can say, okay, that's my viewpoint, but how can we move our discourse on this forwards?[272] (Expert F)

[270] Die Qualität der Gutachter und im Grunde die persönliche Autonomie und Unabhängigkeit ist die eine Situation. Die Politisierung von Gutachtern, also, von Gruppendynamiken in Gutachtergremien ist noch mal etwas anderes. Die gibt es.

[271] Wir [haben] wirklich intensiv und auch explizit diskutiert, was Kriterien sein könnten und das musste halt mal sein, damit man auch unter dem Aspekt von Wissenschaftsforschung, Evaluationsforschung weiter kam und sieht, wo die offenen Fragen sind, denn wenn Sie die Kriterien sich noch mal angucken, die Frage ist ja weitgehend ungelöst, welche Indikatoren man hat, um das einfach zu messen.

[272] Das, was wir als Kriterien angewendet haben, ist eigentlich eine Zukunftsperspektive, also mit der Perspektive der Entwicklung. Und insofern ist es auch in Ordnung, denn man kann sagen, okay,, wir haben hier einen Status quo, wir stellen etwas fest. Wir wünschen aber eine Entwicklung in die und die Richtung. In

Another aspect of the evaluation which everyone learnt something about was the extent of ministerial involvement in Baden-Württemberg which, to some extent we might regard as their *Überwelt*. The experts had to find out how this particular Ministry functioned and to what extent their recommendations would be feasible or not. Those with more experience of working with other ministries guided the less experienced experts through this process, indicating what could or could not work. Expert E tells us more about this aspect:

> Well, it's obvious that the commissioning body wants efficient teacher training. It wants Educational Science to play its part in that, but if this has to go strongly via research it would also mean investing more money, and we went as far as to say, well, if we can't force third-party funding, then at least we can expect the *Land* to do something to get more third-party funding in. The *Land* can do that and this means on the other hand that the commissioning body has to pull more cash out of its wallet and that is probably not what they wanted this evaluation to come up with, but rather for the HEIs to manage this under their own steam. But not much happens as they haven't got the power to do it on their own because the majority of HEIs are given less funding than they need.[273]

It seems fair to say that no-one was under any illusions that the evaluation may not have been a complete success in being able to instigate change, which was the aforesaid intention of the group in a transformative sense. The following statement reveals that the group were well aware that the politics of the day could easily change and then their efforts would be in vain:

> If you look at our recommendations, they are the ones you usually have: institutionalisation of quality assurance, research and creation of main emphases, personnel development, support of young academics, courses, teacher training, resources. Ministries are relatively autonomous in how they deal with recommendations, and especially in cases of Ministries that are fairly self-confident such as in Baden-Württemberg, they will take notice of them, might use them, but, as in the case of many evaluation reports, these will wander into a

---

diesem Sinne waren sie fair, die Kriterien und ich glaube, in diesem Sinne ist am Ende ja auch der Versuch gemacht worden den Bericht zu formulieren und nicht im Sinne von, ihr habt in der Vergangenheit etwas nicht gemacht, was ihr aber gar nicht hättet wissen können. So gesehen waren wir, finde ich, so fair, wie es irgend ging. Aber es wäre nicht fair, an diesen Kriterien, die vergangene Arbeit von erziehungswissenschaftlichen Fakultäten oder erziehungswissenschaftlicher Forschung zu messen, dann würde nämlich... im Grunde genommen würde dann keine, ich kenne keine einzige Erziehungswissenschaftliche Fakultät, die diesen Kriterien hätte standhalten können, zu dem Zeitpunkt. Also standhalten können hätte diesen Kriterien teilweise so eine Einrichtung wie das Max-Planck-Institut. Bloß die arbeiten unter völlig anderen Voraussetzungen als wir in den Universitäten mit einem Massenfach. […]. Ich finde es gut, dass man sich an so etwas reibt, an Innerdisziplin, das man also sagt, okay, da steht jetzt mein Standpunkt und wie entwickeln wir das im Diskurs.

[273] Also, die Auftrag gebende Behörde, das ist klar, die möchte eine effiziente Lehrerbildung, die möchte, dass die Erziehungswissenschaft ihren Beitrag leistet, aber dass das stark über Forschung gehen muss, würde ja auch bedeuten, dann muss ich mehr Geld investieren und wir sind dann ja soweit gegangen, dass wir gesagt haben, ja gut, wir können quasi die Drittmittel nicht zwingen, dann können wir wenigstens vom Land erwarten, dass sie etwas tun, dass mehr Drittmittel kommen. Das kann ja das Land machen und das heißt aber wieder, der Auftraggeber muss tiefer ins Portemonnaie greifen und das war ja vielleicht nicht das, was sie zuerst mit der Evaluation wollten, sondern aus eigenen Kräften. Und da geht dann vieles nicht aus eigenen Kräften, weil die meisten Institutionen waren sicher von der finanziellen Seite her unter dotiert.

> cupboard, in your house, or mine, and will only be partially adopted and then, shortly afterwards, the political or financial-political premises are set differently and then they are devalued because the basis of business is no longer the same.[274] (Expert J)

One of the institutional heads reflecting on the evaluation confirmed that this was indeed the case:

> Well, I'd say that the questions which were posed back then have made way for much more exciting and more current questions: this whole notion of the teacher, professionalisation of school heads and things of that sort, which have come to the fore so much that today, I think, almost everyone would say, that is the situation back then. The catalogue of problems of back then no longer moves us greatly today.[275]

While this may indeed be true, the fact that a discipline is required, at any point in time, to engage in processes of self-reflection which involve the viewpoints of valued peers is an indication that quality assurance cannot be regarded as a static phenomenon and those involved in it need to constantly engage with the possibly changing paradigms which surround them and their own attitudes and values towards these. Although some might agree with the resigned comment of Expert H who said that he couldn't see that the evaluation had encouraged the discipline to make great progress,[276] if we look at what another institutional head said, it would appear that Expert H may have been mistaken:

> To my mind evaluations are always a great opportunity. Even then when you don't agree with the result. To recap: I could have climbed up the wall at the things written in here. But that's not the point. The point is that I now have occasion to work through it all. I can even say why I don't support this result, and I have to explain my reasoning. And that is a quality in its own right and otherwise you wouldn't do it like this, not even be able to, because there wouldn't be a reason to do so. Well, one should always see it, it's not always about saying that that which the experts say has to be true. Absolutely not. Rather: on the basis of the report you have to create a sensible truth for yourself, which is

[274] Denn wenn Sie unsere Empfehlungen angucken, das sind die üblichen, die man hat. Institutionalisieren der Qualitätssicherung, Forschung und Schwerpunktbildung, Personalentwicklung, Nachwuchsförderung, Stu-diengänge, Lehrerbildung, Ausstattung. Die Ministerien sind doch relativ autonom in der Umsetzung und gerade wenn es Ministerien sind, mit einem relativ großen Selbstbewusstsein, wie in Baden-Württemberg, dann nehmen die das zwar zur Kenntnis, können das benutzen, aber wie viele Evaluationsberichte wandern in den Schrank, bei ihnen oder bei mir und werden nur teilweise genommen und kurz danach werden politische Prämissen oder finanzpolitische Prämissen anders gesetzt und dann ist das schon entwertet, weil die Geschäftsgrundlage nicht mehr stimmt.

[275] Also, ich würde sagen, die Fragen, die damals gestellt sind, sind inzwischen für viel spannendere, hoch-aktuellere Fragen, also, dieses ganze Lehrerbild, Professionalisierung von Schulleitern und solchen Dingen, die sind im Moment so vordringend, dass heute, glaube ich, fast alle sagen würden, das ist der Stand eben von damals. Der Problembestand von damals und nicht mehr das, was uns heute wirklich bewegt.

[276] Den großen Fortschritt hat man auch nicht gesehen aus dieser Maßnahme heraus.

> directional for that which you do. […] If I know from the outset that I don't have to change anything and stick by that, then it would be meaningless.[277]

7.10 Summary

This chapter has described some of the diverse situations which the evaluators encountered, the difficulties and challenges they faced, and their attitudes towards these. The evaluation provided a range of complex issues for them to deal with, in particular regarding the contrasting paradigms apparent in this discipline and some of the problems of the methodology of the evaluation, e.g. amount of literature to digest, student numbers in meetings. They also encountered a host of different stakeholders, each of which had different interests. Furthermore, they were challenged by their own personal constellation (*Eigenwelt*), bringing to the evaluation varying levels of subject knowledge and evaluation expertise to discuss with the others (*Mitwelt*).

In my final and penultimate chapter I would therefore like to review the key issues raised by this evaluation, and make suggestions for how evaluations within the context of higher education, and German Educational Science in particular, might look like in the future.

[277] Also, noch einmal, für meine Begriffe sind Evaluationen immer eine große Chance. Selbst dann, wenn man nicht mit dem Urteil einig geht. Also noch einmal, also, ich hätte mich zum Teil grün und blau ärgern können, über das, was da drin steht. Das ist aber überhaupt nicht der Punkt. Der Punkt ist, dass ich jetzt einen Anlass habe, mich daran abzuarbeiten. Ich kann ja auch sagen: Warum teile ich dieses Urteil überhaupt nicht, und ich muss es erläutern. Und das ist eine Qualität für sich und man würde das sonst in dieser Art und Weise nicht machen, nicht einmal machen können, weil es keinen Anlass gibt, das zu tun. Also, das sollte man immer sehen, es geht nicht immer darum, zu sagen, also das, was die Gutachter sagen, das muss so wahr sein. Überhaupt nicht. Sondern man muss auf der Grundlage eines Gutachtens zu einer sinnvollen Wahrheit für sich selber kommen, die dann handlungsleitend ist, für das, was man tut. […] Es hätte nur dann keinen Sinn, wenn ich am Anfang schon weiß, dass man nichts ändern muss und das dann durchhalten würde. Dann wäre es sinnlos.

CONCLUSION

Chapter 8

Review of the evaluation, critical discussion and conclusions

This chapter starts with a brief vignette which is based on what I have been told by the evaluation's organiser and various members of the commission. (Some parts of what follows have been adapted from Harris-Huemmert, 2008)

> On a cold grey morning in November a group of five people dressed warmly in winter coats waited patiently at the bus stop outside their hotel. They would be taken to the local university, where they were meeting the university's chancellor, members of the Education faculty and groups of students. "You know," quipped one of the party, "if this bus has a crash and we all get killed, then Germany's going to lose the elite of its Educational Science in one go". "It's funny you should say that," said one of the others, smiling wryly, "but I'd had exactly the same thought when I saw the list of who was going to be on this commission. Well, let's hope the bus driver knows what he's doing then…"
>
> The bus arrived and they were grateful to step out of the biting wind and snow flurries that surrounded them. Another day of talks and seemingly endless meetings lay ahead of them, added to which came the pressure of sitting down to write an immediate report which summarised what the group had discovered that day. They were aware of the burden their job carried, but remained curious at what each new institution would reveal and were stimulated by what lay ahead of them…

This vignette gives us some notion of what it is like to be involved in evaluation work, much of which involves uncertainty, questioning, revision and, hopefully, resolution. When I started on my research, I posed three general questions (see also Chapter 1):

> What kind of *experience* do evaluators have when they engage in large-scale evaluations of higher education?
>
> What *attitudes* towards evaluation and higher education quality analysis do evaluators have and do these change as work progresses?
>
> What kinds of *difficulties* do evaluators encounter and what strategies do they develop to deal with these?

In the case study, the experts introduced what Harman (1998, p.354) describes as 'outside values and constructs' into their discussions, depending on their personal stance. Indeed, it

seemed that the evaluators were never entirely free of bias, as they were unable to separate their own value systems from their arguments. As Lee (2006, p.168) has pointed out: 'true neutrality is a practical impossibility'. This is not a problem if the evaluation has enough members with differing views to compensate for any bias present, as in this case study, where differences of opinion were manifest. The final report stated that where differences of opinion were present, decisions were reached by consensus (*Gutachterkommission*, 2004, p.8) even if this meant that a substantial group may not have been in agreement. What follows are some of the answers to the above questions and further considerations with regard to the overall evaluation and suggestions for ways of improving the evaluation experience.

8.1 Ministerial contract, impact on evaluation and appointed experts

In 2003 the Ministry of Science, Research and Art in Baden-Württemberg called for a structural evaluation of Educational Science to assess whether the discipline was in alignment with current and future requirements in research, teaching and teacher education (Chapter 4). Experts were asked to present information on the discipline's strengths and weaknesses. They had to assess its capabilities with regard to international developments. Furthermore, they were asked to make recommendations on its structure, subject-specific development and the foci of interest in research, teaching and further education, making particular reference to teacher training and empirical educational research. The evaluation involved 15 HEIs. Eighteen experts were invited to take part. The commission comprised six women and 13 men, one of whom was a ministerial representative. Four members were foreigners and one was a non-specialist in Law.

There was considerable weight behind those participating. The implication was that institutions would possibly take any recommendations that the evaluation report made more seriously as those involved were highly regarded. According to the principle that expertise is necessary 'in order to have a sufficiently deep understanding of how any object of evaluation functions, and how it interacts with the context in which it exists and

operates' (Lee, 2006, p.162), the members were chosen because they represented different areas of the discipline. One of the ministerial representatives told me: 'In this case we assumed that reputable subject experts would be more likely to be accepted'.[278]

As the Ministry of Science, Research and Art was the commissioning body, the immediate question arose as to the extent of its influence as could be documented by the difficulties Röbbecke had convincing the HEIs in advance that the evaluation was trustworthy and that financial cuts would not be made. Documents predating the evaluation indicate that the Ministry already had a very detailed perception of what the institutions were doing and where deficits in their management, research interests, and output lay, although the Ministry was relying on the evaluation for the presentation of 'hard' evidence that could therefore not be refuted (Chapter 4). In addition to the state's perceived poor results in PISA, the Ministry therefore placed emphasis on raising the standard of its teachers, which included their training in Educational Science alongside subject-specific pedagogies. Furthermore, it declared its interest in empirical educational research and specifically appointed an expert in this field into the important sub-committee which was particularly crucial in setting up the commission and specifying the direction it should take.

The sub-committee held meetings with the Ministry in advance of the first plenary meeting. During these meetings, the criteria for evaluation were discussed and the foundations of the ensuing work were established. I have already stated that of the 18 experts six were unable to attend the first plenary meeting (Chapter 5). One of the experts elected to the commission was later obliged to leave on the grounds of conflicting engagements and did not take part in any onsite visits. A new expert was elected before the onsite visit stage to provide support and expertise in the area of special education as the commission felt that it was not sufficiently weighted in this field. The fact that not all members were present from the outset of the commission meant that their opinions were not heard or taken into consideration in the establishment of evaluation criteria. Arguably,

[278] In diesem Fall hier sind wir schon davon ausgegangen, dass fachlich renommierte Leute eine höhere Akzeptanz haben.

the direction therefore might have been somewhat different, although this is speculative, but based on what was said during interviews.

8.2 Connectivity of experts and differing paradigms

The evaluation of Educational Science 2003/2004 was conducted at a time when evaluations in Germany were still not especially well established in higher education. External peer review processes were not yet commonplace, and there was considerable suspicion among those being investigated. As we have seen from the present study, although most of the experts were experienced both in their fields and in university administration, different levels of status could be identified, as some areas of their expertise were mainstream, whereas others were clearly on the periphery of the discipline. According to the previous networking among individuals, we have argued that those who were well-connected *and* in the main disciplinary areas were more likely to have influence within the commission and be more involved in the main decision-making processes. In other words, *Eigenwelt* was not always equal, hence the *Mitwelt* they experienced during their mutual evaluative work was also not on an equal par.

Furthermore, due to the differing paradigms within the discipline (empiricism and hermeneutics) we could witness a conflict of interest. On this occasion and in particular because of the Ministry's involvement, we saw the dominance of empiricism, although it remains to be seen whether this will remain the dominant paradigm within Educational Science in the future. Arguably, this will shift again. The interviews, which were held a few years after the evaluation had been concluded, would seem to indicate as much. Many experts commented that in retrospect they felt that they could (and should) have gone further in being possibly more brave in the establishment of evaluation criteria for their discipline, which at this point were still dominated by existing university-based criteria. The criteria for Educational Science may therefore need to be treated in a possibly different manner than those in other disciplines, allowing for more fairness with regard to the quality of regional research results, for example, which are not necessarily of any wider

interest or impact to an international audience. If we expand this point, the *Lebenswelt* of the teacher training colleges was unfamiliar to the experts who were almost all university-based. Although obviously aware of this fact, they nevertheless applied their hitherto known culture from the university into this slightly different context. As the evaluation stood, they did not provide a separate list of criteria for the teacher training colleges.

8.3 Management of literature and report-writing

The experts had to work extremely hard to complete the evaluation in the time given, which took longer to conclude than expected. The experts also stated at the start of their work that they could not examine the quality of teaching in any great depth, as this would be far beyond that which they could manage. The amount of literature which they had to engage with before going onsite was immense and overwhelming, and we have discovered that there were different tactics for overcoming this (Chapter 7). Some evaluators read in depth the five self reports of those HEIs which they knew they would be visiting, and skim-read the others. A different tactic was to pose certain questions and see if these were covered. Others again attempted to read everything. A detailed understanding of all the institutions by all participants was therefore not realised and regretted, although the experts accepted that those with sufficient knowledge had enough expertise to provide the others with detailed information and a consensus of opinion about standards and quality assurance could be reached. They did not deny, however, that some of the onsite group constellations might be 'tougher' than others and more rigorous as a result, which again raises the issue of comparability of evaluative judgement among the different groups. Furthermore, it could be argued that extensive self reports are unnecessary, overly time-consuming for the HEIs which have to compile them, and are ultimately not even read (in their entirety) by those called upon to do the evaluation.

Regarding their own reports for the evaluation, it was a challenge to combine reports on 15 HEIs written by many different authors into a style, format and manner that was homogenous. I was not able to see all of the individual drafts, so my study concentrated on

the main reports, which went through numerous revisions and drafts (Chapter 7). We can conclude that because the reports were gathered together by *evalag* and then edited by Baumert and Tenorth in particular, alongside Röbbecke, the final style of the report can be said to be theirs, although in consensus with the remaining experts who were involved in the revisions.

8.4 Statistical problems

The institutions had different methods for organising their data and statistics, which led to gaps in the self reports (Chapter 7). This was a problem also identified by Mittag, Bornmann & Daniel (2003) in their analysis of other evaluation agencies in Germany. Although the data were regarded as an essential part of what was needed to determine standards, the experts did not appear unduly critical of the lack of uniform methods in acquiring these data for quality assurance purposes. Most agreed that if gaps could not be closed, in spite of *evalag's* best efforts to acquire the information, the commission would have to accept this fact. None, however, seemed unduly worried by this deficit, which indicates possible lack in rigour in their own evaluation methods.

This also raises the issue of how institutions manage their statistical information. In consideration of comparative quality assurance, which was the case in this evaluation to some extent, it seems important that the statistics provided be comprehensive and complete. Furthermore, there is an argument for institutions undergoing external evaluations to have similar databanks in which data are gathered according to the same criteria. It would appear from the experience of this commission that institutions in Baden-Württemberg do not have uniform methods for acquiring and maintaining statistical information which therefore makes external evaluation of university statistics rather problematic. Although each institution is autonomous, it might be useful to suggest agreement among institutions which would go some way to enhancing their quality assurance mechanisms for external evaluations in the future. It is a time-consuming and unnecessary activity for external evaluators to have to chase up missing data.

8.5 Experiences gained onsite and strategies for dealing with specific situations

The commission members were almost all unanimous that onsite visits were needed to promote a better understanding of the quality of each institution in terms of courses, research, organisation of staff, facilities, etc. Meetings with specific groups were of key interest and these were not problem-free and usually accompanied by a noticeable element of tension on both sides: how the institutions would be perceived (i.e. judged), and how the experts would be accepted. HEIs had been asked, for example, to provide student representatives for discussions. However, this was managed quite differently across institutions with some posting up boards advertising the evaluator visits, leading to large numbers of students in attendance, while at the opposite end of the spectrum it seems that students had been hand-picked. The evaluators were therefore unclear about which scenario they would face when they arrived at any given institution and whether the voices they heard were indeed representative of the wider student body. They were aware that the student voice could also be dependent on the current experience, and so care was needed in interpretation.

On some occasions we know that the experts were witness to some intense battles among staff members and it was apparent that even an external evaluation commission would have little effect in countering ongoing animosities. The evaluators' sense of embarrassment was coupled with a deep sense of frustration that their peers could not go about their academic business in a more civilised manner. One strategy they immediately devised when they had to report back on their visit at the end of a visit was to put forward one of the foreign experts as spokesman. This diffused tension, to some extent, and increased the distance between the German experts and their sparring peers. It may be very difficult to encourage a member of staff to move to another institution if the personal chemistry within a department has severely deteriorated. In terms of providing 'good' or 'excellent' quality within departments there should be some mechanisms available to provide mediation if personal relations have taken a turn for the worse.

Due to the large size of this evaluation, it was impossible for all evaluators to visit every institution, another reason for the large number of experts. The itineraries for onsite visits were tight and organised in such a manner to allow for the greatest number of visits in geographically close locations, where this was possible. The onsite visit plans reveal that the schedule was demanding because large distances had to be covered, with a different institution to be examined on each consecutive day (e.g. from 20 to 24 October 2003 the experts did a round trip of approximately 450 miles (730km) from Mannheim, via Hohenheim, Weingarten, Constance, and Freiburg. It is debatable whether or not a schedule of this kind is entirely conducive to attentive evaluation work, as the effort required to listen to each separate institution is immense, challenging, and draining. This was possibly compensated for, however, by the number of experts present (generally six). Evaluative work did not conclude with the visit, as informal discussions were held during dinners at the end of the day, which provided an opportunity for the experts to exchange notes on their opinions in a more relaxed environment.

8.6 Exploiting particular expertise

If we turn to the role of individuals within our commission we could identify that the role of the foreign experts was a positive factor (Chapter 5). However, it could be argued that the German experts might have encouraged the foreigners to speak out more frequently in a comparative sense, thereby helping the foreigners to lose their possible inhibitions towards their host country. Foreign experts are engaged in evaluations precisely because they may have other experiences to offer and can compare these with those in the host country. However, it is a social finesse to be able to speak out without seeming authoritarian and didactic. This might prohibit the full extent of a foreigner's experience from being given.

8.7 Problems of size

The evaluation involved a large number of experts (18), which some described as being too unwieldy. Both in terms of expenditure and the organisation required, any large evaluation

commission of this kind is a 'cumbersome beast'. However, the experts could offer no proposals as to how the size of the commission could be altered (i.e. reduced) as they supported the idea that a representative from each of the specialist areas should be present, which automatically meant a large number of experts. Logistically, however, it was almost impossible for everyone to attend every meeting. Those who were unable to attend the first plenary meeting indicated that had their views been known, the criteria for the whole evaluation might have been somewhat different (Chapter 5). The evaluation was ambitious, as it was establishing evaluation criteria which should be applied to research, teaching and learning in Educational Science. However, as has been previously indicated, these standards might have been weighted somewhat differently had all the experts been involved from the first meeting.

The majority of experts felt that the commission's size was almost too large. Plenary meetings had to proceed through a number of points and perhaps the minutiae could not be discussed in that situation. However, the smaller onsite groups did allow for more discourse. The experts could not offer an alternative methodology though, given the number of institutions that needed examining. Most agreed that onsite visits were an essential part of the evaluation and could not be replaced with any other method, although one very experienced expert offered an alternative by recommending that the institutions could elect representatives and then come to the *evalag* agency instead of the experts visiting the institutions. He cited expenditure as one main reason for his argument. Nonetheless, the majority regarded onsite visits to be indispensable for their understanding of the reality within each institution:

> The thing that is important to me above all is these site visits, the onsite discussions, where you can ask about specifics with the result that something looks quite different all of a sudden. Only then can you form an opinion, which can be substantiated in talks with the other group members, or need to be revised.[279] (Expert D)

[279] Und was vor allen Dingen für mich sehr wichtig ist, sind diese Site Visits, sind die Gespräche vor Ort, bei denen man einzelne Dinge abfragen kann mit der Folge, dass plötzlich manches ganz anders aussieht. Erst danach kann man sich dann ein erstes Urteil bilden, das man in der Diskussion mit den anderen Mitgliedern der Gruppe noch erhärten oder auch revidieren muss.

8.8 Attitudes to selection, evaluation experience and status

Evaluators of higher education pass judgements on the institutions they visit. Those in the present case study were sensitive to their position as peer judges. Their overall approach to the evaluation was one of providing support to their colleagues, suggesting solutions if they discovered a problem. They did not see themselves as acting punitively and were anxious to word their findings in a conciliatory manner, although we have identified that the discourse in meetings was frequently more polemic and outspoken. It could be argued that more polemic reporting could do a discipline more good. This has to be weighed, however, against the point that the experts might themselves be evaluated later on by those they were now judging. So, there may be an element of self-protectionism here which prevents them from going beyond a safe limit.

The majority of experts were fairly ambivalent about their selection into the commission, although those with previous evaluative experience indicated that because of their activities, their selection did not come as a surprise (Chapter 5). Those with more specialised areas of knowledge believed that they were chosen to fill this gap in expertise. They did not necessarily think that they were required to have a broader vision of the entire field of Educational Science, as has been indicated in interviews with *evalag* representatives when they discussed their criteria for selection. It has been shown that those with more experience of evaluation and more numerous contacts were more likely to be of greater influence in the commission. Equally, we saw the ongoing discussion among the main 'well-connected' members with regard to the two main paradigms within the field, which others viewed as a typical fight for status within the group (Chapter 5). Arguably, although input was given by those in the peripheral areas, the majority of decision-making was therefore made by the key figures. Although one might argue in favour of a smaller commission, due to the number of HEIs under inspection, the logistics for each individual would have needed to be increased and it is debatable whether any of the experts present would have been able to conduct more onsite visits.

Those with previous evaluation experience had an attitude which reflected their status of 'experienced evaluator'; in other words they saw themselves as the experts and therefore did not question the fact that they had again been called upon to work on an expert committee. What is perhaps more interesting in this context is that those members who had not worked as an evaluator before did not analyze their selection more carefully. Most thought that it was their specific subject knowledge that had marked them out for selection, but they did not seem to engage in any discussion of this point with either the agency representatives who had made contact with them, with peers in their institutions, or with other experts on the committee at a later stage. Indeed, most participants were surprised to be asked about why they were selected and the question stimulated reflection on this for the first time. There was also a marked difference in attitude between those with more evaluative experience and those with little or none in that the former expressed a subtle form of arrogance implicit in their answers by listing the high-ranking organisations that they had previously been involved with in an evaluative capacity. Those with no previous evaluation experience were more humble in their statements regarding selection, although they, too, were classified as experts in their fields and could therefore stand shoulder to shoulder with those with more evaluation experience.

Without exception the evaluators were concerned about the HEIs they were investigating. They were curious to see the teacher training colleges and find out more about the modus operandi there. They showed sympathy to their colleagues, and were aware of the financial problems which some were encountering. On occasion they were frustrated to see that some colleagues in the HEIs were unable to engage in normal discourse, and this combined with feelings of incredulity, despair and annoyance, too.

Most evaluators stated that they were driven by a curiosity to see how the institutions within Baden-Württemberg were doing. Although each were only going to visit a handful of institutions, they believed they would learn much about what was happening elsewhere and would therefore exponentially increase their knowledge of the status quo of Edu-

cational Science in one of Germany's largest states. Seen historically, the evaluators were aware that this commission marked a first by conducting such a large cross-sectional analysis of all institutions and they believed they were engaging in a ground-breaking evaluation which was establishing criteria for quality standards in Educational Science in other institutions and states. Overall, the intention was supportive: 'From that point of view my primary motive – I didn't have any illusions about the political effectiveness of it – my intention was to somehow help the subject and its colleagues.[280] (Expert H)

8.9 Choice of expert and impact on evaluation

German institutions of higher education are characterised by marked and complex hierarchical structures (Lohnert & Rolfes, 1997, p.9), and so it seems valid to ask if this was evident within the body of experts, coming as the majority did, from 'within the ranks'. Depending on the experience of each individual in terms of his own evaluation practice, work experience, and publication output, a hierarchical structure did indeed seem to crystallize, albeit mostly subconsciously. The experts admitted that it also depended on the subject areas from which they came, as some parts of the discipline had more kudos than others (empirical educational research being at this point higher in status than special needs education, for example). Those experts with greater evaluation experience were also more likely to take the lead in discussions. When asked about whether they felt they had enough voice within the group, everyone replied in the affirmative, although some members appeared to have more 'voice' than others.

Reference has already been made above to the fact that one of those in the sub-committee was a predominantly empirical educational researcher, which he regarded as the 'gold standard'. At the time of the evaluation, this research paradigm impacted on how the experts saw the discipline, e.g. those who were more hermeneutically oriented (which had been the dominant research paradigm in the past) had less influence than the empiricists.

[280] Von daher war mein erstes Motiv- also über die politische Wirksamkeit habe ich mir keine Illusionen gemacht, mein Ziel war, dass ich einfach dem Fach und den Kolleginnen und Kollegen irgendwie würde helfen können.

Nonetheless, they acknowledged the ministry's emphasis on determining how much educational empirical work was being conducted.

8.9.1 Evaluator types

In Chapter 2 I referred briefly to the literature on evaluator types, which have already been described. Linking what I have discovered about the evaluators in this particular case study and their previous networking which we could identify at the start, I would like to suggest the following loose categories, which seem to have developed as the evaluation progressed.

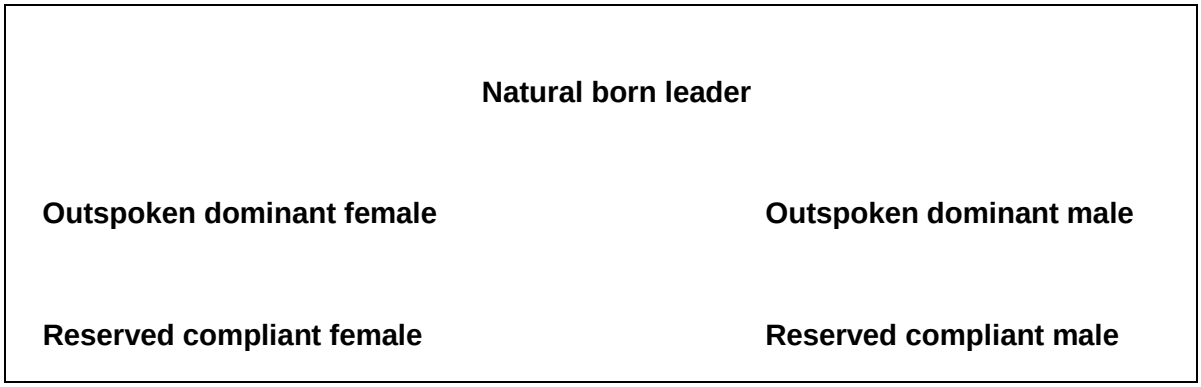


Figure 3: Evaluator types

A 'Natural born leader' type is immediately identified as such by all members. He may not lead the commission, but he is the one everyone turns to for advice. His input is therefore crucial on the outcome of the evaluation. The 'Outspoken dominant female' is well-networked and will voice her opinion freely. She stands her ground against her predominantly male colleagues. The 'Outspoken dominant male' may not be at the head of a high status sub-area within his discipline. Nonetheless, he, too, is well-connected and provides a considerable amount of input into the proceedings. The categories 'Reserved compliant female' and 'Reserved compliant male' are not as well-connected from the start; they may also not have as much evaluation experience to offer. Hence they will be reluctant to voice their opinions on areas that are outside their main fields. They could also be regarded as the learners within the commission and their input will be less extensive. If we consider these types against the fact that all of the members within the commission were at the top of their relevant areas within Educational Science, then we do see a

hierarchical structure emerging. It might therefore be argued that the extent of input could be estimated from the start of a commission according to networking, gender and sub-disciplinary status. This suggestion would need to be verified by examining other disciplines.

8.10 Evaluatory impact and perception of experts

The decision to ask high-ranking experts on to the commission was thought to increase the impact of the evaluation, in terms of institutions taking it seriously and engaging with systems of quality assurance and enhancement. The sample of interviews conducted with institutional heads was small, but covered both universities and teacher training colleges. Here we see interesting structures of behaviour emerging. The PH Weingarten, for example, had just appointed a new rector in 2002. It is customary for institutions to turn a blind eye to the activities of any incoming heads, who are allowed a one hundred day period of grace. The new rector pushed for a whole institution evaluation, which was unheard of in German higher education up to this point. As his institution was perceived as being 'at risk', he decided to launch headfirst into evaluation so that he would be prepared in advance of any externally commissioned evaluation. Fortunately, the faculty of Educational Science had also seen the need for restructuring. Many of the problems which might have come to light during the evaluation had been solved beforehand (e.g. removal of one chair in favour of new lectureships and an additional half-post). In this example we could see that the head was a positive factor and he took the evaluation as a welcome opportunity for reform. With regard to how the other institutional heads saw the evaluation there were conflicting messages, and in some cases the evaluation was seen quite negatively. It could be argued, then, that some heads engage more with the disciplines under examination, whereas others are perhaps somewhat less concerned. I have already indicated that the institutional leadership needs to express a desire for quality improvement and possible reform if the results of an evaluation are to be implemented. If the HEI leadership is disinclined to listen to results, then an evaluation will fall on fallow ground.

8.11 Conclusions

This research has examined the attitudes and experiences of a particular group of experts and revealed some of the difficulties they faced during their evaluation of Educational Science in Baden-Württemberg 2003/2004. We have been able to follow the path of the evaluation by examining the documentary evidence available to us in the archives of the *evalag* agency (thereby revealing much of the *Lebenswelt*), and we were also fortunate to examine all of the personal views (*Eigenwelt*) in greater detail, e.g. the diary of one expert. This study has presented the 'messy' side of evaluation work as it has not attempted to disguise the problematic nature of decision-making along the way, the multiple re-addressing of the selection of evaluation criteria and the disciplinary identities which play a part in who takes on a more decisive role. We have seen how the culture of German higher education (*Überwelt*) and the politics of a particular time impacted on the manner in which experts reached their conclusions on evaluation criteria. We have also seen the juxtaposition of each individual world (*Eigenwelt*) with those of each other expert (*Mitwelt*).

Overall, it seems fair to say that the higher education evaluation landscape in Germany will remain far from homogenous owing to the present federal system. Standards are available; however, it is up to each evaluation agency and the external experts as to how far these are employed. This again is influenced by prevailing research paradigms, and we have seen that the hermeneutic research paradigm, which had for so long been the dominant force in German Educational Science, had to make way for an empirical research paradigm. Evaluations also depend on the engagement of the experts employed as to whether they familiarise themselves with these standards before embarking on evaluation work. To date, the standards do not specify how experts should be recruited or which 'soft' skills they should ideally possess. Discussions with employees from *evalag* and the experts themselves when asked which qualities they believed they should have as evaluators, indicated several areas of importance.

Hard skills	Soft skills
In-depth subject knowledge	Courage
International orientation	Flexibility
In-depth knowledge of the discipline as a whole	Ability to work quickly under pressure
High profile publication output	Tact
Experience of university management	Empathy
Evaluation experience	Dedication to quality enhancement

Table 15: Skills of evaluators

Although the evaluators in the case study endeavoured to conduct a sound and fair evaluation, owing to the relative newness of quality assurance methods and their implementation in German higher education at this time it could be argued that evaluators were only 'fit for purpose' to a certain degree, as this concept is still being developed within the field. If being fit for purpose requires evaluators to have in-depth knowledge of the field, experience of leadership in higher education, reputation, and strong publishing credentials, then the evaluators in this case study certainly all qualified. Being fit for purpose to conduct evaluation work could arguably be expanded to include knowledge of evaluation techniques and encompass being provided with evaluation training by the evaluation agencies according to the particular standards in use, in addition to possessing the other 'soft' skills as described above, such as tact and sensitivity towards the needs of the stakeholders.

The majority of experts in the case study had experience of working as evaluators, which could be regarded as beneficial as they had a preconception of what was being asked of them. There is a danger, however, that people who accept evaluation tasks could lead a discipline into a certain direction, to the detriment of more marginalized groups. As one expert said: 'It was not the case that this was an objective, completely objective procedure. It was shaped by the profiles of those who were doing it.'[281] (Expert E) There are numerous influences at play on all of the groups involved in an evaluation (Mertens, 1998).

[281] Also, es ist überhaupt nicht so, dass das eine irgendwie objektive, völlig objektives Verfahren wäre, sondern das wird geprägt, es gibt Profil von den Leuten, die das machen.

Even the power constellation within the expert group can have an influence on the outcome. This can be combated in part by bringing in more marginalized groups within a discipline, which was so in this case study, in which representatives from special needs education or vocational education participated. However, due to the powerful influence of the empirical educational camp, it appears that at least some of the smaller groups were marginalized. Experts, therefore, need to be highly self-reflective and separate themselves from their own 'camp', if they wish to avoid some of the problems as highlighted above (Schön, 1993). This surely remains one of the main challenges that evaluators face.

8.12 Limitations of study

This research has been broadly concerned with the process of evaluation within a particular context. It has illuminated a specific evaluation in a specific context from a number of angles. We have gained some insight into how experts reached their decisions and made judgements, and have learnt about some of the aspects that may stand in the way of voicing concerns, or documenting these. What this study does not do, as this would form the basis of another piece of research, is to examine the outcomes of the evaluation and discuss its effect on the HEIs under investigation, although I do make reference to some of the points contained in the final report. There is clearly scope for more research into the effectiveness of evaluations of higher education, as it has been argued that evaluations do not necessarily bring about great change (Patton, 1997).

8.13 Implications for policy and future research

This study could only point at some of the issues that arose in one, albeit large evaluation conducted in German higher education. Although the evaluation agency in this case provided an immense amount of support and the experts were highly satisfied with the work of the evaluation's main organiser in particular, one suggestion is that evaluation agencies provide far more induction for experts. This could help resolve many of the problems that we could identify in this evaluation, particularly with regard to the establishment of the evaluation criteria and the *Referenzrahmen* for assessment. However,

this may be difficult to achieve as those conducting evaluations are usually extremely busy in their respective posts. A more extensive induction would nonetheless be a means of examining the culture of the discipline under investigation, and it would possibly give the experts more time to engage in critical self-reflection on their own values before they start their evaluative work.

Although this study has provided insight into the evaluation process within the particular field of Educational Science in Germany, it might also be useful to examine evaluations of different disciplines in a similar manner, to investigate if the experts involved in those have a similar experience, or whether their attitudes are different depending on their disciplinary background. It might be the case that experts in the natural sciences, for example, might perceive of evaluation differently. If comparable analyses were provided, then it might be possible to provide a manual of self-reflection, which would prime experts to the kinds of self-analyses that they need to be aware of during evaluation work in their particular academic environment. Although numerous evaluation manuals exist, which do provide help and guidance, perhaps we need to start thinking more about differentiated evaluation guides that take into account more of the culture under investigation. This, in turn, would raise self-awareness of possible bias and thereby lead to even more open and fair quality debate.

POSTSCRIPT

Chapter 9

Developments in German HE quality assurance since 2004

In this final and very brief chapter I would like to highlight some of the events that have occurred in Germany since the evaluation of Educational Science was concluded in 2004. It is probably true to say that evaluations of the kind conducted in German higher education at the start of the 21st century will no longer take place. They have been identified as being too unwieldy, costly, and to some extent too all-encompassing. While external review has become an accepted instrument in quality assurance, the context in which this occurs has changed quite markedly.

9.1 Introduction of evaluation standards and the 'Exzellenzinitiative'

With the addition of university rankings by journals such as *Spiegel, Stern* or *Focus* and organisations such as the CHE (Center for Higher Education Development) or the DeGEval (German Society for Evaluations) [http://www.che.de and http://www.degeval.de], German universities have been forced to address their standards and what they can offer, not only on a local dimension, but against the wider global academic community. In 2004, during Chancellor Gerhard Schröder's term of office (1998-2005), the question of establishing elite universities in Germany comparable to those in the US Ivy League and Oxford or Cambridge was raised. This immediately caused concern in the ranks of his Social Democratic Party - SDP, a party predominantly given to more egalitarian policies. The debate led nonetheless to the introduction of a competition for funding, the so-called Excellence Initiative (*Exzellenzinitiative*) through the DFG, which has encouraged clusters of research excellence to emerge.[282] Some of those universities expected to be winners of the first wave failed to do so, notably the Humboldt University in Berlin as one major example. However, as universities become more familiar with the competition process, it is

[282] See http://www.bmbf.de/pub/Exzellenzinitiative.pdf for results of the first competition.

expected that more will be able to be given funding in future. The competition is scheduled to run until 2011 with results of each wave published in October of each year.

The university landscape in its entirety has changed quite considerably as a result of the *Exzellenzinitiative*. Some universities have decided to close some of their smaller, possibly less successful departments and now concentrate on those which are doing well. We can no longer see the comprehensive cross-section of all disciplines represented in most universities, as once was the case. Although some of the larger universities still offer this breadth of disciplines, others are becoming far more specialised.

The Bologna Process has also had a huge impact. Although the original intention was to see a three-tier degree structure throughout Europe (BA – MA – Doctorate), this has caused rather large organisational problems in the German higher education sector as HEIs had to decide what should constitute a Mathematics BA programme, to name but one example. As each federal state has slightly different ideas on the subject and each HEI is effectively able to create its own course programmes, we see a diverse range of BAs throughout Germany which, although modular in character, do not necessarily offer the same parts. This arguably makes it harder for students to transfer between institutions rather than easier, which was one of the original notions behind Bologna. Furthermore, each new degree requires official accreditation before it can be implemented, and we can identify a backlog in the accreditation process, although once again, all new degrees were originally scheduled to be accredited by 2010.

Another development in the year following the conclusion of the *Erziehungswissenschaft* evaluation in 2004 was the adoption of the *Standards and Guidelines for Quality Assurance in the European Higher Education Area* (ESG). The standards put forward in Bergen (2005) were incorporated into those of the German *Akkreditierungsrat*. As Hopbach (2006, p.236) states, the 'ESG contain standards that are intended to apply to quality assurance processes in the European Higher Education Area (EHEA). They are divided into three parts:

internal quality assurance within HEIs

external quality assurance of HEIs

external quality assurance of quality assurance agencies.

Hopbach believes that one of the results of the ESG will be that Europe's HEIs will be able to establish a mutual 'language' on quality assurance enabling a discourse that addresses the same principles.

9.2 Organising quality assurance in HE

Recent developments in quality assurance systems in German higher education indicate that much is being done to improve standards. Many universities are introducing their own systems of quality assurance. However, in some cases this is a small operation run by a few people who are given the daunting task of implementing university-wide systems of quality assurance.[283] Arguably, unless they have expertise in quality assurance or have backgrounds in management and organisational decision-making, their knowledge of the diverse academic cultures present within their institutions, in addition to their lack of knowledge of decision-making theories, could lead to situations in which quality assurance is only addressed on a fairly superficial level, although lip service is being paid to maintaining and enhancing quality. For reasons of convenience, a one-size-fits-all system might be introduced which does not pay tribute to the diversity of tasks present or the different cultures and needs within different departments. It would seem more appropriate therefore to think in terms of accommodating quality assurance mechanisms to the respective cultures present. This does not mean, however, that the methods which are used to present data on various aspects of the universities (e.g. student intake, leavers, gender, length of study, publications in reputable journals) should not be standardised for all. Indeed, the present study indicates that far more needs to be done both on a local and a

[283] Following a recommendation by the Accreditation Council (*Akkreditierungsrat*) on 15.06.07, the KMK announced that this had been accepted, thereby allowing HEIs to set up their own internal quality assurance systems (*Systemakkredierung*) on 15.06.07. See also: http://www.kmk.org/fileadmin/veroffentlichungen_beschluesse/2007/2007_06_15-Grundsatzbeschluss-Systemakkreditierung.pdf

federal German level to encourage data compatability to enable cross-comparisons that are accurate and compiled according to the same methods.

Accreditation of degree programmes is a costly and time-consuming factor which, to some extent, has meant that HEIs have found it difficult to address other aspects of quality assurance such as external evaluation simultaneously. Following suggestions by the *Akkreditierungsrat* and the KMK, universities are now able to create their own university-internal quality assurance schemes: 'system accreditation' (*Systemakkreditierung)*. The notion here is that an HEI sets up its own internal quality assurance system which analyses its standards in teaching and research on a regular basis and is also able to accredit degree programmes. External peers are occasionally invited to look at the HEIs and examine so-called random samples (*Stichproben*). If the HEI has had its *Systemakkreditierung* accredited by the *Akkreditierungsrat* (or an accreditation agency which has the capacity to bestow the same licence), then it is assumed that certain standards will be met.

A recent conference in Bonn showed that universities are still in the infancy of *System-akkreditierung* and that this new system of quality assurance has yet to become established (if indeed it does so).[284] We might see different schemes being established throughout the country and although minimum standards may be met, it remains to be seen if this new idea of quality assurance will actually maintain or preferably raise the standards of Germany's HEIs overall. It could be argued that a sense of complacency might enter into the process once the system has been installed. Furthermore, will *Systemakkreditierung* really take into account the different cultures present within the higher education arena? A one-size-fits-all approach does not seem to be the way forward here, as the present study shows, where each evaluation requires particular evaluation criteria suitable to particular academic cultures to be established.

[284] See http://interchange.university-partners.de for further information on conference "Erste Gehversuche mit der Systemakkredierung", Bonn, 13-14 September 2009

From the experiences of those working in quality assurance in Germany it seems that individual university heads can be very influential in improving the quality of their institution as a whole, but also the individual disciplines within it. Institutional leadership can and does have a huge impact on what a particular discipline is able to achieve. System accreditation, therefore, needs to work very closely with the university heads to ensure that the process remains not only stable, but can move the HEI forward.

9.3 Evaluations in *Erziehungswissenschaft*

With regard to the quality of *Erziehungswissenschaft* and the manner in which it has been assessed over the past few years, Tippelt (2008) concludes that the majority of HEIs are now working according to internationally accepted standards and that external evaluations have become a regular feature of an ongoing engagement with the maintenance of good standards. He reinforces the argument that external peers remain an essential part of quality assurance in this field (and elsewhere), and that they conduct onsite visits and engage with the variety of stakeholders to be found there. However, he stresses that HEIs need to maintain a balance in their search for high quality in both research *and* teaching. Standards can only be improved if the faculty members actively engage with the recommendations which evaluators provide them. If HEIs are over-evaluated, there is a danger that they may become immune to suggestions for improvement (ibid., p.186). This would be 'counter-productive'.

9.4 Changing role of *evalag*

Evalag is still fairly young in terms of evaluation agencies in Germany. Like the ZEvA, it has evaluated disciplines across the state concurrently, but there have been no follow-up procedures to date. In 2005 *evalag* started experiencing difficulties as universities decided to step outside the evaluation procedure. This came about for two reasons. As mentioned above, universities were struggling under the burden of accreditation procedures and secondly, there was a lack of clarity on *evalag's* role itself. As Dr Gerhard Schreier told me in our interview (19.12.07), *evalag* was supposed to undergo a meta-evaluation to examine

how well it was operating. A commission was established to undertake this task; however, the universities refused to take part, therefore making the entire evaluation operation redundant. Kohl, who was one of VW's chief executives and a member of the commission, wrote to Schreier saying that '*evalag* is failing because of the way it is governed'.[285] The perceived problem was the Ministry's involvement and the fact that *evalag's* governing body was not entirely independent. More support could have come from Minister Frankenberg; however, although he had been a former university president in Mannheim and knew the HEI scene extremely well he failed to provide it. It was possibly for this reason that Schreier decided to vacate his position as *evalag's* director, as he could not envisage that any great changes would occur. The Ministry/autonomy problem was so acute that it also led to the resignation of two other members of the board of directors at the end of 2005. It was therefore necessary to temporarily put on hold all regular, comparative evaluations of teaching and research in Baden-Württemberg.

The following year saw a new orientation take place. Running projects were completed during this time and on 31 December, its founding director, Dr Gerhard Schreier, left the company. Dr Martina Röbbecke took over as acting head in the first half of 2007 before making way for Dr Anke Rigbers on July 1 and she has remained in post since then. Towards the middle of 2007 the project 'Institutional Quality Assurance (*evalag* IQ)' was restarted in which an international committee of experts started to develop a guidebook on institutional quality assurance.[286] Today, *evalag's* operations also include acting in an advisory capacity by supporting institutions in the creation of their own quality assurance mechanisms within the framework of system accreditation. The agency is also on the verge of accreditation by the *Akkreditierungsrat*, so that it will be able to provide accreditation for degree programmes. It will be able to work outside the borders of Baden-Württemberg and assist any HEI that asks for its support.

[285] *Evalag* scheitert an der Governance-Struktur.
[286] see http://www.evalag.de/dedievl/projekt01/index.php?idcat=7&sid=496489a91144e3caa884173a78350e8a

9.3 A vision for the future?

If we look into the very distant future we may see external quality assurance agencies of *evalag's* type becoming redundant as quality assurance mechanisms within institutions become so refined as to make any external agency support superfluous. It could be argued that *evalag* would have therefore achieved the objective stated on its website that 'The purpose of evaluation is not to prove, but to improve' (Egon Guba). If by accompanying institutions towards their own autonomous quality assurance, it would indeed have served the community to improve itself.

REFERENCES

Alderman, G. & Brown, R. (2005) Can Quality Assurance survive the Market? Accreditation and Audit at the Crossroads, *Higher Education Quarterly*, Vol.59:4, pp.313-328

Alkin, M.C., Daillak, R. & White, R. (1979) *Using Evaluations. Does Evaluation Make a Difference*? (Beverly Hills & London: Sage)

Altrichter, H., Schratz, M., & Pechar, H. (eds.) (1997) *Hochschulen auf dem Prüfstand* (Innsbruck: Studienverlag)

Aylett, R. & Gregory. K. (eds.) (1996) *Evaluating Teacher Quality in Higher Education* (London: Falmer)

Bailey, R. (1999) The Abdication of Reason: Postmodern Attacks Upon Science and Reason. In J. Swann & J. Pratt (eds.) (1999) *Improving Education: Realist Approaches to Method and Research* (London & New York: Cassell), pp.30-38

Ball, C. (1985) *What the Hell is Quality?* (Guildford: SRHE)

Ball, S. (1990) *Politics and Policy Making in Education. Explorations in Policy Sociology* (London & New York: Routledge)

Balzer, L. (2005) *Wie werden Evaluationsprojekte erfolgreich? Ein integrierender theoretischer Ansatz und eine empirische Studie zum Evaluationsprozess* (Landau: Verlag Empirische Pädagogik)

Bao, X.-M. (1999) A Phenomenology of Teaching with Laptop Computers: A Case Study through the Eyes of a Trainer. In H.E. Klein (ed.) *Interactive Teaching and the Multimedia Revolution. Case Method and Other Techniques* (Wisconsin: OMNI PRESS), pp.43-52. Available online: <http://pirate.shu.edu/~baoxuemi/bao9808.pdf> Accessed 14 June 2009

Barnett, R. (1988) Limits to academic freedom. In M. Tight (ed.) (1998) *Academic Freedom and Responsibility* (Milton Keynes: Open University Press), pp.88-103

Barnett, R. (1992) *Improving Higher Education: total quality care* (Buckingham: SRHE & Open University Press)

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

Beck, D.E., & Cowan, C.C. (1996) *Spiral Dynamics: Mastering Values, Leadership, and Change* (Malden: Blackwell)

Becher, T. & Trowler, P.R. (2001) *Academic Tribes and Territories*, 2nd edition (Maidenhead & New York: SRHE)

Becker, C.H.F. (1996) State Secretary and Prussian Minister of Culture, source unknown. In P. Glotz (1996) *Im Kern verrottet? Fünf vor zwölf an Deutschlands*

Universitäten (Stuttgart: Deutsche Verlagsanstalt GmbH), p.7

Bednarz (1985) Quantity and Quality in Evaluation Research. A Divergent View. In S.B. Merriam (1995) *Case Study Research in Education. A Qualitative Approach* (San Francisco: Jossey-Bass), pp.170-171

Bewyl, W. (1988) *Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation* (Frankfurt, Bern, New York: Peter Lang)

Bewyl, W. (2003) Selected Comments to the Standards for Evaluation of the German Evaluation Society, English edition, (Cologne & Berne: DGEval). Available online: <http://www.evaluation.org.uk/Pub_library/German%20Evaluation%20Standards_030409.pdf> Accessed 8 June 2006

Böllert, K. & Tippelt, R. (2004) Evaluationsverfahren in der Erziehungswissenschaft. In R. Tippelt et al. (eds.) (2004) *Datenreport Erziehungswissenschaft 2004* (Wiesbaden: VS Verlag für Sozialwissenschaften, Leske & Budrich), pp.139-152

Bourdieu, P. (1988) *Homo Academicus* (Cambridge: Polity)

Brandt, T. (2007) Sozialer Kontext der Evaluation. In R. Stockmann (ed.) (2007) *Handbuch zur Evaluation. Eine praktische Handlungsanleitung* (Münster: Waxmann), pp.164-194

Brint, S. (1994) *In an Age of Experts. The Changing Role of Professionals in Politics and Life* (New Jersey: Princeton University Press)

Burgess, R.G. (ed.) (1989) *The Ethics of Educational Research* (New York & London: Falmer)

Chamberlain, D.H. (1885) The Educational Debate: The German University and Gymnasium and the American College, *The New York Times*, 27 February. Available online: <http://query.nytimes.com/mem/archivefree/pdf?_r=1&res=9407E7D6143FE533A25754C2A9649C94649FD7CF&oref=slogin> Accessed 9 November 2008

Chiseri-Strater, E. & Sunstein, B.S. (1997) Fieldworking: Reading and writing research. In J.A. Hatch (2002) *Doing Qualitative Research in Education Settings* (Albany: State University of New York Press), p.95

Csizmadia, T., Enders, J., & Westerheijden, D. (2008) Quality management in Hungarian higher education: organisational responses to governmental policy, *Higher Education*, Vol.56:4, pp.439-455

Davies, J. (2000) Cultural Change in Universities in the Context of Strategic and Quality Initiatives. In P. Tabatoni et al., (2000) *Strategic Management and Universities' Institutional Development* (EUA Geneva & Brussels). Available online: <http://www.unige.ch/eua> Accessed 7 May 2006

Denzin, N.K. & Lincoln, Y.S. (eds.) (1994) *Handbook of Qualitative Research* (Thousand

Oaks, London, New Delhi: Sage)

Denzin, N.K. & Lincoln, Y.S. (eds.) (2000) *Handbook of Qualitative Research*, 2nd edition (Thousand Oaks, London, New Delhi: Sage)

Dressel. P.L. (1982) Values (Virtues and Vices) in Decision-Making, *New Directions for Higher Education*, No.37, Vol.10:1, pp.31-43

Dube, C. (2006) *Vergleich der Schul- und Hochschulsysteme in der Bundesrepublik Deutschland und England/Wales unter besonderer Berücksichtigung der Bildungsfinanzierung* (Munich: GRIN Verlag)

Elton, L. & Partington, P. (1991) Teaching Standards and Excellence in Higher Education: Developing a Culture for Quality, *Occasional Green Paper No 1* (Sheffield: Training Unit)

Enders, J. (1997) *The Academic Profession in Europe: A View from the Status Perspective*. CERT Discussion Paper 97/18 (Kassel: Centre for Research into Higher Education and Work)

Enders, J. (2001) A Chair System in Transition: Appointments, Promotions, and Gate-keeping in German Higher Education, *Higher Education*, Vol.41:1-2, pp. 3-25

Erkens, H. (ed.) (2004) *Evaluation in der Erziehungswissenschaft* (Wiesbaden: VS Verlag für Sozialwissenschaften, Leske & Budrich)

Evans, G.R. (1999) *Calling Academia to Account. Rights and Responsibilities* (Buckingham: SRHE & OUP)

Federkeil, G. (2004) Evaluation und Akkreditierung, *wissenschaftsmanagement 1*, January/February 2004 (Bonn: Lemmens Verlags- und Mediengesellschaft mbH)

Fetterman, D.M. (1991) A Walk Through the Wilderness. Learning to Find Your Way. In W.B. Shaffir & R.A. Stebbins (eds.) (1991) *Experiencing Fieldwork. An Inside View of Qualitative Research* (Newbury Park, CA: Sage), Chap.6, pp.87-96

Fetterman, D.M., Kaftarian, S.J. & Wandersman, A. (eds.) (1996) *Empowerment evaluation: Knowledge and tools for self-assessment & accountability* (Thousand Oaks: Sage)

Fischoff, B. (1996) The Real World: What good is it? *Organisational Behavior and Human Decision Processes*, Vol.65, pp.232-248. In R. Hastie & R.M. Dawes (2001) *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision-Making* (Thousand Oaks, CA: Sage)

Fournier, D.M. (2005) Evaluation. In S. Mathison (ed.) *Encyclopedia of Evaluation* (Thousand Oaks: Sage), pp.139-140

Frackmann, E. (1992) The German Experience. In A. Craft (ed.) (1992) *Quality Assurance in Higher Education*. Proceedings of an International Conference, Hong Kong 1991 (London & Washington: Falmer), pp.47-64

Führ, C. & Furck, C.-L. (eds.) (1997) *Handbuch der deutschen Bildungsgeschichte, Bd.VI. 1945 bis zur Gegenwart* (C.H. Beck: Munich)

Garvin, D.A. (1984) What Does Product Quality Really Mean? *Sloan Management Review*, Vol.26:1, pp.25-43. Available online: <http://doku.iab.de/externe/2006/k060210f02.pdf> Accessed 30 June 2006

Gibson, A. (1986) Inspecting education. In G. Moodie (ed.) (1986) *Standards and Criteria in Higher Education* (Guildford: SRHE), pp.128-135

Giorgi, A. (1985) *Phenomenology and Psychological Research* (Pittsburgh: Dusquesne University Press)

Giorgi, A. (1992) Description versus interpretation: competing strategies for qualitative research, *Journal of Phenomenological Research*, Vol.23, pp.119-135

Giorgi, A. (1994) A phenomenological perspective on certain qualitative research methods, *Journal of Phenomenological Psychology*, Vol.25, pp.190-220

Glotz, P. (1996) *Im Kern verrottet? Fünf vor zwölf an Deutschlands Universitäten* (Stuttgart: Deutsche Verlagsanstalt GmbH)

Goetze, W. (2005) Zwei Jahrzehnte Bildungsevaluation: Erwartungen, Angebote und Realität, paper, *Annual Conference of SEVAL*, Berne, 3 June. Available online: <http://www.seval.ch/documents/kongresse/2005/Unterlagen/Referat_Walter_Goetze.pdf> Accessed 8 June 2006

Good, B. (1994) *Medicine, Rationality and Experience. An anthropological perspectiv*e. Lewis Henry Morgan Lectures (Cambridge: Cambridge University Press)

Gottschalk, L. Kluckhohn, C. & Angell, R. (1945) The Use of Personal Documents in History, Anthropology and Sociology. In C. Robson (2002) *Real World Research* (London: Blackwell), p.350

Grant, R. (2008) A Phenomenological Case Study of a Lecturer's Understanding of Himself as Assessor, *Indo-Pacific Journal of Phenomenology*, Special Edition February 2008. Available online: <http://www.ipjp.org/index.php?option=com_jdownloads&Itemid=25&task=view.download&cid=47> Accessed 30 May 2009

Green, D. (ed.) (1994) *What is Quality in Higher Education*? (Buckingham: SRHE & Open University Press)

Gregory, K. (1996) The Evaluation of the Teaching of Individual Academics in Higher Education: Progress towards a Construct. In R. Aylett & K. Gregory (eds.) (1996) *Evaluating Teacher Quality in Higher Education* (London: Falmer)

Gruber, K.-H. (2006) The German PISA-shock: some aspects of the extraordinary impact of the OECD's PISA study on the German education system. In H. Ertl (ed.) (2006) *Cross-national Attraction in Education. Accounts from England and Germany* (Didcot: Symposium Books), pp.195-208

Gutachterkommission (2004) *Evaluation der Erziehungswissenschaften an den Universitäten und Pädagogische Hochschulen des Landes Baden-Württemberg* (Bonn: Lemmens Verlags- und Mediengesellschaft mbH)

Guba, E.G. & Lincoln, Y.S. (1981) *Effective Evaluation* (San Francisco/London: Jossey Bass)

Guba, E.G. & Lincoln, Y.S. (1989) *Fourth Generation Evaluation* (Newbury Park, London & New Delhi: Sage)

Habermas, J. (1979) *Communication and the Evolution of Society,* T. McCarthy (trans.) (Boston: Beacon)

Hahn, H.-J. (1998) *Education and Society in Germany* (Oxford: Berg)

Hammersley, M. & Atkinson, P. (2007) *Ethnography: Principles in Practice*, 3rd edition (London: Routledge)

Handbook for Evaluation Teams, 6th Edition (2004), Middle States Commission on Higher Education (Philadelphia)

Handbook for Commission Members (2003) Middle States Commission on Higher Education (Philadelphia). Available online: <http://www.msche.org/publications/CommHandbook04050310080257.pdf>. Accessed 23 January 2007

Hanft, A. (2003) Evaluation und Organisationsentwicklung, *EvaNet-Positionen 10/2003*, Available online: <http://evanet.his.de> Accessed 14 June 2006

Harman, G. (1998) The management of quality assurance: a review of international Practice, *Higher Education Quarterly*, Vol.52:4, pp.345-364. Available online: <http://www3.interscience.wiley.com/cgi-bin/fulltext/11927994/PDFSTART> Accessed 14 June 2006

Harris-Huemmert, S. (2007a) Researching External Evaluators of Higher Education in Germany. Methods and Techniques, *Research in Comparative and International Education*, Vol.2:2, pp.135-143. Available online: <http://doi:10.2304/rcie.2007.2.2.135> Accessed 30 May 2007

Harris-Huemmert, S. (2007b) Evaluations of Higher Education in Germany: A Question of Size, presentation, QAA/SRHE conference *Emerging landscapes: research into quality management in higher education*, Birmingham, 27 June 2007

Harris-Huemmert, S. (2008) Evaluators of Higher Education in Germany: Are they 'fit for purpose?' *Quality in Higher Education*, Vol.14:1, pp.55-65 Available online: <http://dx.doi.org/10.1080/13538320802011722> Accessed 19 May 2008

Harvey, L. & Green, D. (1993) Defining quality, *Assessment and Evaluation in Higher Education*, Vol.18:1, pp. 9-34

Hastie, R. & Dawes, R.M. (2001) *Rational Choice in an Uncertain World: The Psychology*

of Judgment and Decision-Making (Thousand Oaks, CA: Sage)

Hatch, J.A. (2002) *Doing Qualitative Research in Education Settings* (Albany: State University of New York Press)

Heidegger (1927) *Basic Problems of Phenomenology*, trans. A. Hofstadter (Bloomington: Indiana University Press)

Henkel, M. (2000) *Academic Identities and Policy Change in Higher Education* (London: Jessica Kingsley)

Hertz, R. & Imber, J.B. (eds.) (1995) *Studying Elites using Qualitative Methods* (Thousand Oaks/London/New Delhi: Sage)

Hillier, Y. (1999) Exploring Informal Practitioner Theory in Adult Basic Education. In J. Swann & J. Pratt (eds.) (1999) *Improving Education: Realist Approaches to Method and Research* (London & New York: Cassell), p.154

Hopbach, A. (2006) The European Standards and Guidelines and the Evaluation Agencies in Germany, *Quality in Higher Education* Vol.12:3, pp.235-242 Available online: <http://dx.doi.org/10.1080/13538320601051044> Accessed 18 September 2008

Hornbostel, S., & Simon, D. (eds.) (2006) *Wie viel (In-) Transparenz ist notwendig? Peer Review Revisite*d, iFQ Working Paper, No.1, December 2006. Available online: <http://www.forschungsinfo.de/Publikationen/Download/working_paper_1_2006.pdf> Accessed June 9 2007

Howard, D.C.P. (1994) Human-computer interactions: A phenomenological examination of the adult first-time computer experience. *Qualitative Studies in Education*, Vol.7, pp.33-49

Huber, K. (2003) *Evaluation an Hochschulen: Evaluationen an der Universität Bern*, presentation at Annual SEVAL conference, 3 June. Available online: <http://www.advd.unibe.ch/uls/2_Ebene/Eval_Aufgaben.htm> Accessed 2 September 2006

Huisman, J. (2003) Higher Education in Germany. In *CHEPS* – higher education monitor 2003. Available online: <http://www.utwente.nl/cheps/documenten/germany.pdf> Accessed 10 November 2008

Humboldt, Wilhelm von (1810) Über die innere und äußere Organisation der höheren wissenschaftlichen Anstalten in Berlin. In *Die Idee der deutschen Universität: Die fünf Grundschriften aus der Zeit ihrer Neubegründung durch klassischen Idealismus und romantischen Realismus* (1956) (Darmstadt: Wissenschaftliche Buchgesellschaft) Unvollendete Denkschrift, geschrieben 1810 (vielleicht 1809) Erstmals veröffentlicht 1896 durch Bruno Gebhardt, Wilhelm von Humboldt als Staatsmann, Bd. I, S. 118. Sodann W. v. Humboldt, *Werke*, Akademieausgabe Bd. X., S. 250

Johnes, G. & Rothschild, R. (1975) Assessing Assessment. The Evaluation of Teaching and Research in UK universities, RES newsletter *Evaluating Teaching in Higher*

*Education: a collection of conference paper*s (London: UTMU)

Jung, C. (1921/1971) Psychological Types, *Collected Works*, Vol.6, Chapter XI (London: Routledge & Kegan Paul), p.414

Kalkwijk, J. (1992) The Netherlands: The Inspectorate Perspective. In A. Craft (ed.) (1992) *Quality Assurance in Higher Education*. Proceedings of an International Conference, Hong Kong 1991 (London & Washington: Falmer), pp.97-108

Keiner, E. (ed.) (2001) *Evaluation (in) der Erziehungswissenschaft* (Weinheim & Basel: Beltz)

Kieser, A. (1998) Das Idealbild der Evaluationskonstrukteure. Was lehren niederländische Erfahrungen mit der Evaluation universitärer Forschung? *Forschung und Lehre*, Vol.8, pp.408-411. Available online: <http://www.forschung-und-lehre.de/pdf/08_98.pdf> Accessed 12 June 2006

Kieser, A. (1999a) Über Marktmechanismen nachdenken. Aspekte zum Wettbewerb an Universitäten, *Forschung und Lehre* 6/99, pp.284-285. Available online: <http://www.forschung-und-lehre.de/pdf/06_99.pdf> Accessed 14 June 2006

Kieser, A. (1999b) Viel Lärm um nichts? Evaluation von Studium und Lehre und ihre Folgen, *Beiträge zur Hochschulpolitik: Projekt Qualitätssicherung 4/1999*, paper, Rostock University, 6-8 September 1998

Kincheloe, J.L. and Berry, K.S. (2004) *Rigour and Complexity in Educational Research. Conceptualising the bricolage* (Maidenhead: Open University Press)

Klieme. E. & Tippelt, R. (eds.) (2008) Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz, *Zeitschrift für Pädagogik*, 53. Beiheft, Dezember 2008, (Weinheim & Basle: Beltz)

Kogan, M. (1986) *Educational Accountability* (London: Hutchinson)

Kohler, J. (2003) Institutionelle Qualitätssicherung statt Programmevaluation? Vortrag zur Tagung des Projekts Qualitätssicherung der HRK „Qualitätssicherung an Hochschulen Neue Herausforderungen nach der Berlin-Konferenz“ am 20./21. November 2003 in Bonn. *EvaNet-Positionen* 12/2003-2004 Available online: <http://www.hrk.de/de/download/dateien/12-2003_-_Institutionelle_Qualitaetssicherung_statt_Programmevaluation_-_Kohler.pdf> Accessed 10 November 2008

Kolb, D.A. (1981) Learning Styles and Disciplinary Differences. In. A. Chickering (ed.) (1981) *The Modern American College* (San Francisco, CA: Jossey-Bass)

Kristofferson, D., Sursock, A. & Westerheijden, D.F. (1998) *Quality assurance in higher education – manual of quality assurance: procedures and practice* (Turin: European Training Foundation)

Kromrey, H. (2001a) Evaluation von Lehre und Studium –Anforderungen an Methodik und Design. In C. Spiel (ed.) (2001) *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck* (Munster, New York, Munich & Berlin:

Waxmann)

Kromrey, H. (2001b) Evaluation – ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung – Empfehlungen für die Praxis, *Sozialwissenschaft und Berufspraxis*, Vol.24:2, pp.105-131. Available online: <http://www.bibb.de/dokumente/pdf/a11_vielschichtiges_konzept.pdf> Accessed 5 November 2008

Kromrey, H. (2004) Evaluation oder Selbstevaluation? Dilemmata im peer-review-Konzept. In: H. Merkens (ed.) (2004) *Evaluation in der Erziehungswissenschaft* (Wiesbaden: VS Verlag für Sozialwissenschaften, Leske & Budrich)

Kromrey, H. (2006) Qualität und Evaluation im System Hochschule. In R. Stockmann (ed.) (2006) *Evaluationsforschung, Grundlagen und ausgewählte Forschungsfelder* (Opladen: Leske & Budrich) pp.233-258

Künzel, R. (1997) Political Control and Funding: The Future of State Support. In M.G. Ash (ed.) (1997) *German Universities Past and Future – Crisis or Renewal*? (Providence & Oxford: Berghahn Books)

Law, A.I. & Bronson, W.H. (1977) *Evaluation Trainer's Guide. The Evaluation Improvement Program* (Princeton, NJ: Educational Testing Service, Evaluation Improvement Program)

Law, A.I. & Bronson, W.H. (1977) *Program Evaluator's Guide* (Princeton, NJ: Educational Testing Service, Evaluation Improvement Program)

Lee, B. (2006) Theories of Evaluation. In R. Stockmann (ed.) (2006) *Evaluations-forschung, Grundlagen und ausgewählte Forschungsfelder* (Opladen: Leske & Budrich), pp. 137-178

Lincoln Y.S. & Guba, E.G. (1985) *Naturalistic Inquiry* (Thousand Oaks, CA: Sage)

Lohnert, B. & Rolfes, M. (1997) Handbuch zur Evaluation von Lehre und Studium an Hochschulen. Ein praxisorientierter Leitfaden, *Schriftenreihe „Evaluation der Lehre"* 3/97 (Hannover: ZEvA)

Lonsdale, A. (ed.) (1990) *Academic Staff Appraisal in Australian Higher Education* (Canberra: Australian Government Publishing Service)

Maats, P. (2007) Einführung in das Datenmanagement und die Datenauswertung. In R. Stockmann (ed.) (2007) *Handbuch zur Evaluation. Eine praktische Handlungs-anleitung* (Munster, New York, Munich & Berlin: Waxmann), pp.278-313

Magd, H. & Curry, A. (2003) An empirical analysis of management attitudes towards ISO9001:2000 in Egypt, *The TQM Magazine*, Vol.15, pp.381-90

McClain, C.I, Krueger, D.W., & Taylor, T. (1989) Northeast Missouri State Universities Value Added Assessment Profile. A Model for Educational Accountability. In M. Kogan (ed.) (1989) *Evaluating Higher Education* (London: Hutchinson)

McCulloch, G. & Richardson, W. (2000) *Historical Research in Educational Settings*

(Buckingham & Philadelphia: Open University Press)

Meltsner, A. (1976) *Policy Analysts in the Bureaucracy* (Berkeley: University of California Press)

Merleau-Ponty, M. (1945) *Phenomenology of Perception* (translated & reprinted 2008) (Abingdon: Routledge)

Merriam, S.B. (1988) *Case Study Research in Education. A Qualitative Approach* (San Francisco: Jossey-Bass)

Merkens, H. (ed.) (2004*) Evaluation in der Erziehungswissenschaft* (Wiesbaden: VS Verlag für Sozialwissenschaften, Leske & Budrich)

Mertens, D. (1998) *Research methods in education and psychology: intergrating diversity with quantitative and qualitative approaches* (Thousand Oaks, CA: Sage)

Middlehurst, R. (2004) Changing Internal Governance: A Discussion of Leadership Roles and Management Structures in UK Universities, *Higher Education Quarterly*, Vol.58:4, pp.258-279

Minelli, E., Rebora, G. & Matteo, T. (2008) How Can Evaluation Fail? The Case of Italian Universities, *Quality in Higher Education*, Vol. 14:2, pp.157-173 Available online:<http://dx.doi.org/10.1080/13538320802278412> Accessed 27 November 2008

Mittag, S. (2006) *Qualitätssicherung an Hochschulen. Eine Untersuchung zu den Folgen der Evaluation von Studium und Lehre* (Munster, New York, Munich & Berlin: Waxmann)

Mittag, S., Bornmann, L., & Daniel, H.-D. (2003) *Evaluation von Studium und Lehre an Hochschulen. Handbuch zur Durchführung mehrstufiger Evaluationsverfahren* (Münster, New York, Munich & Berlin: Waxmann)

Morley, L. (2003) *Quality and Power in Higher Education* (Maidenhead & Philadelphia: SRHE)

Morley, L. (2004) *Theorising Quality in Higher Education* (London: Institute of Education)

Müller-Böling, D. (ed.) (1996) *Qualitätssicherung in Hochschulen* (Gütersloh: Verlag Bertelsmann Stiftung)

Müller-Böling. D. (1997) Evaluation zur Rechenschaftslegung oder Qualitäts-verbesserung? Eine Bestandsaufnahme der Evaluation an deutschen Hochschulen. In H. Altrichter, M. Schratz, & H. Pechar (eds.) (1997) *Hochschulen auf dem Prüfstand* (Innsbruck: Studienverlag), pp.88-107

Müller-Böling, D. (2000) *Die entfesselte Hochschule* (Gütersloh: Verlag Bertelsmann Stiftung)

Murphy, E. & Dingwall, R. (2001) The Ethics of Ethnography. In Atkinson et al. (eds.)

(2001) *Handbook of Ethnography* (London/Thousand Oaks/New Delhi: Sage), pp.339-351

Newman, J.H. (1899) *The Idea of a University*, reprinted 1996 (New Haven & London: Yale University Press)

Nickel, S. (2005) EFQM an Universitäten – Nachhaltige Qualitätsentwicklung oder Management-Mode? *evaNet-Positionen* 08/2005, pp.1-19. Available online: <http://www.hrk.de/de/download/dateien/08-2005_-_EFQM_an_Universitaeten_-_Nickel.pdf> Accessed 10 November 2008

Norris, P. (ed.) (1999) *Critical Citizens. Global Support for Democratic Government* (Oxford: Oxford University Press)

O'Grady, J. & French, G. (2000) The Definition of a University. In *Northrop Frye's Writings on Education,* Vol.7., (Toronto, Buffalo & London: University of Toronto Press), pp.406-421

Orr, D. & Pätzold, M. (2006) Procedures for Research Evaluation in German Higher Education: Current Fragmentation and Future Prospects, *ACCESS. Critical Perspectives on Communication, Cultural & Policy Studies*, Vol.25:2, pp. 16-30

Ory, J. (2000) Teaching Evaluation: Past, Present, and Future. In K. Ryan (ed.) (2000) Evaluating Teaching in Higher Education: A Vision for the Future, *New Directions for Teaching Learning* Vol.83, (San Francisco: Jossey-Bass) pp.15-19

Ostrander, S.A. (1995) 'Surely You're Not in This Just to be Helpful'. Access, Rapport and Interviews in Three Studies of Elites. In R. Hertz & J.B. Imber (eds.) (1995) *Studying Elites using Qualitative Methods* (Thousand Oaks/London/New Delhi: Sage) pp.133-150

Otto, H.-U., Krüger, H.-H., Merkens, H., Rauschenbach, T., Schenk, B., Weishaupt, H., & Zedler, P., (2000) *Datenreport Erziehungswissenschaft* (Opladen: Leske & Budrich)

Ouichi, W.G. (1979) A Conceptual Framework for the Design of Organisational Control Mechanisms, *Management Science*, Vol.25, p.9

Owen, J. & Rogers, P.J. (1999) *Program Evaluation. Forms and Approaches* (London & New York: Sage)

Patton, M.Q. (1982) *Practical Evaluation* (Beverley Hills: Sage)

Patton, M.Q. (1987) *How to use Qualitative Methods in Evaluation* (Newbury Park: Sage)

Patton, M.Q. (2002) *Qualitative Research and Evaluation Methods*, 3rd edition (Thousand Oaks & London: Sage)

Pellert, A. (2002) Hochschule und Qualität. In T. Reil & M. Winter (eds.) (2002) *Qualitätssicherung an Hochschulen: Theorie und Praxis* (Bielefeld: Bertelsmann), pp.21-29

Phillips, D. & Filmer-Sankey, C. (1993) *Diversification in Modern Language Teaching. Choice and the National Curriculum* (London & New York: Routledge)

Picht, G. (1964) *Die deutsche Bildungskatastrophe. Analyse und Dokumentation* (Olten/ Freiburg: Walter)

Pirsig, R.A. (1974) *Zen and the Art of Motorcycle Maintenance* (New York: Bantam), pp.163-164

Pollitt, C. (1998) Evaluation in Europe: Boom or Bubble? *Evaluation*, Vol.4:2, pp.214-224

Popper, K. (2004) *Conjectures and Refutations* (London and New York: Routledge Classics)

Power, M. (1994) *The Audit Society Rituals of Verification* (Oxford: Oxford University Press)

Pring, R. (2000) *Philosophy of Educational Research* (London & New York: Continuum)

Pritchard, R. (1990) *The End of Elitism? The Democratisation of the West German University System* (New York, Oxford & Munich: Berg)

Pritchard, R. (2004) Humboldtian values in a changing world: staff and students in German universities, *Oxford Review of Education*, Vol.30:4, pp.509-527

Raiffa, H., Richardson, J. & Metcalfe, D. (2002) *Negotiation Analysis. The Science and Art of Collaborative Decision-Making* (Cambridge, Mass. & London: Belknap Press of Harvard University Press)

Ramsden, P. (1991) A performance indicator of teaching quality in higher education: The Course Experience Questionnaire, *Studies in Higher Education*, Vol.16:2, pp.129-150

Readings, B. (1996) *The University in Ruins* (Cambridge, Massachusetts & London: Harvard University Press)

Reil, T. & Winter, M. (eds.) (2002) *Qualitätssicherung an Hochschulen: Theorie und Praxis* (Bielefeld: Bertelsmann)

Rindermann, H. (2001) *Lehrevalaution: Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen mit einem Beitrag zur Evaluation computerbasierten Unterrichts* (Landau: Verlag Empirische Pädagogik)

Ringer, F.K. (1969) *The Decline of the German Mandarins* (Cambridge, Mass.: Harvard University Press)

Robson, C. (2002) *Real World Research* (London: Blackwell)

Rossi, P.H., Lipsey, M.W., & Freeman, H.E. (2004) *Evaluation. A Systematic Approach*, 7th edition (Thousand Oaks, London & New Delhi: Sage)

Ryan, K (ed.) (2000) Evaluating Teaching in Higher Education: A Vision for the Future, *New Directions for Teaching Learning*, No.83 (San Francisco: Jossey-Bass)

Sadler, D.R. (1981) Intuitive Data Processing as a Potential Source of Bias in Naturalistic Evaluation, *Educational Evaluation and Policy Analysis*, Vol.3:4, pp.25-31

Saunders, L. (ed.) (2007) *Educational Research and Policy-Making. Exploring the border country between research and policy* (New York & London: Routledge)

Schade, A. (2004) Shift of Paradigm in Quality Assurance in Germany: More Autonomy, but Multiple Quality Assessment? In S. Schwarz & D.F: Westerheijden (2004) *Accreditation and Evaluation in the European Higher Education Area* (Dordrecht: Kluwer Academic Publishers), pp.175-196

Schleiermacher, F. (1808) Gelegentliche Gedanken über Universitäten im deutschen Sinn. In E. Anrich (1956) *Die Idee der deutschen Universität* (Darmstadt: Hermann Gentner), pp.219-308

Schön, D. (1983) *The Reflexive Practitioner: How Professionals Think in Action* (New York: Basic Books)

Schöne, W. (1966) *Kampf um die deutsche Universität* (Hamburg: Hans Christian Verlag)

Schwarz, S. & Westerheijden, D.F. (2004) *Accreditation and Evaluation in the European Higher Education Area*, (Dordrecht: Kluwer Publishers)

Schweitzer, R. (2002) Editorial, *Indo-Pacific Journal of Phenomenology*, Vol.2:2, pp.1-2

Shaffir, W.B. & Stebbins, R.A. (eds.) (1991) *Experiencing Fieldwork. An Inside View of Qualitative Research* (Newbury Park: Sage), pp.87-96

Silvers, P.J. (1982) Regional Accreditation Team Members: How They Perceive Their Role and How They Make Decisions, *ASHE Annual Meeting 1982.* Paper presented at the Annual Meeting of the Association for the Study of Higher Education, Washington DC, 2-3 March, 1982

Simon, D. (2003) Selbststeuerung der Wissenschaft durch Evaluationen? Optionen und Grenzen institutioneller Forschungsbewertungen. In K. Fischer & H. Parthey (eds.) (2003) *Evaluation wissenschaftlicher Institutionen: Wissenschaftsforschung Jahrbuch 2003* (Berlin: Gesellschaft für Wissenschaftsforschung 2004), pp.9-18 Available online: <http://www.wissenschaftsforschung.de/JB03_9-21.pdf> Accessed 12 June 2006

Smith, N.L. (1990) Cautions on the use of investigative case studies in meta-evaluation. In R.K. Yin (2003) *Case Study Research: design and methods*, 3rd edition (Thousand Oaks/CA, London, New Delhi: Sage), pp.373-378

Spiegelberg, H. (1975) *Doing Phenomenology: Essays on and in Phenomenology* (The Hague: Martinus Nijhoff)

Spiel, C. (ed.) (2001) *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck* (Munster, New York, Munich & Berlin: Waxmann)

Spinelli, E. (1989) *The interpreted world: An introduction to phenomenological psychology* (London: Sage)

Sproull, L. & Larkey, P. (1979) Managerial Behaviour and Evaluator Effectiveness. In H.C. Schulberg & J.J. Jerrell (eds.) (1979) *The Evaluator and Management* (Beverly Hills & London: Sage) pp.89-104

Stake, R. (1975) Programm Evaluation. Particularly Responsive Evaluation (University of Illinois at Urbana-Champaign: Center for Instructional Research and Curriculum Evaluation), from a paper presented at a conference on „New Trends in Evaluation", Göteborg, Sweden, October 1973. Available online: <http://www.wmich.edu/evalctr/pubs/ops/ops05.html> Accessed 8 June 2006

Stake, R.E. (1986) Quieting reform: social science and social action in an urban youth program. In R.K. Yin (2003) *Case Study Research: design and methods*, 3rd edition (Thousand Oaks/CA, London, New Delhi: Sage), p.15

Stake, R.E. (2004) *Standards-based and Responsive Evaluation* (Thousand Oaks/London: Sage)

Stensaker, B. (2003) Trance, Transparency and Transformation: the impact of external quality monitoring on higher education, *Quality in Higher Education*, Vol.9:2, pp.151-159. Available online: <http://dx.doi.org/10.1080/1353832308158> Accessed 18 September 2008

Stensaker, B. (2008) Outcomes of Quality Assurance: A Discussion of Knowledge, Methodology and Validity, *Quality in Higher Education*, Vol.14:1, pp.3-13

Stock, M. (2004) Steuerung als Fiktion. Anmerkungen zur Implementierung der neuen Steuerungskonzepte an Hochschulen aus organisations-soziologischer Sicht, *die Hochschule* 1/2004. Available online: <http://www.hof.uni-halle.de/journal/hs2004_1.pdf> Accessed 4 May 2006

Stockmann, R. (ed.) (2004) *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*, 2nd edition (Opladen: Leske & Budrich). *evaNet-Positionen* 5/2004, pp.1-15. Available online: <http://www.evanet.his.de/evanet/forum/pdf-position/StockmannPosition.pdf> Accessed 14 June 2006

Stockmann, R. (ed.) (2006a) *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*, 3rd edition (Munster, New York, Munich & Berlin: Waxmann)

Stockmann, R. (2006b) *Evaluation und Qualitätsentwicklung. Eine Grundlage für ein wirkungsorientiertes Qualitätsmanagement* (Munster, New York, Munich & Berlin: Waxmann)

Stockmann, R. (ed.) (2007) *Handbuch zur Evaluation. Eine praktische Handlungs-anleitung* (Munster, New York, Munich & Berlin: Waxmann)

Struhkamp, G. (2005) Evaluation in Germany: An Overview, *Journal of Multi Disciplinary Evaluation*, Nr.3, October 2005. Available online: <http://evaluation.wmich.edu/jmde/content/JMDE003content/13_Evaluation>

Accessed 8 June 2006

Stufflebeam, D.L., (2001) *Evaluation models: New Directions for Evaluation*, No.89, Spring (San Francisco: Jossey Bass)

Swann, J., & Pratt, J. (eds.) (1999) *Improving Education: Realist Approaches to Method and Research* (London & New York: Cassell)

Szanto, T.R. (2005) Evaluations of the Third Kind: External evaluations of external quality assurance agencies, *Quality in Higher Education*, Vol.11:3, pp.183-193. Available online: <http://dx.doi.org/10.1080/135383205500329509> Accessed 18 September 2008

Teichler, U. (2005) Quantitative und strukturelle Entwicklungen des Hochschulwesens. In U. Teichler & R. Tippelt (eds.) (2005) Hochschullandschaft im Wandel, *Zeitschrift für Pädagogik*, 50. Beiheft, Oktober, pp.8-24

Teichler, U. & R. Tippelt (eds.) (2005) Hochschullandschaft im Wandel, *Zeitschrift für Pädagogik*, 50. Beiheft, Oktober

Thompson, B. & King, J.A. (1981) *Evaluation Types: Do Differences Matter*? paper at 65th AERA annual conference, Los Angeles, 13-17 April

Tippelt, R. (2004) Standards und Kriterien von Forschungsevaluationen in der Erziehungswissenschaft. In H. Merkens (ed.) (2004) *Evaluation in der Erziehungswissenschaft* (Wiesbaden: Leske & Budrich), p.92

Tippelt, R., Rauschenbach, T., & Weishaupt, H. (eds.) (2004) *Datenreport Erziehungswissenschaft 2004* (Wiesbaden: VS Verlag für Sozialwissenschaften)

Tippelt, R. (2008) Qualitätsstandards und Forschungsevaluationen in der Erziehungswissenschaft. In E. Klieme & R. Tippelt (eds.) (2008) Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz, *Zeitschrift für Pädagogik*, 53. Beiheft, (Weinheim & Basle: Beltz), pp.171-187

Tooley, J. (1999) The Popperian Approach to Raising Standards in Educational Research. In J. Swann & J. Pratt (eds.) (1999) *Improving Education: Realist Approaches to Method and Research* (London & New York: Cassell), p.169

Toombs, S. K. (ed.) (2002) Handbook of Phenomenology and Medicine. In *Philosophy And Medicine*, Vol.68 (Boston: Kluwer ACA)

Trotman, D. (2005) Looking for Sparks: A phenomenological study of educational practice and the imaginative lifeworld in primary schools, *Teaching Education*, Vol.16:1, No.1, pp.61-72

Ursin, J., Huusko, M, Aittola, H., Kiviniemi, U. & Muhonen, R. (2008) Evaluation and Quality Assurance in Finnish and Italian Universities in the Bologna Process, *Quality in Higher Education*, Vol.4:2, pp.109-120
Available online: <http://dx.doi.org/10.1080/13538320802278222>
Accessed 27 November 2008

Van Berkel, H.J.M. & Wolfhagen, H.A.P. (2002) The Dutch System of External Quality Assessment: Description and Experiences, *Education for Health*, Vol.15:3, pp.335-345

Van den Berg, J.H. (1972) *A different existence: Principles of phenomenological psychopathology* (Pittsburgh: Dusquesne University Press)

Van der Mescht (2004) Phenomenology in Education: A Case Study in Educational Leadership, *Indo-Pacific Journal of Phenomenology*, Vol.4:1, July

Van Kemenade, E., Pupius, M. & Hardjono, W. (2008) More Value to Defining Quality, *Quality in Higher Education*, Vol.14:2, pp.175-185. Available online: <http://dx.doi/10.1080/13538320802278461> Accessed 18 September 2008

Van Vught. F. (1997) The Humboldtian University under Pressure. New Forms of Quality Review in Western European Higher Education. In H. Altrichter et al. (eds.) (1997) *Hochschulen auf dem Prüfstand* (Innsbruck: Studienverlag), pp.48-87

Vedung, E. (2004) Evaluation Research and Fundamental Research. In R. Stockmann & R.M.S. Visser (2005) *Trends in Program Evaluation Literature: The Emergence of Pragmatism*, TCALL Occasional Research Paper No.5, Texas Center for Adult Literacy & Learning. Available online: <http://www.tcall.tamu.edu/orp/orp5.htm> Accessed 22 May 2006

Volkov, B. & Volkova, E. (2005) More than 'The Least Liked Person at a Cocktail Party': Ethnography of Program Evaluators. A paper presented at the *Joint Conference of the Canadian Evaluation Society and the American Evaluation Associatio*n, Toronto, Canada, 27 October, received from authors via personal email

Vouzas, F.K., & Gotzmani, K.D. (2005) Best Practices of the new ISO9000:2000 organizations on their road to business excellence, *The TQM Magazine*, Vol.17, pp.259-266

Vroeijenstijn, T. (1992) External Quality Assessment, Servant of Two Masters? The Netherlands University Perspective. In A. Craft (ed.) (1992) *Quality Assurance in Higher Education*. Proceedings of an International Conference, Hong Kong 1991 (London & Washington: Falmer), pp.97-108

Walford, G. (ed.) (1994) *Researching the Powerful in Education* (London: UCL)

Walford, G. (2001) *Doing qualitative educational research: a personal guide to the research* (London: Continuum)

Weiss, C. (1972) Utilization of Evaluation: Toward Comparative Study. In C. Weiss (ed.) (1972) *Evaluating Action Programs: Readings in Social Action and Education* (Boston: Allyn & Bacon), pp.318-326

Wenneras, C. & Wold, A. (1997) *Nature* 387, 22 May, pp.341-344

Widmer, T. (2006) Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird. In R. Stockmann (ed.) (2006b) *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*, 3rd edition (Munster, New York, Munich &

Berlin: Waxmann), p.8

Winter, M. (2002) Studienqualität durch Evaluation und Akkreditierung – vier Entwicklungsszenarien, *Qualitätssicherung an Hochschulen: Theorie und Praxis* (Bielefeld: HRK)

Witte, J. (2007) The changing political framework of quality assurance in German higher education: National debates in European context, *2nd European Quality Assurance Forum 'Implementing and Using Quality Assurance: Strategy and Practice*', 15-17 November, Rome, pp.48-53

Wolcott (1995) *The Art of Fieldwork* (Walnut Creek: AltaMira)

Yin, R.K. (2003) *Case Study Research: design and methods*, 3rd edition (Thousand Oaks/CA, London, New Delhi: Sage)

Zhou, Y. (2005) *Chinese Higher Education at the Turn of the Century: Undergraduate Education Evaluation* (Beijing: Higher Education Press)

The following are additional **INTERNET REFERENCES**

Bologna Declaration:
<http://www.bologna-berlin2003.de/pdf/bologna_declaration.pdf> Accessed 7 June 2006

Berlin communiqué (12 September 2003):
<http://www.bologna-berlin2003.de/pdf/Communique1.pdf> Accessed 7 June 2006

Centre for Higher Education Development:
<http://www.che.de> Accessed 4 April 2007

Die *Zeit*: <http://www.zeit.de/campus/2007/02/aufruf-lehre> Accessed 24 September 2007

DeGEval German Society for Evaluations: <http://www.degeval.de Accessed 2 July 2008

DeGEval evaluation standards (in English):
<http://www.degeval.de/calimero/tools/proxy.php?id=72> Accessed 2 July 2008

Dutch Evaluation Model: <www.tkk.fi/Misc/H3E/wg2/NLgeneral.doc>
Accessed 13 March 2007

European University Institute:
http://www.eui.eu/ProgrammesandFellowships/AcademicCareersObservatory/AcademicCareersbyCountry/Germany.aspx#GrossSalaries Accessed 11 May 2009

Evalag evaluation agency:
<http://www.evalag.de> Accessed 4 February 2007

Evalag evaluation report of PH Weingarten:
<http://www.evalag.de/pdf/Kurzf/Summary_WG_Gutachten.pdf> Accessed 4 February 2007

Evalag evaluation report *Erziehungswissenschaft*:
<http://www.mwk.baden-wuerttemberg.de/uploads/tx_ttproducts/datasheet/Erziehungswissenschaft.pdf> Accessed 5 April 2006

Evalag IQ:
<http://www.evalag.de/dedievl/projekt01/media/pdf/qm/eckpunkte_evalag_iq.pdf> Accessed 24 February 2008

Evaluation of *Akkreditierungsrat*:
<http://www.akkreditierungsrat.de/fileadmin/Seiteninhalte/Dokumente/Sonstige/Evaluationsbericht.pdf> Accessed 10 May 2008

Excellence Initiative (results of first round of competition):
<http://www.bmbf.de/pub/Exzellenzinitiative.pdf> Accessed 1 April 2007

Forschung und Lehre:
<http://www.forschung-und-lehre.de/pdf/08_98.pdf> Accessed 21 June 2006

German federal law on use of interview data:
<http://www.gesetze-im-internet.de/bundesrecht/bdsg_1990/gesamt.pdf> Accessed 5 May 2009

KMK decision on system accreditation:
<http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2007/2007_06_15-Grundsatzbeschluss-Systemakkreditierung.pdf> Accessed 11 November 2007

Leibniz Association:
<http://www.wgl.de> Accessed 9 September 2008

Programme for International Student Assessment (PISA):
<http://www.pisa.org> Accessed 25 June 2006

Research Assessment Exercise (history):
<http://www.rae.ac.uk/aboutus/history.asp> Accessed 6 May 2009

Research Assessment Exercise (changes in methodology):
<http://www.rae.ac.uk/aboutus/changes.asp> Accessed 09 October 2008

System accreditation:
<http://www.university-partners.de/veranstaltungen/veroeffentlichungen/>
"Erste Gehversuche mit der Systemakkredierung", conference Bonn, 13-14 September 2009

Taiwanese use of ISO9000:
<http://www.icdf.org.tw/web_pub/20080619084537%E7%AC%AC%E5%9B%9B%E7%AB%A02.pdf> Accessed 4 February 2008

Zentrale Evaluations- und Akkreditierungsagentur Hannover (ZEVA):
<http://zeva.uni-hannover.de> Accessed 15 September 2006

APPENDIX A

Evaluations of *Erziehungswissenschaft* in Germany until 2004

Tab. 7.1: Bestandsaufnahme von Evaluationsverfahren erziehungswissenschaftlicher Studiengänge (Stand 1.12.2003) [1]

Hochschule	Evaluationsverfahren
U Augsburg	Evaluationen i.A. des bayer. Staatsministeriums für Wiss., Forschung Kunst[2]
Otto-Friedrich-U Bamberg	Evaluationen im Auftrag des bayerischen Staatsministeriums für Wissenschaft, Forschung und Kunst
U Bayreuth	Evaluationen i.A. des bayer. Staatsministeriums für Wiss., Forschung u. Kunst
FU Berlin	Fachbereichsinterne Evaluation[3]; hochschulinterne Evaluation[4]
U Bielefeld	keine Evaluation
U Bonn	keine Evaluation
TU Braunschweig	Evaluation durch ZevA5
TU Chemnitz	keine Evaluation
TU Darmstadt	Fachbereichsinterne Evaluation
U Dortmund	Hochschulinterne Evaluation
TU Dresden	Fachbereichsinterne Evaluation
U Düsseldorf	keine Evaluation
U Duisburg-Essen, Standort Duisburg	keine Evaluation
U Duisburg-Essen, Standort Essen	Hochschulinterne Evaluation
Kath. U Eichstätt	keine Evaluation
U Erfurt	Hochschulinterne Evaluation; Evaluation durch den Wissenschaftsrat
U Erlangen-Nürnberg	Evaluationen i.A. des bayer. Staatsministeriums für Wiss., Forschung u. Kunst
U Flensburg	Evaluation im Auftrag der Landesrektorenkonferenz Schleswig-Holstein
U Frankfurt a. M.	Hochschulinterne Evaluation
PH Freiburg	Fachbereichsinterne Evaluation; Evaluation durch evalag6
Georg-August-U Göttingen	Evaluation durch die Wissenschaftliche Kommission Niedersachsen; Evaluation durch ZevA
U Greifswald	keine Evaluation
U Halle-Wittenberg	geplante Evaluation im Verbund mit Jena und Leipzig
U der BW Hamburg	Evaluation im Auftrag des Wissenschaftssenators
U Hamburg	Evaluation im Verbund Norddeutschen U; fachbereichsinterne Evaluation
U Hannover	Evaluation durch ZEvA; Evaluation durch Wiss. Kommission Niedersachsen
PH Heidelberg	Fachbereichsinterne Evaluation; Evaluation durch evalag
U Heidelberg	Evaluation durch evalag
U Hildesheim	Evaluation durch ZevA
U Jena	geplante Evaluation im Verbund mit Halle und Leipzig
U Kaiserslautern	keine Evaluation
PH Karlsruhe	Evaluation durch evalag
TU Karlsruhe	Evaluation durch evalag
U Kassel	Hochschulinterne Evaluation; fachbereichsinterne Evaluation
Christian Albrecht U zu Kiel	Fachbereichsinterne Evaluation; Evaluation im Verbund Norddeutscher U
U Koblenz-Landau Abteilung Landau	keine Evaluation
U zu Köln	Evaluation geplant
U Leipzig	geplante Evaluation im Verbund mit Halle und Jena
U Lüneburg	Evaluation durch ZEvA; Evaluation durch Wiss. Kommission Niedersachsen
U Magdeburg	Fachbereichsinterne Evaluation
U Mainz	Fachbereichsinterne Evaluation
U Mannheim	Evaluation durch evalag
U Marburg	Hochschulinterne Evaluation
Ludwig-Maximilians-U München	Evaluationen i.A. des bayer. Staatsministeriums für Wiss, Forschung u Kunst
U Münster	Evaluation geplant
U Oldenburg	Evaluation durch ZEvA
U Osnabrück	Evaluation durch ZEvA; Evaluation durch Wiss. Kommission Niedersachsen
U Paderborn	Hochschulinterne Evaluation
U Passau	Evaluationen i.A. des bayer. Staatsministeriums für Wiss, Forschung u. Kunst
U Potsdam	Evaluation durch Wissenschaftsrat
U Regensburg	Evaluationen i.A. des bayer. Staatsministeriums für Wiss, Forschung u. Kunst
U Rostock	Evaluation im Verbund Norddeutscher Universitäten
PH Schwäbisch-Gmünd	Evaluation durch evalag
U GH Siegen	Evaluation geplant
U Trier	Fachbereichsinterne Evaluation
U Tübingen	Fachbereichsinterne Evaluation; Evaluation durch evalag
U Ulm	Evaluation durch evalag
Hochschule Vechta	Evaluation durch ZEvA
U Würzburg	Evaluationen i.A. des bayer. Staatsministeriums für Wiss., Forschung u. Kunst
U Wuppertal	keine Evaluation

1 Datengrundlage sind die zweimaligen Befragungen der Delegierten des EWFT. In einer ersten Runde sind diese gebeten worden, Evaluationserfahrungen mitzuteilen und entsprechende Materialien zur Verfügung zu stellen. In einer zweiten Runde sind die Delegierten per standardisierten Fragebogen befragt worden. In den Auswertungen nicht erfasst sind einmalige und nicht systematische Evaluationen einzelner Lehrveranstaltungen. Des weiteren haben die „Evaluationen' des Expertenrates NRW keine Berücksichtigung gefunden, da diese von den Delegierten des EWFT sehr unterschiedlich in Hinblick auf ihre Erfassung als Evaluation bzw. als Verfahren, die nicht als Evaluation gewertet werden können, eingeschätzt wurden.

2 sog. Speck-Kommission und sog. Baumert-Kommission

3 fachbereichsinterne Evaluation = institut. Verfahren der (Lehr-)Evaluation auf Fachbereichs-/Fakultätsebene

4 hochschulinterne Evaluation = Evaluationsverfahren im Auftrag der Hochschulleitung bzw. Senat der HS

5 Zentrale Evaluations- und Akkreditierungsagentur des Landes Niedersachsen

6 Evaluationsverfahren Erz.wiss. in Baden-Württemberg mit einer landeseigenen Evaluationsagentur (evalag)

Table courtesy of Böllert & Tippelt (2004, p.142-143)

APPENDIX B

Map of 15 HEIs in evaluation of *Erziehungswissenschaft*

The map originally presented here cannot be made freely available via ORA because of copyright. It was sourced from "Evaluation der Erziehungswissenschaft an den Universitäten und Pädagogische Hochschulen des Landes Baden-Württemberg 2004 (Bonn: Lemmens Verlag), p.14

Map courtesy of *Erziehungswissenschaft* 2003-2004 published report, p.14

APPENDIX C

Informed consent for participation in a research project

Understanding evaluators: a study of evaluations of German higher education

Susan Harris-Huemmert, Department of Education, Oxford University, is conducting doctoral research on the work of external evaluators who participate in evaluations of higher education in Germany. The aim of the study is to find out how evaluators reach their decisions. Do their backgrounds play a part? Are they affected by the hierarchical structures of German universities and if so, how much? Do subject experts, who already have experience of earlier evaluations, offer more input than those who have less or no evaluatory experience? Does their cultural or social capital play a role? If so, does this have an impact on the outcome of the evaluation?

You are requested to contribute through interview and discussion to this study. Data will be digitally recorded and saved on a computer for the duration of the project. All data will be kept safe and not given to third parties. Data will also not be used for further research purposes.

Participation in this project is entirely voluntary. You may terminate participation at any point and without explanation.

Your anonymity will be maintained as far as possible. Quotations will only be used following your consent.

If you have questions or would like to be informed on the results of the study when it is concluded, please contact Susan Harris-Huemmert for further details: email: susan.harris-huemmert@new.ox.ac.uk or the Director of the faculty, Professor John Furlong: Telephone +44 1865 274023. If you have questions on the manner in which this project is being conducted, you may also contact Professor Ingrid Lunt (Director of Graduate Studies). Telephone: +44 1865 274023.

If you would like to consent to participate in this research project, then please sign below. Please keep a copy for your records.

_____

Participant Date

APPENDIX D

Questions for evaluators

Did you know about *evalag* before this evaluation?

How did you react when you were invited to become a member of the evaluation commission?

Do you have an idea why you were selected? (Have you had a similar experience before?)

Did you know other members of the commission before, as colleagues, personally, or through hearsay?

Was this more a help or a hindrance? How?

Which competences or personal characteristics to you believe evaluation experts should possess?

Did you have a clear idea before, what would be asked of you? Or not so much?

Did you occupy a particular role in the commission?

Your chairperson was a foreigner. How was this?

What did you think of the commission's modus operandi, how it was put together? And its development?

During the evaluation evalag provided you with institutional self appraisals. How did you find these?

What was it like being commissioned into this evaluation by the Ministry of Science, Research and Art? Apart from specific requests by the Ministry, what was it like having a permanent representative on board?
How did the decision-making culture in the evaluation come about? What or who played a role?

Do you think that everyone's values were similar, or did you sense that these occasionally diverged? If so, why?

Were there specific problems that needed addressing? Or could these not be resolved?

How far were the self appraisals accurate presentations of what you encountered onsite?

As you are an educationalist, how did you feel about what you saw onsite? Were you surprised by anything in particular?

As all of the commission's members were experts in slightly different areas, do you think everyone was able to bring their expertise into the evaluation equally? Or were there marked differences? (If so, how were these dealt with?)

When the onsite reports were being prepared, there might have been differences of opinion and also during discussions with pro vice-chancellors. Were you satisfied with the final report?

How would you describe the role of the evaluation's organiser?

In retrospect, would you wish to conduct the evaluation in the same way or would you modify your work in any way?

Did you have any particular worries during the evaluation? If so, what was the reason for these? (practical/organisational))

What have you learnt personally from this experience?

Normally you yourself are employed by an institution. Has the fact that you placed judgement on other institutions had an impact on the way you perceive your own institution, or the way you deal with it?

Would you do another evaluation again?

Have you been able to follow events in educational science in Baden-Württemberg since this evaluation?

APPENDIX E

Interview participants, place, and date of interview

Fölling-Albers	Regensburg	25.10.06
Mandl	Munich	06.12.06
Röbbecke	Mannheim	09.01.07
Ossner	Weingarten	13.02.07
Beck	Weingarten	13.02.07
Baumert	Berlin	10.07.07
Tenorth	Berlin	16.07.07
Horstkemper	Potsdam	17.07.07
Gogolin	Hamburg	18.07.07
Schelten	Munich	24.07.07
Simmel-Joachim	Wiesbaden	30.07.07
Prenzel	Kiel	16.08.07
Terhart	Münster	17.08.07
Erichsen	Münster	17.08.07
Achtenhagen	Göttingen	12.09.07
Wijnen	Lüneburg	16.09.07
Phillips	Oxford	28.09.07
Fornefeld	Cologne	23.10.07
Schomann	Mannheim	26.10.07
Hammer	Mannheim	26.10.07
Schreier	Gütersloh	19.12.07
Bolius	Stuttgart	10.04.08
Lübke	Stuttgart	10.04.08
De Corte	Oxford	08.07.08
Albers	Schwäbisch Gmünd	29.10.08
v.Grävenitz	Constance	30.10.08

Visits to *evalag* archives

Mannheim	18.07.06
Mannheim	14.08.06
Mannheim	26.10.06
Mannheim	02.01.08

APPENDIX F

QUESTIONNAIRE A (final version)

1. Structures and quantitative data

1.1. Brief synopsis

1.1.1. Please forward (already available) documents which can inform us about the historical development of Educational Science in your institution. Please include structural plans.

1.1.2. How is the subject organised and bound into your institution (faculty/subject area/institute/position of subject didactics)?

1.2 Study programmes, students, examinations

1.2.1 Which courses do you offer? Please indicate terminating courses and provide end date (table 1A or table 1B).

1.2.2 Wherein are the strengths of the courses you provide?

1.2.3 Are there, or have there been, restrictions in admission onto specific Educational Science courses? What did these restrictions consist of?

1.2.4. Please complete Tables 2 to 6 (freshmen, students, examinations, average length of course, number of doctorates, *Habilitation*) [287]

1.2.5. How do you observe operational courses and if so gauge their success?

1.2.6. In which areas do graduates in Diploma and Masters degrees work in? How do you gain information on the career paths of graduates?

1.3 Personnel

1.3.1 In order to clarify personnel structures (of the subject and the institute), please complete Tables 7-10 (in the statistical section).

1.3.2 Please indicate professorial honours and achievements (Table 11)

1.3.3 Is research in Educational Science subdivided into research groups? If so, please provide divisions (e.g. research programmes or working groups) and complete Overview 2 and Table 12.

1.3.4. Explain absences.

1.3.5. Which professorships will become available and when? Does the discipline or institute intend to reallocate these? Please provide reasons for this.

1.3.6. How would you rate the human resources in your subject (number, age structure, job structure)?

1.4 Allocation of finances, space and materials

1.4.1 Please provide an overview of income development over the past 5 years (Table 13).

1.4.2 Please present allocation of space (Overview 1) and explain exceptions.

1.4.3 How does your teaching staff rate the allocation of finances, space and resources?

[287] The *Habilitation* is peculiar to Germany in that in order to become a professor it is necessary to write the equivalent of a second doctorate in an unrelated area.

2 Research

2.1 Research foci and organisation

2.1.1 Please describe the research programme in Educational Science and include explanatory documents if necessary.

2.1.2. Please describe the research project work currently being undertaken.

2.1.3. Which research strategy does the institute pursue (decisions on projects, funding applications etc.)?

2.1.4. Which development programme (school development etc.) does the institute pursue?

2.1.5. How are students integrated into research and empirical (research) methods in particular?

2.2 HEI internal cooperations

2.2.1. Which research cooperation exists with other disciplines in your institution?

2.2.2. Please describe the types of cooperation. Are there joint programmes, and in which research areas?

2.3. External cooperations

2.3.1. With which HEIs or external research institutions have research projects been conducted (or funding applications placed) over the last 5 years (for example, with the DFG)? What was the aim of these projects?

2.4. Visiting scholars

2.4.1 How many scientists/researchers (from Germany and abroad) visited your institution and were hosted by Educational Science for more than a few days over the last 5 years?

2.5. Events and congresses

2.5.1. Which significant scientific events were carried out over the last 3 years?

2.6. Status quo and perspectives for research

2.6.1. Are there any serious restrictions preventing your institution from carrying out efficient research?

2.6.2. Which positive and supportive frameworks would you like to point out?

2.6.3. Which are currently the most important research topics in Educational Science?

2.6.4. How would you rate the research output of Educational Science in your HEI in international comparison?

2.6.5. What are the interim plans for your discipline's development over the next 5 years? Which (structural and contextual) aims in the research area should be particularly promoted by your discipline and HEI?

3. Educational Science

(Main degree programme, foundation degree programme, Diploma, Magister, BA & MA)

3.1 Into which stages of study are the different courses organised and which aims and conceptual key areas are related to this?

3.1. Which areas of specialisation are offered?

3.2. Are your aims and concepts specific to your faculty and do you liaise with other HEIs on the conceptualisation and further development of these aims?

3.3. Which compulsory, optional and open events does the foundation and main study course provide? Which events are offered in the foundation course? *Please include the respective course and exam requirements.*

3.4. Which contents are conveyed within the core curriculum? Is there coordination between the courses in teacher training and other diploma courses?

3.5. How are the different courses structured (e.g. through timetabled courses or modules)? Please provide additional information if available.

3.6. How are courses link to exam requirements (e.g. via comments in course study booklets)?

3.7. Which didactics of higher education or innovations in teaching methods are being used? Are programmes of higher educational didactics on offer and how are they constructed?

3.8. Are there indications of teaching research/learning research? Do students have the possibility to take part in current research projects?

3.9. To what extent are students prepared for the demands of later practice (teaching opportunities, internships etc.)?

3.10. In which phases of different courses are internships envisaged? Are internships obligatory? What is your concept and how are these internships supervised?

3.11. Please describe the type and amount of international cooperation in teaching (e.g. internships, international examinations). How are students prepared for this (foreign language skills)? How many students have taken part and with what success over the last 5 years? How many students have taken part in EU programmes (e.g. Socrates, Erasmus)?

4. Teacher training

4.1. In which phases of study are the different courses integrated? What are their aims and conceptual targets?

4.2. Which areas of specialisation do you offer? Do you liaise on these with other HEIs?

4.3. Which compulsory, optional and open events are provided for the foundation and main courses in teacher training programmes? Which events are offered in the preparatory course? *Please include the respective course and exam requirements.*

4.4. Do you have a core curriculum for Educational Science in your institution? What does this consist of? Are combinations of the core curriculum envisaged with Diploma and main course study programmes?

4.5. Is the order of teaching courses prescribed in the different degree programmes? How is this made transparent to students (e.g. course plans)?

4.6. Can you see problems in the transfer between foundation and main degree programmes, in particular caused by a change of institution? If so, what are these? Have supportive measures proven successful?

4.7. How do courses relate to exam requirements (e.g. via comments in course study booklets)?

4.8. Which didactics of higher education or innovations in teaching are being used? Are teaching development programmes on offer and how are they organised?

4.9. Do you have research on teaching and learning research? Do students have the possibility to take part in current research and development projects?

4.10. To what extent are students prepared for the demands of later practice (teaching opportunities, internships etc.)?

4.11. Please describe the type and amount of international cooperation in teaching (e.g. internships, international examinations). How are students prepared for this (foreign language skills)? How many students have taken part and with what success over the last 5 years? How many students have taken part in EU programmes (e.g. Socrates, Erasmus)?

4.12. What is the quantitative contribution of Educational Science in the training of teachers?

4.13. To what extent is there curricular and organisational liaison between the study programme of Educational Science, its related subjects and subject didactics?

4.14. In which phases of the different courses are internships envisaged? How are these internships supervised? How are these internships coordinated, and who is responsible?

4.15. Is it possible to take the State Examination in Educational Science? How many theses were written over the last 3 years?

4.16. To what extent is school practice a criterion for appointments to teach in courses that offer teacher-training?

5. Universal questions

5.1 Supervision and counselling

5.1.1 Are information events on offer to freshmen? What is the nature of these events?

5.1.2 What is the extent of information and advice to students at the different stages of their degrees? Please differentiate between general and subject-specific advisory provision.

5.1.3 Who is the point of contact for students taken Educational Science degrees (Diploma, Magister) and in the teacher-training courses?

5.1.4 Please provide the internet addresses under which students can find information.

5.1.5 Which difficulties or supervisory problems do students most frequently report back on? How can the discipline or the institution help alleviate these problems?

5.1.6 Which particular forms of supervision are offered by Educational Science (e.g. tutorials, mentoring etc.)?

5.1.7 Do you have programmes for international students?

5.1.8 What support are students wishing to study abroad given?

5.2. Academic development

5.2.1 What development programmes do you have in place for early career academics (e.g. doctoral studies and/or graduate colloquia)?

5.2.2 How do doctoral students fund themselves?

5.2.3 What strategies do you have to support early career academics and which are you planning? How are doctoral students incorporated into teaching and research (e.g. integration into running projects?)

5.2.4 Are you planning on introducing junior professors?

5.3 Support of women

5.3.1 Do you see particular necessities to support female students and early career academics? What possibilities do you see to make Educational Science more attractive to men?

5.3.2 Which specific provision does your institution have to enable the communion of family and career, or family and studying (e.g. nursery, part-time courses)?

5.4 Further education

5.4.1 Which further education programmes does Educational Science offer?

5.4.2 Which teaching staff is involved in this?

5.4.3 Which media are used in programmes of further education?

5.4.4 Do you offer further education programmes together with cooperation partners?

5.4.5 How do you cooperate with the State Institute for Education and Teaching (Landesinstitut für Erziehung und Unterricht)? How do the programmes you provide differentiate from each other?

5.5 Status and perspectives in teaching, degree courses and further education

5.5.1 How does your discipline (institute) rate its own achievements in teaching, degree courses and further education?

5.5.2 What are the interim plans of your discipline for developments over the next 5 years? Which plans regarding teaching, degree courses and further education should be implemented as a matter of urgency by your discipline/institution?

5.6 Quality of teaching and research

5.6.1 Please indicate the methods you use to evaluate teaching and research (student statements, graduate questionnaires etc.)?

5.6.2 How are teaching reports compiled by your institution? Who is responsible for them? Please provide the latest example if possible.

5.6.3 How do you ensure the quality of teaching and research at the junction of teacher training and Educational Science in its wider context?

5.6.4 How do you rate your discipline's achievements in teaching and research?

5.7 Current points of concern

5.7.1 How has your discipline discussed international studies which judge the capabilities of educational systems and their consequences (in organisational events, with students)? Which consequences need to be drawn in relation to teacher education?

5.7.2 What conclusions in relation to future research activities has your discipline drawn from the debate on PISA results and similar studies?

QUESTIONNAIRE B (final version)

1 Funding

Please indicate the degree of funding according to each funding body.

1.1 Please summarise the aforenamed funded projects.

1.2 Please list the type, frequency and duration of all posts, which have been/are being funded by external bodies.

2 Publications

2.1 Please compile a list of all publications produced by your institute over the last 5 years, differentiated into the following categories:

- Publisher
- Monograph
- Chapters in regularly appearing journals/volumes
- Papers in internationally peer-reviewed scientific journals
- Papers in German peer-reviewed scientific journals
- Papers in non-reviewed scientific journals
- Papers in journals produced by professional bodies

2.2 Please list the three most important publications of your chief scientific staff (alphabetically differentiated according to author)

3 Prizes and awards

Please list all prizes and awards given over the last 5 years (listed alphabetically in the order of named individuals)

4 Sabbaticals

Which important domestic or foreign institutions did your main staff visit over the last 5 years (differentiated according to person and institution)? Please name all institutions listed in questionnaire B.

l

5 Day events and congresses

Please list all day events and congresses (differentiated according to persons involved)

6 Scientific positions and functions

Please list all named positions and functions (differentiated according to persons involved)

7 Examinations

Please list the extent of examining over the past 3 years

8 Teaching duties

Please list the extent of teaching duties

9 Participation in staff development

Please describe the extent of staff engagement in further and continuing development (differentiated according to persons involved/activities.

APPENDIX G

Sample interview

I: Kannten Sie überhaupt die Arbeit von *evalag*, bevor Sie damit anfingen? 00:00:17-5

H: Ich wusste, dass es diese Agentur gibt, aber ich hatte bis dahin noch nie Kontakte dazu und kannte auch das Konzept nicht näher. 00:00:26-5

I: Ja. Und wie kamen Sie dann zu der Kommission? Wurden Sie angeschrieben? 00:00:33-8

H: Es kam ein Anruf, von der Kollegin, die das Ganze betreut hat. Und... ja, mit der Frage, doch mal darüber nachzudenken, ob ich mir vorstellen könnte, in einer solchen Kommission, also, der Auftrag wurde beschrieben, sämtliche an der erziehungswissenschaftlichen Ausbildung beteiligten Institutionen in Baden-Württemberg, also an der Lehrerbildung beteiligten Institutionen, soweit es dann eben den erziehungswissenschaftlichen Anteil der Ausbildung angeht, zu evaluieren, und ob ich mir vorstellen könnte, da mit zu arbeiten. Und das fand ich eine ganz interessante Sache. Und habe dann einmal darüber geschlafen und dann gesagt, ja, das möchte gerne machen. Wobei mich eben besonders interessiert hat, die Tatsache, dass es Baden-Württemberg ist, weil es ja das einzige Bundesland ist, in dem es auch noch die Pädagogischen Hochschulen gibt, während die in allen anderen Bundesländern integriert sind inzwischen in die Universität. Und mich hat sehr interessiert, gibt es da, lassen sich da Unterschiede beobachten, in den jetzigen Konzeptionen und was kann man daraus lernen. 00:01:49-8

I: Ja. Also, das war ein Reizpunkt…00:01:52-2

H: Das war die Motivation. 00:01:53-2

I: …wo Sie gesagt haben, da möchte ich unbedingt jetzt rein… 00:01:55-9

H: Ja, ganz genau. 00:01:57-1

I: …weil das ein anderes System ist. Hatten Sie schon vorher an Evaluationen ähnlicher Art teilgenommen, oder war das dann in dieser Art das erste Mal, oder... 00:02:07-6

H: Ich war zu diesem Zeitpunkt schon seit mehreren Jahren Mitglied der Strukturkommission der Deutschen Gesellschaft für Erziehungswissenschaft, das ist die DGFE, und zu deren Aufgaben gehört unter anderem die Beratung von Hochschulen in solchen Fragen. Wenn die etwa darüber

nachdenken, wie Studiengänge vielleicht umorganisiert werden könnten, oder in welcher Weise das Konzept an der eignen Institution noch mal reflektiert und durchdacht werden könnte. Im Rahmen dieser Tätigkeit ist mir so etwas begegnet und mich hatte die Universität Halle, hier im Nachbarland Sachsenanhalt auch gebeten, sie zu beraten, gemeinsam mit einem Kollegen, der ebenfalls dort in der Strukturkommission mitgearbeitet hat, mit Rudolf Tippelt. 00:03:01-4

I: Ah ja. Okay. 00:03:05-2

H: Das war vorangegangen. Und ansonsten ist mein, bin ich ja schon seit vielen Jahren so im Bereich wissenschaftlicher Begleitung von Modellversuchen im Bildungsbereich. Das ist etwas verwandtes, aber liegt nicht. Wir beraten da eher Schulen, oder begleiten eher Schulen, weniger Hochschulen. 00:03:28-7

I: Kannten Sie dann andere Mitglieder schon? 00:03:36-3

H: Ja, also, für mich war schon ganz wichtig. Also, die Bitte war direkt, ob ich mir vorstellen könnte, den stellvertretenden Vorsitz zu übernehmen. 00:03:42-9

I: Das kam gleich am Anfang? 00:03:44-4

H: Das kam gleich zu Anfang. Mit dem Hinweis darauf, dass Jürgen Baumert den Vorsitz übernehmen wird, und das völlig klar ist, dass das der Leiter des Max-Planck-Institutes und der damals ja noch ganz stark in die Pisa-Forschung eingebundene Jürgen Baumert. Dass da eine Stellvertretung eben auch mit viel Arbeit verbunden sein würde. Gleichzeitig arbeite ich aber sehr gerne mit ihm zusammen. Das war mit auch ein Grund, da rein zu gehen. Und es gab vor allen Dingen auch die Möglichkeit, noch Mitglieder zu benennen und insofern hatte ich vorgeschlagen, den Kollegen Beck aus der Schweiz mit aufzunehmen. Wen kannte ich denn noch? Also, es war eine, durchaus eine Mischung. Ich kannte viele nicht, viele Mitglieder. Es war ja auch eine große Kommission, aber konnte eben auch noch mit Vorschläge machen. 00:04:45-5

I: Fanden Sie das... oder... 00:04:49-2

H: Der Kollege Tenorth war auch schon klar als Mitglied... 00:04:52-6

I: Zu dem Zeitpunkt, als Sie. ? 00:04:54-0

H: Zu dem Zeitpunkt. Und den kenne ich natürlich auch. Terhart war auch Mitglied. Also, es waren schon einige, mehr aus der überregionalen Zusammenarbeit, aus den Bereichen Lehrerbildung

vertraute Kollegen dabei, die ich auch sehr schätze. 00:05:12-5

I: Ja. Und würden Sie sagen, dass es insgesamt für so eine, so eine Arbeit immer wichtig ist, dass man immer einen Stamm, sozusagen, hat die sich gut kennen vorher. Oder würden Sie sagen, es würde auch genauso gehen, vielleicht, ja, ich weiß es nicht... 00:05:29-5

H: Wichtig ist es, dass es ein arbeitsfähiges Gremium wird und dazu braucht man, glaube ich, dazu ist es zumindest eine erleichternde Bedingung, wenn es einige Menschen gibt, von denen man weiß, die können gemeinsam arbeiten. Das heißt nicht, dass die sich immer einig sein müssen, ganz im Gegenteil, das kommt ja auch drauf an, dass man sich nicht zu früh einig ist, aber es müssen hinreichend integrationsfähige Menschen zusammenkommen, wo dann nicht so Selbstdarstellungsbedürfnisse und konkurrenzmäßige Darstellungen irgendwann die Sachfragen überlagern. Das finde ich schon wichtig. 00:06:12-2

I: Ja. Ja, ja. Und in diesem Fall war das ja etwas Besonderes, eben weil das vom Ministerium beauftragt wurde. 00:06:18-4

H: So ist es, ja 00:06:18-4

I: Wie war das für Sie? Oder wie hat sich das manifestiert? 00:06:25-7

H: Es gab einen Auftrag, von Anfang an und es gab ganz zu Anfang eine sehr klare Abgrenzung auch noch mal, wo Herr Baumert und ich noch einmal einen sehr deutlichen Brief verfasst haben, an das Ministerium, um die Rolle der ministerialen Vertreter, beziehungsweise, assoziierten Mitglieder dieser Kommission zu klären, weil es natürlich klar ist, dass wenn ein Ministerium einen Auftrag erteilt, dass es da einen Informationsfluss geben muss. Es muss aber auch geklärt werden in solchen Gremien, dass es wirklich unabhängige Gremien sind, und dass da nicht eine, ich sag mal, Gefälligkeitsbegutachtung bei rauskommt, sondern, die Kommission muss die Freiheit haben, wirklich an alle Informationen heranzukommen, und in ihrer Arbeit auch nicht durch irgendwelche Beschränkungen schon in eine bestimmte politisch wünschenswerte Richtung gedrängt zu werden. Also diese Unabhängigkeit des Gremiums zu garantieren und zu sichern das fand ich in der Anfangsphase ganz wichtig. Und das haben wir da auch geklärt. Also, das ist dann die Frage, wo braucht die Kommission beispielsweise dann auch mal Zeiten, wo die Ministerien, ministeriellen assoziierten Mitglieder nicht teilnehmen, sondern wo die Kommission unter sich ist. Und wo wir dann aber gleichzeitig aber wieder der Informationsfluss gesichert. 00:07:56-6

I: Ja. Wie war das vom Verhältnis? Inwiefern waren die Ministerien, Ministerienmitglieder meistens dabei, oder kann man prozentuell sagen, vielleicht achtzig Prozent oder sechzig Prozent

dabei? 00:08:08-3

H: Nein, ich würde schon denken, mindestens achtzig Prozent. Sie waren ja auch beteiligt an den Begehungen, die wir überall gemacht haben. Es gab einige interne Beratungen und die lagen nicht höher als zwanzig bis dreißig Prozent, würde ich mal so schätzen. 00:08:26-9

I: Und war das gut, weil sie vom Ministerium natürlich die Institutionen in ihrem Bundesland, ich gehe davon aus, sehr gut kannten. War das andererseits für die Kommission dann von Vorteil, sie dabei zu haben, in manchen Fällen, dass man sagen konnte, oder speziell Fragen ans Ministerium dann direkt richten konnte? 00:08:50-0

H: Das war sehr wichtig, um Hintergrundinformationen zu bekommen, um auch noch mal Klärungsaufträge deutlicher formulieren zu können, und um nachhaken zu können. Also, das war vom Informationsfluss her auf jeden Fall sinnvoll. Wir hatten ja die Besonderheit, dass es zwei Ministerien gibt, sowohl das Wissenschaftsministerium als auch das Bildungsministerium. So, dass man immer mit beiden Seiten zu tun hatte. 00:09:18-2

I: Hat es da Probleme gegeben, manchmal, eben weil das verschiedene Perspektiven gibt, oder... 00:09:23-7

H: Also, ich habe da Interessenskonflikte nicht bemerken können. Also, es gibt deutlich unterschiedliche Aufmerksamkeitsrichtungen, natürlich, weil die Arbeitsgebiete unterschiedlich sind, aber das Wissenschaftsministerium war stärker auf den Teil Lehrerbildung ausgerichtet, das Bildungsministerium noch stärker so auf Richtung Schulqualität und in welcher Weise muss das denn, durch die Art und Weise, der Anlage der Ausbildung, wie weit kann so etwas befördert werden? Also, das waren Akzentverschiebungen, aber kein Interessengegensätze. 00:09:56-9

I: Okay. Wenn Sie jetzt speziell die Arbeit als Evaluatorin betrachten, was würden Sie sagen, sind Kompetenzen oder Fähigkeiten, die man unbedingt haben muss, wenn man in so eine Arbeit hineinsteigt? 00:10:16-1

H: Also ich denke, nützlich ist gerade in unserem föderalen System in der Bundesrepublik, dass man einigermaßen einen Überblick hat über die verschiedenartigen Regelungen, damit man nicht seine Arbeitswelt, in der man selber steckt, also in meinem Fall die Brandenburger Wirklichkeit, in der ich jetzt gerade arbeite, damit man die nicht für den Normalfall hält. Sondern damit man also weiß, an welchen Stellen muss man jetzt nach den Variationen suchen und nach den spezifischen Ausprägungen, also, eine Übersicht über die Bildungslandschaft und die Vielfalt der geltenden Regelungen und das möglichst auch nicht nur auf das eigene Land, sondern… Ich fand sehr, sehr

hilfreich die Mitarbeit ausländischer Kollegen und Kolleginnen, um dann noch mal den Blick zu weiten, und zu schauen, wie wird das eigentlich woanders gehandhabt und deren Frageperspektive auch noch mal mit aufzunehmen. Das war sehr hilfreich. Und die, wenn ich das jetzt auf die Kompetenzebene bringe, die Offenheit eben auch solche anderen Variationen, erst einmal nicht nur als denkbar, sondern als möglicherweise auch Erweiterung des eigenen Blicks zu betrachten. Und die eigene Normalität nicht zu verabsolutieren, das finde ich eine sehr wichtige Qualifikation von Evaluatoren.

Eine zweite ganz wichtige Fähigkeit für Evaluatoren, finde ich, ist die Perspektivübernahme. Wie ist das, wenn man selbst evaluiert wird, mit welchen Ängsten, oder mindestens ja, strategischen Überlegungen ist das immer auch bei denjenigen verknüpft? Welche Bedürfnisse die eigene Identität auch ein Stück zu stützen und die eigene Arbeit positiv darzustellen und den neugierigen Blick von außen dann eben nicht nur als Kontrolle, - natürlich wird der immer als Kontrolle wahrgenommen, da muss man sich drüber im Klaren sein - aber die Fähigkeit dann zu haben, wie muss ich jetzt mein Verhalten als Evaluator, als Evaluatorin möglichst gestalten, dass ich eben nicht nur als der „defizitorientierte Kontrollblick von außen“ wahrgenommen werde, sondern als jemand, der erst einmal verstehen will, wie laufen hier bestimmte Dinge, was sind die Begründungen dafür, welche Varianten sind vielleicht auch abgewogen worden bereits, und in welche Richtung könnte man sich perspektivisch vorstellen, etwas zu verändern? Also, die schwierige Aufgabe, jemanden zu evaluieren und den gleichzeitig fröhlicher darüber werden zu lassen, dass er jetzt auch unter eine kritische Lupe genommen wird, den Weg, glaube ich, gibt es noch nicht. Wenn wir den hätten, das wäre nobelpreisverdächtig. 00:13:12-9

I: Das wäre ja eine ideale Welt! (Gelächter) 00:13:14-2

H: Aber da eben mit dem notwendigen Taktgefühl ranzugehen und gleichzeitig aber auch, deutlich genug, ja, Sachverhalte zu benennen, die man entweder kritisierenswert oder überlegenswert findet, und dann auch darauf zu beharren, dass man da jetzt noch weitere Informationen, Detailinformationen braucht, sich nicht so schnell mit so „Fassaden-präsentationen“ zufrieden zu geben. Diese Balance zu halten, das finde ich, ist eine wichtige Qualifikation für alle diejenigen, die so eine Evaluation auf Augenhöhe machen wollen. 00:13:47-8

I: Also, ist das eine Art, eine Kombination aus Diplomatie und Kritik? Könnte man das so vielleicht formulieren? 00:13:56-4

H: Hmmm, ja. Diplomatie und Kritik widersprechen sich nicht. 00:13:59-0

I: Eben. 00:14:01-0

H: Es ist eine Balanceleistung, das sehe ich auch so. Es ist eine schwierige Abgrenzung. Was ist jetzt eigentlich auch schon Beratung? Wir haben ja Empfehlungen gegeben. Also, insofern kommt der Beratungsaspekt noch dazu und die Schwierigkeit ist aber eben immer, Beratene erst einmal aufzuschließen für das, was zu beraten ist, und dazu muss man einerseits immer den Beratenen vermitteln, dass man über bestimmte Kompetenzen verfügt und bestimmte Dinge auch vergleichend einordnen kann, eben weil man sich intensiv mit der Materie befasst hat. Man muss nachweisen können, dass man sich auf die regionalen und lokalen Besonderheiten hinreichend einlässt, und dass man gemeinsam dann versucht, noch mal zu überlegen, in welche Richtung könnte man denn bestehende Ansätze entweder verstärken, oder aber auch umsteuern und da muss man den Mut haben, auch zu sagen, an welchen Stellen man wirklich jetzt, ja, auch radikal kritisiert, den Finger in offene Wunden zu legen, muss man schon trauen. Sonst ist man als Evaluator eine Fehlinvestition. 00:15:10-1

I: Man muss auch mutig sein. 00:15:11-1

H: Ja, genau. Also, wer nur auf Akzeptanz und geliebt werden aus ist, das ist keine gute Rolle dann. 00:15:22-5

I: Aber eben, das auch die, weil Sie kommen ja als Kollegin. 00:15:26-2

H: So ist es. Und als eine Kontrollinstanz. 00:15:28-7

I: Das ist auch so eine... 00:15:30-1

H: Und das ist der Vorteil und der Nachteil, weil man auf der einen Seite als Kollegin auch weniger gefürchtet ist. Wir sind eben kein staatliches Kontrollgremium, was direkt mit Anweisungen oder so irgendwo eingreifen könnte, oder über Ressourcenvergabe oder ähnliches verfügt. Mittelbar ist es natürlich trotzdem nicht unwichtig, was ein solches Evaluationsteam äußert und es ist immer eine Bewertung der eigenen Arbeit, die da gemacht wird. Und insofern ist man als Berater natürlich auch nicht in der Lage, in der man ansonsten ist, wenn freiwillig Expertise eingeholt wird, wie das etwa bei den Beispielen, die ich zu Anfang zu meinem Hintergrund genannt habe, wenn uns eine Institution von Außen direkt holt und sagt, bitteschön schaut euch bei uns um, und wir erwarten von euch Ratschläge. Dann muss man vieles an Vorarbeiten gar nicht so sehr leisten, sondern dann wissen die häufig auch schon um Probleme und man bemüht sich gemeinsam um eine Klärung. In diesem Falle war es eben ein von Außen eingesetztes Gremium und dann muss dann erst so eine Arbeitsbasis hergestellt werden. Und dazu gehört eben ein gehöriges Maß an kommunikativen Fähigkeiten und an Sachkompetenz und das muss dann beides zusammentreffen, Idealerweise.

00:16:56-7

I: Haben Sie von Anfang an gespürt, dass die Institutionen eher der Sache positiv gegenüberstanden, oder war da eine absolute Abwehrhaltung am Anfang zu spüren, die sich dann erst gelegt hat, oder was waren so die ersten Eindrücke von den Institutionen? 00:17:15-1

H: Das waren ja sehr viele Institutionen, sehr viele Standorte und da waren durchaus Variationen zu finden. Das hängt stark ab, vom Stellenwert auch der Universitätsspitze, die überhaupt diesen Bereich in den Universitäten oder, zugemessen werden. Bei den PHs ist es eindeutiger. Da ist die Aufgabe Lehrerbildung und bei den Universitäten ist es auch in Baden Württemberg nicht sehr viel anders als in den meisten Bundesländern, dass die Lehrerbildung, oder überhaupt die erziehungswissenschaftliche Ausbildung auch in den Hauptfachstudiengängen dort, dass das in aller Regel nicht so die „Prestige-Kinder" sind der Universitäten, sondern eher so etwas wie Pflichtaufgaben. Da haben wir Erfahrungen aus der Integration der PHs in diese Bereiche und da ist… Baden Württemberg unterscheidet sich da nicht sehr von anderen Bundesländern. Es gab aber durchaus deutliche Unterschiede an einzelnen Standorten. Bei den Pädagogischen Hoch-schulen hatte ich schon den Eindruck, dass da stärker die Hoffnung dran geknüpft war, dass jetzt ein objektivierender Blick von außen, vielleicht auch noch mal die besonderen Probleme, Belastungen, Überlastungen und Ausstattungsmängel zur Kenntnis nehmen und mit Nachdruck auch formulieren könnte. Das waren Erwartungen, die da mehr oder weniger deutlich formuliert zu Tage traten.

Bei den Universitäten habe ich also solche Hoffnungen weniger wahrgenommen, sondern da hatte ich bei einigen Universitäten deutlich den Eindruck, dass das in der Universitätsspitze nicht im Vordergrund des Interesses stand und bis hin zu einer Universität, die im Laufen, mit der es dann zum Teil relativ konflikthaft verlief, die eigentlich erst nachdem die Begehungen stattfanden und wir dann noch mal über die, das Resultat der Begehungen, ja, noch mal in gemeinsamen zentralen Veranstaltungen auch diskutiert haben, da brach dann ein Stück weit Konflikt auf, weil unsere schriftlichen Rückmeldungen dort mit Missfallen zur Kenntnis genommen wurden. Da tauchte dann erst, im Laufe des Verfahrens gab es dann aber, nachdem so der erste Unmut geäußert worden war und wir uns darauf eingelassen haben, aber auch durchaus nicht zurückgewichen sind, an Stellen, wo wir fanden, da muss man jetzt ganz klar eine kritische Position benennen. Also, da wo sämtliche erziehungswissenschaftlichen Stellen einkassiert werden. An einer Universität. 00:20:09-9

I: War das (name of institution)? 00:20:12-4

H: Zum Beispiel. Das war ein heftiger Konfliktfall. Da muss man sich dann auch fragen dürfen, ob denn, wenn diese Stellen alle wegfallen... 00:20:21-9

I: Was passiert! 00:20:21-9

H: .... kann man da Ausbildung überhaupt noch verantworten und dann muss man das laut und deutlich so benennen und da, denke ich mir, ist dann auch etwas in Bewegung gekommen. Also, das war sehr unterschiedlich. 00:20:32-2

I: Aber da, ich denke, das, wenn man das so offensichtlich sieht, dann eben, dann muss man dran bleiben 00:20:36-4

H: Ja. Klar. 00:20:38-1

I: Man kann das einfach nicht gehen lassen. Okay. 00:20:42-5

H: Da fand ich zum Teil auch wirklich die Mitarbeit auswärtiger Kollegen noch einmal sehr hilfreich. Weil die selber interessenunverdächtig sind. Als Kollege aus der deutschen Szene ist man ja immer ein Stück weit in diesem konkurrenzorientierten, durch Ranking- Vergleiche angefeuertem Klima, ein Stück weit mit betroffen und da ist die Mitarbeit von auswärtigen Kollegen noch einmal sehr, sehr hilfreich. 00:21:10-4

I: Haben die dann in dem Moment eine spezielle Rolle übernommen? Oder einfach, weil sie nun mal diese, wenn man so will, fremde Perspektive mit eingebracht haben, oder konnte man dann sagen, es wäre gut, wenn ihr da jetzt ein bisschen verstärkt vermitteln oder, eben weil ihr aus dem Ausland kommt und keine Agenda habt... 00:21:33-6

H: Also, es ist also nicht nur die Herkunft, sondern das ist ja immer auch ein Personenfaktor. Da fand aber gerade die Kopplung etwa von (name of evaluator), den fand ich wegen seiner integrativen Art, seiner sehr verbindlichen und trotzdem glasklaren, in der Sache glasklaren Rückmeldung, aber gleichzeitig sehr verbindlich und in keiner Weise konfliktverschärfend. Also, diese Kopplung, fand ich da sehr, sehr hilfreich und da haben aber die Kollegen auch zuweilen sehr bewusst noch einmal gesagt, also wir stehen jetzt ja nicht unbedingt in irgendeinem Interessenverdacht und vielleicht sollten wir unsere, aus unserer Perspektive noch mal eine Rückmeldung geben. Das fand ich also ausgesprochen hilfreich. Und an einigen Stellen haben wir sie auch speziell gebeten. Und auch eben noch mal internationale Erfahrungen einzubringen, wie werden bestimmte Probleme woanders gelöst. 00:22:31-3

I: Ja. Was eben kann man da jetzt vielleicht anders machen. Und eben der Sprecher war ja... 00:22:36-7

H: Was sind Standards auch woanders? 00:22:39-9

I: Als Vergleich. Und was, oder wie empfanden Sie das, dass Professor Wijnen eben als Sprecher tätig war? 00:22:49-0

H: Ausgesprochen hilfreich auch. Also auch in seiner sehr freundlichen Art. Also, das war auch die Bitte damals. Also, Baumert und ich waren ja beide Stellvertreter und also, wir fanden, dass das eine sehr glückliche Wahl war in dieser Mischung von Verbindlichkeit, aber eben auch sehr, sehr erfahren... in den Niederlanden sind solche evaluativen Bestrebungen ja einfach schon sehr viel besser verankert. Das fand ich also, eine sehr... Also diese Person war da einfach sehr gut gewählt. 00:23:28-4

I: Und er hat auch ja vermittelnd eingegriffen, oder? 00:23:34-4

H: Nein, er hat Positionen sehr klar benannt und auch, für meine Begriffe, sehr deutlich noch mal gemacht, dass an manchen Stellen ein Konsens nicht herzustellen ist, sondern dass da, wo das der Fall ist, man dann eben die entsprechenden Positionen verdeutlichen muss und wir hatten ja unsererseits auch nicht zu entscheiden, sondern wir haben dann die Aufgabe als Kommission, die Unterschiedlichkeit der Positionen und die daran gebundenen verschiedenen Alternativen zu verdeutlichen. Gefordert ist dann ja in der Tat, die politische Entscheidung. 00:24:14-2

I: Die danach kommt. 00:24:14-2

H: Die danach kommen soll, ja. Da ist das, was wir dann nicht mehr verantworten. Aber erst einmal die verschiedenen Positionen zu verdeutlichen, und dann auch irgendwann zu sagen, das muss man jetzt als verschiedene Möglichkeiten stehen lassen. Nur das auch nicht immer nur etwas mit, also, das hat auch noch mal innerhalb unserer Disziplin häufig auch etwas damit zu tun, dass beispielsweise die sonderpädagogischen Perspektiven im Augenblick ja sehr im Fluss sind, und dass es das sehr unterschiedliche Ideen dazu gibt, wie denn etwa das Studium der Sonderpädagogik, in welche Richtung sich das entwickeln soll. Ich mache es mal an dem Bereich deutlich. Die Berufspädagogik kann man da genauso als einen Pfeil nehmen und da gibt es dann auf der einen Seite Positionen, die das eher klassisch sehen, so als Zweitfach Ausbildung mit zwei verschiedenen Schwerpunkten in der Ausrichtung eines sonderpädagogischen Lehramtstudiums. Es gibt aber auch die Position zu sagen, wir müssen uns hier in die Richtung von Rehabilitationswissenschaften entwickeln und wir müssen stärker so etwas, wie Querschnittkompetenzen vermitteln und es ist gar nicht einzusehen, dass also, jede einzelne Verhaltensstörung oder so, dann nun auch noch einen eigenen pädagogischen Psychologen braucht,

sondern dann muss man vielleicht die Organisationsstrukturen ganz anders definieren. Also, da kann man nicht immer auf einen Konsens kommen, sondern da muss man dann die verschiedenen Positionen, die in der Kommission auch vertreten sind ja, möglichst deutlich ausführen und dann aufzeigen. Was hat die Kommission entweder... wir haben es, glaube ich, nie, irgendwann auch mal, es ist nie irgendwo ein Minderheitenvotum formuliert worden. Aber wir haben das durchaus zwischenzeitlich ins Auge gefasst. Das muss dann möglich sein, auch ohne dass darüber also die Arbeitsfähigkeit einer Kommission leidet. Das ist guter demokratischer Brauch. Wünschenswert ist sicherlich, dass man über intensive Diskussion sich eher auf gemeinsame Empfehlungen einigen kann. Ohne dass es vermischt wird zu einem Kompromiss, mit dem eigentlich keiner richtig zufrieden ist. Und das fand ich in unserer Kommission recht geglückt. Was ich wenig geglückt fand, ist dann die Umsetzung durch die politische Seite, das fand ich eher enttäuschend. Da habe ich schon den Eindruck, dass viel Arbeit in diese Kommission gegangen ist. Ohne, dass das später deutlich geworden wäre, wo das denn eigentlich auch umgesetzt worden wäre von den Entscheidungsträgern. 00:27:01-0

I: Haben Sie das speziell verfolgen können, also, im Anschluss an die Arbeit? 00:27:05-4

H: Wir haben ja noch mal... Herr Wijnen und ich im Ministerium in Stuttgart die Empfehlung überreicht, und wir haben die Presse verfolgt. Ich bin später auch noch mal eingeladen worden, mehrfach von Verbandsvertretern oder von verschiedenen Gremien, die das intern sehr intensiv diskutiert haben. Also, die Aufnahme vor Ort, an den einzelnen Standorten, da habe ich den Eindruck gehabt, da ist vieles in Bewegung gekommen. Da ist später auch noch mal, Erfahrung noch mal abgefragt worden bei der Gestaltung von Studiengängen und insofern kann man nicht sagen, dass diese Arbeit umsonst gewesen wäre. Aber so von den politischen Gremien könnte ich jetzt nicht sagen, wo ist da wirklich von dem Empfehlungsteil, den wir formuliert haben, auch wirklich etwas umgesetzt worden. Etwa, was Besserstellung der, hinsichtlich der Ressourcen von Pädagogischen Hochschulen angeht, die Stärkung der Forschungskapazität und ähnliche Dinge. So, das fand ich ein Stück weit enttäuschend. 00:28:19-0

I: Da werde ich bestimmt noch nachhaken beim Ministerium. Gut, zu denen komme ich ja noch. Wie, würden Sie sagen, dass die Zusammenarbeit in der Kommission gelaufen ist, insgesamt gesehen? 00:28:34-8

H: Also, dafür, dass eine sehr große Kommission war, ist es ausgezeichnet gelaufen. Ich fand, dass die *evalag* das hervorragend organisatorisch betreut hat, und dass wir also da wirklich eine ganz ausgezeichnete Zuarbeit bekommen haben und auch unsere Zuarbeiten, die wir geliefert haben, sind von denen ganz wunderbar integriert worden. Das war wirklich ganz, habe ich noch nie so top und kompetent wirklich aufgenommen erfahren, wie dort. 00:29:15-6

I: Lag das an der Person von Frau Dr. Röbbecke speziell? 00:29:19-2

H: Ja, also da würde ich sagen, da liegt der Löwenanteil. Ich meine, auch eine noch so kompetente Person kann, wenn die Institution, für die sie handelt, das nicht mitträgt und das nicht mitgestaltet, man kann das nicht alleine hinkriegen, aber Frau Röbbecke war da wirklich ganz, ganz ausgezeichnet. Auch in ihrer Kombination von Sachkompetenz und von angenehm verbindlicher, kommunikativer Kompetenz. Also, wie sie da doch diese schwerbeschäftigten und also solche Gremien, in denen sich so viele Menschen aus allen Ecken dann auch der Republik jeweils treffen, wie es ihr gelungen ist, die an einen Tisch zu bringen, beziehungsweise, da wo sich unüberwindliche Terminschwierigkeiten auftaten, dass dann über schriftliche Kommunikation einzubinden. Das war schon toll. Muss ich wirklich sagen, und auch den Arbeitsdruck, der gegen Ende ja immer unvermeidlich entsteht, Fertigstellung des Berichts, das war, das habe ich sehr, sehr geschätzt. 00:30:30-7

I: Ich meine, ich stelle mir das als ziemlich kompliziert vor, wenn eben, weil in, bei der *evalag* im Gegensatz zu anderen Evaluationsagenturen, die Berichte ja von Verschiedenen geschrieben werden und dann zusammengebracht wurden. Anderswo eben wird das nur von einer Person geschrieben und hier hat man ja verschiedene Stilrichtungen und das muss da alles irgendwie vereinigt werden. 00:30:56-2

H: Also, die Endredaktion hat uns dann ja auch mehrere Klausurtage gekostet. 00:31:01-0

I: Ich habe die Entwürfe gesehen. (lacht) Das muss schon ganz viel Arbeit gewesen sein. 00:31:08-5

H: Ja, das war wirklich, in der Tat... Es waren eben auch alles Menschen, Herr Prenzel oder Herr Baumert, und ich denke, ich selber habe auch genügend Arbeit an anderen Stellen, und dafür dann immer noch so die Power zu finden, dass dann hinreichend soweit, bis hin in die Formulierungen zu überarbeiten. Erstens so, dass dann auch keine Verletzungen entstehen. Ersten, dass es sachlich angemessen ist, und zweitens, dass es dann eben auch noch kommunikativ verträglich abgefasst ist. Das ist schon wirklich sehr aufwendig, und ich finde, dass das dann geklappt hat, nach meinem Eindruck. 00:31:49-5

SH Würden Sie sagen, dass das in irgendeiner Form, vielleicht ein bisschen abgeschwächt wurde? Also, dass es vielleicht manchmal einen gegeben hat, der gesagt hat, also, das müssen wir stärker ausdrücken, weil sonst werden sie es nicht, oder aber dass da... 00:32:04-5

H: Also, an einigen Stellen polemische Spitzen, die, von denen ich denke, dass die in einem solchen Bericht aber auch nicht vertretbar sind. Also, man kann in einem Kommentar, in einer Zeitschrift vielleicht, sich so verhalten und man kann sich als Einzelperson so verhalten, aber als Mitglied, als Kommissionsbericht, denke ich mir, ist man verpflichtet, sich jeglicher Polemik zu enthalten, und diese Textsorte ist eine schwierige, eines solchen Berichtes und da manchmal Kollegen davon zu überzeugen, dass jetzt die Grenze zur Polemik, die sich zwar vergnüglich liest und wo man inhaltlich eigentlich gute Lust hätte, sich dem anzuschließen, dass das aber in dieser Form jetzt so nicht geht. Das ist, glaube ich, überall gelungen, und das war dann manch-mal nicht ganz ohne Aufwand. 00:33:00-5

I: Das glaube ich. 00:33:02-1

H: Aber das finde ich nötig. Also, das finde ich wirklich nötig, da muss man auch von der Zunft erwarten, dass sie sich dann konfliktfreudig genug an solchen Produkten abarbeitet, aber das ist schon keine alltägliche Aufgabe. Wir sind ja eigentlich alle sehr einsame Formulierer und hängen auch alle an unseren Formulierungen und da denke ich, dass Frau Röbbecke sich dann auch noch mal wirklich ungeheure Verdienste erworben hat, durch ihre verbindliche Art und durch Vorschläge. Also wir haben in der Schlussphase noch mal auch bilateral zusammen-gearbeitet. Das fand ich eine ganz ausgesprochen gute Erfahrung. 00:33:46-9

I: War das wichtig, dass sie… sie ist ja Politikwissenschaftlerin. Wäre es vielleicht noch einen Tick besser gewesen, wenn sie selber Erziehungswissenschaftlerin gewesen wäre, oder war das nicht, in dem Fall, von Belang? 00:33:55-7

H: Also, ihre Kompetenzen, die sie in diesem Bereich hatte waren enorm. Also, da glaube ich nicht, dass es viele Erziehungswissenschaftler gibt, die über ein ähnliches Detailwissen inzwischen verfügen. Weil wir ja sehr häufig auch nur unseren unmittelbaren Bereich intensiv kennen. Also, alleine die Zugehörigkeit zur erziehungswissenschaftlichen Zunft ist in dem spezi-fischen Bereich noch nicht, oder... Also da war, glaube ich, Frau Röbbecke sehr satisfaktions-fähig. (Gelächter) Das ist ein Witz. Und die Tatsache, dass sie eben aber auch noch mal diesen politikwissenschaftlichen Blick jeweils hatte, und sich da auch noch mal so ein Stück ja, reflektorische Distanz dann da auch dazu reingebracht hat, das war noch mal sehr hilfreich. Und das war bei denjenigen, die etwa wie Herr Tenorth dann mit einer historischen Perspektive noch mal drauf gucken. Also, die Vielfalt an Kompetenzen, die da zusammenkam war schon wirklich hilfreich. 00:35:06-9

I: Vor allen Dingen, soviel ich sehen kann, in der Kommission, hat man wirklich auch versucht bei den achtzehn Mitgliedern, alle möglichen Bereiche abzudecken. 00:35:12-8

H: Ja. 00:35:13-3

I: Trotz allem, also, jeder hat zwar seinen Schwerpunkt gehabt, konnte aber noch ein bisschen weiterschauen, als nur über sein Feld hin. Okay. 00:35:21-7

H: Also, das finde ich auch, also bei Ihrer Eingangsfrage eben, was ist wichtig bei der Zusammensetzung solcher Kommissionen, also, ich glaube, es wäre nicht gut, wenn man die sehr stark, ich sag mal so, auf technokratisch oder, das ist jetzt schon zu scharf, so auf instrumentelle Fertigkeiten, was Evaluationstechniken und Erfahrungen angeht, wenn dass das Haupt- oder einzige Kriterium wäre, sondern gerade die, der etwas breitere Blick, natürlich gehören dazu methodische Kompetenzen und, wir haben uns ja eigene Instrumente gestrickt, wo noch keine Vorlagen, um ganz bestimmte Urteile ein Stück objektivierbar zu machen. Und dazu sind solche methodischen Kompetenzen wichtig, aber gerade die Breite der Qualifikationen und den Überblick etwa über berufspädagogische, über sonderpädagogische, da haben wir sehr drauf geachtet, dass da ein unterschiedliches Bündel an Fähigkeiten, da muss auch nicht jeder über alle verfügen, sondern, dass ist ja der Vorteil von Arbeitsteilung. 00:36:30-3

I: Würden Sie sagen, dass alle gleichgewichtig dann ihre Stimme zum Ausdruck bringen konnten, oder waren, gab es da spezielle Mitglieder, sie haben stark für ihr Feld gekämpft und über die anderen war das dann weniger aussagekräftig? 00:36:51-4

H: Doch, das gibt es natürlich, also, ich würde jetzt nicht sagen, dass Menschen da Lobbyarbeit betrieben haben. Das wäre ein Missverständnis, dem möchte ich gerne vorbeugen. Aber es gibt ganz klar, natürlich Bereiche, in denen man selber sehr engagiert ist, in denen der eigene Informationsstand höher ist und wo auch das eigene, wo man stärker für brennt, und das war schon zu spüren. Das finde ich auch in Ordnung, dass man zu bestimmten Punkten, sehr gezielt auch noch mal die Menschen angesprochen hat, und aber gleichzeitig auch da wieder gesagt hat, bei allem Engagement für diesen Bereich, bei den Empfehlungen muss es jetzt aber schon auch Abgewogenheiten geben. Das hat eben auch Auseinandersetzungen zur Folge gehabt. 00:37:47-5

I: Ja. Okay. 00:37:47-9

H: Aber es gab wenig so Fachegoisten. Also schon eine große Aufgeschlossenheit, sich da mit der Gesamtlage auseinander zu setzen und etwa solche Querschnittfragen, wie Förderung des wissenschaftlichen Nachwuchses, Qualität von Lehre, das sind ja wirklich auch Fragen, die quer zu solchen... 00:38:07-4

I: Gegenüber allen. 00:38:11-0

H: .... solchen Überlegungen stehen. 00:38:11-4

I: Sie haben jetzt auch sehr, sehr viele Berichte, haben die Self Reports von den Universitäten und PHs bekommen. Wie war das als Sie die bekamen? Ich denke, jeder hat sie alle bekommen. Und wie reagierten Sie darauf? 00:38:26-5

H: Also, ich fand erst einmal, einen für mich unerwartet schwierigen Punkt, wirklich Klarheit über die Verlässlichkeit der eingeholten Auskünfte zu bekommen. Also, wenn man Abgleich vorgenommen hat zwischen statistischen Auswertungen, die wir aufgrund von anderen Aggregationen, die bereits vorlagen im Land und den Auskünften der einzelnen Institutionen, dann ergaben sich ganz unvorhergesehene Widersprüche. Also, auch das kennt man.

Ich bin auch als Dekanin eines Fachbereichs, der evaluiert wurde, war mir das nicht ganz unvertraut so etwas, aber das potenziert sich natürlich, wenn man so viele Standorte hat und wenn es so differenziert ist, wie in einem so großen Land wie Baden Württemberg . Da fand ich es schon sehr, sehr schwierig und sehr verdienstvoll eben, da auch wieder die Unterstützung durch *evalag*, wirklich noch mal Daten kritisch zu hinterfragen und Widersprüche aufzudecken und dann zu schauen, weil man verliert gegenüber den Rückmeldungen bei den jeweiligen Institutionen sofort an Glaubwürdigkeit als Kommission, wenn die einem nachweisen, die wissen gar nicht, über welche Zahlen sie reden, was die Quellen sind und da sind Unsicherheiten über die Güte der Daten. Das ist also tödlich. Insofern war das ein Punkt, mit dem ich nicht allzu arbeitsaufwendig eingeschätzt hätte, der aber wichtig, glaube ich, war und da gab es an einigen Stellen halt stärkeren Widerstand, eine stärkere Abwehr. Und an einigen Stellen waren das wirklich ganz klare Interessengegensätze.
Also da , wo wir etwa sehr klar die Situation der Fachdidaktiken angesprochen haben, wie, das ist gar kein besonderer Fall unserer Baden Württembergischen Kommission, sondern das spiegelt sich in der Entwicklung der letzten Jahre an den meisten Universitäten, dass die Fachdidaktiken häufig relativ stiefmütterlich behandelt worden sind, dass da Stellen gestrichen worden sind. Dass es so das Bestreben gab, dass nicht als Professuren auszuschreiben, sondern eher erfahrene Praktiker in die Schulen zu holen und Expertise weiterzugeben, aber nicht eine forschungsorientierte Fachdidaktik wirklich aufzubauen. Wenn wir also solche Rückmeldungen gegeben haben, dann hat das weder die Universitätsleitung erfreut, noch häufig auch die Fachbereiche selber, für die das ja immer auch ein, in der Außendarstellung ein unerwünschtes Ergebnis ist, wenn so etwas publik wird. Obwohl die in der Sache häufig die Einschätzung geteilt wird, aber ob man das freudig begrüßt, als endlich mal klare Unterstützung bei der Benennung eines Missstandes, oder ob man sich sagt, dass muss jetzt aber ausbalanciert werden, in dem Interesse hinreichend attraktiv zu sein für Studierende und zwar auch für Studierende mit Anspruch, die wir ja hier an unserer Hochschule

gerne sehen möchten. Dass hat nicht immer Freude ausgelöst, also, insofern ist das genau dieses Balanceverhältnis. Das ist unterschiedlich. Also, weil es überwiegend, fand ich, dass wir mit großer Offenheit behandelt worden sind, also wir sind wenig auf so massive Abwehr gestoßen. Bei einigen auch ja, renommierten Universitäten, hat man dann schon gemerkt, dass, als, dass relativ enttäuscht festgestellt wurde, dass einige verdienstvolle Arbeiten, ihrer Meinung nach, nicht hinreichend gewürdigt worden sind, oder dass es eine stärkere Verlagerung gibt. Ja, die so hinsichtlich der Hochbewertung eher sozialwissenschaftlich orientierter Bildungsforschung, die da ein Stück durchschlägt. Also, das muss man sicherlich deutlich machen. In der Erziehungswissenschaft hat ja doch ein Paradigmenwechsel stattgefunden und die Universitäten, an denen genau solche Auseinandersetzungen noch laufen, und wo stärker darauf verwiesen wurde, dass aber eben Erziehungswissenschaft anders als etwa solch technisch- naturwissenschaftliche Fakultäten, eben immer auch eine Wissenschaft ist, die auf die Gestaltung einer pädagogischen Praxis gerichtet ist, und dass insofern die Forschungsorientierung dort auch unterschiedliche Ausrichtungen noch mal hat. Also, um es ganz schlicht zu sagen, nicht die Frage der eingeworbenen Mittel bei der deutschen Forschungsgemeinschaft als das Prestigeträchtigste Forschungspotential, nicht nur die Ausschöpfung dieses Bereichs ist wichtig, sondern wenn da regional mit Bildungseinrichtungen vor Ort zusammengearbeitet wird und dort Problemlösungen entwickelt werden, gemeinsam mit Praktikern und die Umsetzung dieser Problemlösung wissenschaftlich evaluiert wird, dass diese auch einen wichtigen Stellenwert hat und haben muss. Darüber haben wir durchaus kontrovers diskutiert, welchen Stellenwert das hat und in welcher Weise man aber auch sagen muss, es gibt auch Einrichtungen, an denen eben einfach zu wenig Forschung stattfindet. Weil die im Tagesgeschäft untergehen und in der Überlast der Studierenden, und dass, wenn man Forschung haben will, man auch Forschung ausbauen muss. Und nicht denken muss, das geht alles mit Bordmitteln. Das geht eben dann nicht, wenn man mit solchen Überlastzahlen zu kämpfen hat und wenn man auch, etwa wie in Baden Württemberg alle Pädagogischen Hochschulen am Mittwoch sind die immer alle im Praxisfeld. Also, wenn sie an einer Pädagogischen Hochschule am Mittwoch irgendeinen Kollegen finden wollen, ist es aussichtslos. Dann ist der mit seinen Studierenden, sitzt der in Klassen und hospitiert. Und ob das so bleiben muss, oder ob man nicht auch andere Formen von forschendem Zugang zu Praxis daneben stellen muss, das sind kontroverse Fragen und wenn wir die dann diskutiert haben mit den Praktikern und da die Gewichtung anders gesetzt wurde, dann waren die schon zum Teil auch verschnupft. Und haben ihre Arbeit verteidigt und die Anregung, mal darüber nachzudenken , ob man das auch anreichern kann, durch andere Formen...etwa durch forschenden- Lernens, Projektorientiertes -Lernen, da gab es bei einigen in Situationen eine ausgesprochene Aufgeschlossenheit und auch gute Ansätze, die da jetzt verstärkt werden. Und bei anderen eher eine hohe Distanz und die kann man auch durch eine solche Evaluation nicht von heute auf morgen umdrehen. Das wäre eine Überschätzung dessen, was man durch Evaluation erreichen kann. Wer sich die Messlatte so hoch setzt, muss daran scheitern als Kommission. 00:45:59-5

I: Und ich denke, das bedarf einfach einer gewissen Zeit, bis manche Sachen sich umändern. Das geht nicht von heute auf morgen. 00:46:04-5

H: So ist es. 00:46:04-3

I: Und würden Sie sagen, dass die Self Reports eigentlich das wiederspiegelten, was Sie dann vor Ort gesehen haben, oder gab es da große Differenzen zum Teil? Wo Sie sagen, aber das stimmt doch gar nicht, was auf dem Papier steht. Schauen Sie sich doch das an, oder war es doch ähnlich? Also, die Self Reports waren, wie die Institutionen dann auch wirklich waren? 00:46:26-7

H: Also, die Begehungen und die Gespräche mit den verschiedenen Gruppen waren schon sehr wichtig, um diesen Selbstreport zu arrondieren. Und da fand ich es auch sehr wichtig, getrennt mit den Gruppen zu reden. Also, insbesondere die Gespräche mit Studierenden waren ausgesprochen wichtig. Da fand ich es auch gut, dass das nicht handverlesen ausgesuchte Studierende lediglich waren, sondern , wir haben ja zum Teil ja mit großen Gremien- also mit zweihundert Leuten in einer Pädagogischen Hochschule ... 00:46:58-4

I: Das habe ich gehört. 00:46:58-4

H: Haben wir Gespräche geführt und das ging. Das fand ich also eine ganz ausgesprochen interessante Erfahrung. Und Selbstreports tendieren dazu, Probleme als, Probleme zwar zu benennen, aber vor allen Dingen den Aspekt der auf dem Weg befindlichen Lösung zu betonen. Wohingegen die Gespräche mit den Gruppen zum Teil doch die Probleme sich noch mal geschärft haben, und noch mal sehr deutlich auch, ja, Dinge zu Tage getreten haben- ich mache es an einem Beispiel mal deutlich- es wurde in dem Selbstreport, wird immer betont, die Bereitschaft zur Kooperation mit den verschiedensten Institutionen. Beispielsweise auch zwischen Pädagogischer Hochschule und Universität, wenn die denn an einem Ort sind ist das ja ein naheliegender Gedanke. Und da wo so etwas im Selbstreport immer zunächst mal als selbstverständliche Kooperationsbereitschaft geäußert wird, und zuweilen dann auch auf Austausch hinsichtlich Bibliotheksnutzung oder ähnlichen institutionellen Verkopplungen verwiesen wird, da haben wir schon in den Gesprächen vor Ort doch sehr deutlich die Vorbehalte entweder direkt mitgeteilt bekommen, oder aber auch von den verschiedenen Beteiligten sehr unterschiedliche Perspektiven dann letztlich mitgeteilt bekommen. Also da, wo die Pädagogischen Hochschulen etwa sehr stark ihr Interesse an einer Kooperation mit der Universität stark gemacht haben, und gesagt haben, wir würden gerne auch Austausch auf Dozentenebene, gemeinsame Forschungsansätze haben, da wurde zum Teil bei den, wo dieser hierarchische Aspekt schon deutlich, dass die Universitäten sehr gerne mit ihrer Partnereinrichtungen in Michigan oder mit der internationalen Szene in schieß-

mich- tot kooperieren, aber die Kooperation vor Ort ihnen doch weitaus weniger naheliegend erscheint. Und das ist unter der Frage, des harten Konkurrenzdrucks im Wissenschaftsbetrieb auch alles verständlich, trotzdem ist es im Sinne einer gedeihlichen Zusammenarbeit zwischen, also, wenn man schon so eine Doppelstruktur hat, aus meiner Sicht, schon ein Problem und das taucht in den Selbstreports bei weitem nicht so scharf auf, wie in den Gesprächen, die wir vor Ort geführt haben. Wo man dann eben auch die Realitäten besser einholen kann. Wer macht denn jetzt tatsächlich die Praxis. Betreut da die Praxisübungen, oder so. Sind es da nicht doch die Studienräte, die erfahrenen, die da jetzt reingeholt werden. Auch wenn da eigentlich eine Stelle ausgewiesen ist, aber auf dieser einzigen Stelle, die es überhaupt gibt, geht da doch wieder ganz viel Kapazität möglicherweise in ehrgeizige Forschungsprojekte und solche Sachen deckt man nicht über Self Reports auf, sondern da kriegt man schon noch mal deutliche Hinweise aus diesen Begehungen und aus den Gesprächen mit verschiedenen Gruppen, wenn man hinreichend nachbohrt, das ist dann ja aber auch die Aufgabe einer solchen Kommission. 00:50:19-8

I: Ja, sicher. Wenn Sie mit, oder als Sie mit den Studenten gesprochen haben, Sie sagten schon, einmal waren das an die zweihundert. Jetzt, von ihren Erfahrungswerten, würden Sie sagen , es ist besser wenn man eine Gruppe von, sagen wir mal, zwanzig bis fünfzig hat, oder fünf bis zehn, oder haben Sie da irgendwelche Erfahrungen aus, da rausholen können, wo Sie sagen, ja das war eine ideale Gruppengröße für diese Gespräche. 00:50:50-1

H: Ich glaube, es ist weniger die Gruppengröße. Sondern es wäre wünschenswert, dass die Gespräche mit Studierenden ein Stück weit stärker vorbereitet werden können. Also, als wir... wir waren nicht vorbereitet auf ein Gespräch in der Aula, haben aber akzeptiert, dass engagierte Studierende das in ihrer Vorbereitung gesagt haben, wir möchten das aber auf demokratische Füße stellen und insofern möchten wir das eben als eine Art Podiumsdiskussion, wo dann die, möglichst Viele auch zu Wort, mit einbezogen werden können. Das wäre natürlich günstig gewesen, wenn man das vorbereitet hätte. Dass da eine etwas verteiltere Moderation möglich gewesen wäre. Vom Ertrag fand ich das trotzdem sehr ergiebig. Wenn man stattdessen fünf studentische Vertreter, von denen man nicht ganz sicher sein kann, sind das jetzt nicht doch die studentischen Hilfskräfte in den Prestigeprojekten, deren Perspektive man angeliefert kriegt. Da dominiert dann also die Frage der Authentizität, die der Gruppengröße (long pause). Aber ein vertieftes Gespräch mit einigen wenigen und auch mit demokratisch legitimierten Vertretern dann, darum sollte man sich in jedem Fall bemühen, das würd ich vorschlagen. Nicht nur so auf Zuruf, sondern da auch wirklich über die gewählten Vertreter der Studentenschaft zu gehen und von denen dafür zu sorgen, dass die mit einer Nominierung beteiligt sind. Zu kleine Gruppengrößen können manchmal, wenn dann solche erschlagenden Kommissionen von außen kommen auch wieder erschlagend wirken. Also, wenn da lediglich vier Studenten sitzen und, auch wenn die selbstbewusst sind... 00:52:53-8

I: Da sitzen drei Professoren gegenüber, vielleicht ist das auch ein Problem. 00:52:57-0

H: Genau. 00:52:58-0

I:...ganz offen... 00:52:59-4

H: Und alles wird gerade aufgeschrieben und so, da kann auch die Unterstützung einer größeren Gruppe durchaus ganz sinnvoll sein. Würde ich jetzt keine technischen Empfehlungen draus ablesen, sondern es kommt auch da wieder noch mal auf die Aspekte an, auf die, wir haben in, ausgehend von den Self Reports, immer auch ganz bestimmte Fragerichtungen schon gehabt, wo wir mit den verschiedenen Gruppen, also, wenn sich irgendwo andeutet, hier gibt es offensichtlich auch ein Problem von Nachwuchsrekrutierung, haben wir das stärker zum Thema gemacht. Etwa bei der, bei den Gesprächen mit wissenschaftlichen Mitarbeitern, oder wo es ganz deutliche Hinweise von Studierenden gab, dass insbesondere in manchen Studiengängen Engpässe sind, und die Erreichbarkeit von Lehrenden sich etwa schon als Problem hat zeigen lassen, oder bestimmte Ausstellungsmerkmale. Dann waren das ja schon mal Hinweise für die Vorbereitung auf diese Gespräche. Also die Kommission hat sich arbeitsteilig immer auch auf solche Gespräche vorbereitet. Und ich glaube, dass die Qualität der Vorbereitung auch noch mal mindestens so wichtig ist, wie solche Faktoren, wie Gruppengrößen. 00:54:17-6

I: Ja. Die Zusammensetzung von den Kommissionsmitgliedern, die dann die Vor-Ort- Begehungen gemacht haben, war das von Evalag organisiert, oder zum Teil von der Kommission selber vorgeschlagen? 00:54:30-6

H: Das ist immer abgestimmt worden. Und so etwas ist ja immer eine Mischung von Interessen und pragmatischen Überlegungen. Es muss sich dann auch einigermaßen terminlich vereinbaren lassen und die Kompetenzen müssen vorhanden sein. Also, dass... das war wieder eine dieser wunderbaren Zuarbeiten von evalag, denn auch noch... muss man ja auch die Grippeausfälle noch kompensieren und kurzfristig Dinge reparieren. Was weiß ich was, bis hin zu den Fehlleistungen der Deutschen Bundesbahn. Weil man irgendwo auf einem Abstellgleis gelandet ist, statt an dem Zielort. 00:55:12-0

I: Ja. Das hatte ich ja letzte Woche. 00:55:18-2

H: Also, auch dabei war dann, waren die Kollegen sehr findig und sehr hilfreich. Und das erfordert dann auch flexibles, sich dann noch mal schnell darauf einlassen, auf eine Sache, auf die man eigentlich nicht so vorbereitet war, aber dadurch, dass es eben immer Leitfäden vorher gab, Zusammenfassungen noch mal, Waschzettel, wenn man so will, auf denen die wichtigsten Sachen

standen, konnte das gut aufgegriffen werden. 00:55:45-5

I: Ich glaube, es waren an die drei Mitglieder jedes Mal bei einer Vor-Ort- Begehung...Drei Mitglieder? 00:55:52-3

H: Mindestens. Drei bis vier. Es kam auf die Größe des Standortes an. Also, wenn man so eine große Einrichtung wie Pädagogische Hochschule Freiburg hat, waren es mehr, als wenn wir bei diesen ganz kleinen Einrichtungen, die ja fast schon wie größere Gymnasien sich ausnehmen. Also, Schwäbisch Gmünd beispielsweise. 00:56:16-1

I: Also, das wurde da auch ein bisschen angepasst. 00:56:18-2

H: Ja, das wurde da schon angepasst. 00:56:22-0

I: Okay. Jetzt zu der Kommission an sich. Wie hat sich da eine Art Entscheidungskultur innerhalb der Kommission, gut, einige Mitglieder kannten sich schon vorher, sie verstanden sich mindestens von der Arbeitsweise, vielleicht von früher her. Wie war das dann, das zusammen zu bringen? Hat es da mehr, hat es da schon Personen gegeben, die mehr das Sagen hatten, immerzu, oder wie war das denn? Und wie war das für die einzelnen Mitglieder? Vielleicht, vom Gefühl? 00:56:58-8

H: Also, wir haben keine gruppendynamischen Sitzungen eingeschoben. Wo wir das noch mal zum Thema gemacht hätten. Es hat aber auch nie, und ich glaube, weil wir uns auch darüber ausgetauscht haben, dass schon relativ sensibel wahrgenommen haben. Es hat an einigen Stellen diplomatische Vermittlungen gegeben und natürlich ist in einem solch großen Gremium die Beteiligung nie völlig ausgeglichen. Das, was uns geholfen hat, war die ungeheure sprachliche Kompetenz unserer auswärtigen Mitglieder. Wenn wir viele Dinge da auch noch in Englisch hätten führen müssen, viele Diskussionen, wäre das sicher schwieriger geworden. Das war nun bei allen so, dass die ein ganz wunderbares Deutsch gesprochen haben und das hat viel Arbeit abgenommen haben. Die schriftliche Vorbereitung war wichtig und die Rückmeldung, die dann in großer Offenheit gegeben wurde. Einschließlich auch der Überarbeitung von Texten, da wo wir dann gesagt haben, offensichtlich kann jemand sich jetzt von seinem Ursprungstext nicht lösen, das müssen wir jetzt mal über Weitergabe regeln, und dann muss man das auch aushalten. Dass eben da eigene Ursprungspapier sich jetzt sehr entfernt von der Ursprungsfassung. 00:58:24-2

I: Wie kamen die damit klar? Also, diejenigen, die das aus der Hand dann geben mussten? 00:58:29-7

H: Also, ich habe den Eindruck gehabt, dass da keine Feindschaften für das Leben bei entstanden

sind. Dass einige...also, es war an einigen Stellen vorrausagbar, wenn X etwas gesagt hat, wird jetzt als nächstes die Wortmeldung von Y kommen. Das ist auch so, wie man das immer kennt. Insofern war da nicht alles die reine Harmonie. Es gab aber keine unüberbrückbaren Gräben, sondern es gab eine sehr sachorientierte Diskussion. In einem verhältnismäßig großem Gremium, ich denke, dass an sehr vielen Stellen, da der Kollege Baumert aus meiner Perspektive noch mal sehr hilfreich war, in seiner wirklich ganz glänzenden Art, Dinge auf den Punkt zu bringen, zu strukturieren von Vorneherein, noch mal auch die Aufmerksamkeit sehr zu fokussieren, und zu sagen, also, bei allen Uneinigkeiten im Detail, zeichnet sich jetzt aber in den wesentlichen Punkten für meine Begriffe XYZ ab, und dann hat man sich ein Stück weit daran abgearbeitet. Es hat eigentlich nie einen solchen Konsens gegeben, einen solchen Dissens gegeben, dass man mal gesagt hat, also, nein, an der Stelle fühle ich mich nun völlig falsch dargestellt, oder hier fühle ich mich jetzt aber völlig untergebügelt. Oder so. Sondern es hat an einigen Stellen Mäkeleien gegeben, aber sie sind zu einem guten... sie sind integriert worden. Es hat an einigen Stellen Fußnoten gegeben. So, es gibt dann immer solche Karrieren, denn wird etwas in einer Fußnote zunächst thematisiert und beim letzten Überarbeitungsversuch erweisen sich einige Fußnoten dann doch noch als wieder auflösbar. Wir haben sehr konsensorientiert diskutiert und das hat gut geklappt. Und ich glaube, dass wir auch eine ganz gute Arbeitsteilung in unserer Kommission hatten, wo so eine Rollenteilung von Tieren auf der einen Seite und verbindlich wieder einholen auf der anderen Seite, sich dann so ein Stück...Also, das war eine konzertierte Aktionen, die gut funktioniert haben. Das hilft manchmal einfach auch. 01:01:03-4

I: Die Tatsache, dass Herr Professor Baumert speziell auch Erfahrungen mit dem niederländischen Evaluationssystem gemacht hat, war das sehr von Einfluss auf die Arbeit insgesamt? Hat man das gespürt? 01:01:17-7

H: Also, es war nicht nur Herr Baumert. 01:01:19-1

I: Ich weiß, es waren noch andere dabei, aber er... 01:01:20-5

H: Herrn Terhart, fand ich, zum Beispiel auch ausgesprochen...also, immer wenn es um solche verbindliche, um Brückenbau ging, fand ich den Kollegen Terhart auch unglaublich hilfreich. Bei aller Deutlichkeit seiner Positionen und die Bereitschaft, auch mal Dinge nebeneinander stehen zu lassen, und zu sagen, da gibt es eben hier unterschiedliche Spielarten. Diese Toleranz war hoch ausgeprägt. Und das hat die Sache erleichtert. 01:01:49-8

I: Aber es war, oder, es war auch für die ursprünglichen Kriterien, also für diesen Katalog, welcher am Anfang geschrieben wurde. War das da schon sehr nützlich, das man diese Perspektive schon hatte? 01:02:07-4

H: Das war nützlich, es war gleichzeitig aber auch noch mal hilfreich, dass Herr Baumert da sehr offen war, also die ersten, von der Formulierung der ersten Kriterien hat es massive Verschiebungen noch mal gegeben, die eben nicht nur einseitig die Forschungsorientierung so sehr hoch gehängt haben, sondern die Qualität von Lehre, die eben dafür ein ausgewogeneres Verhältnis gesorgt haben. Und die Tatsache, dass das immer ganz besonders schwer methodisch durch solche Kommissionen aber zu erfassen ist. Da haben wir aber ganz gut, glaube ich, für Indikatoren, Systeme und so gesorgt. Also, das war hilfreich. Und das da auf der einen Seite eine hohe Aufgeschlossenheit war, eben sich sowohl was die Hereinnahme weiterer Kriterien, als auch dieser unterschiedlichen Standards, also, dieser auf dieser praxisgestaltenden Wissenschaft, diesem Spezifikum gerade dieses Wissenschaftsbereichs, gerade sich darauf sehr einzulassen, und dann nicht nur so die allgemeinen Evaluationskriterien im Wissenschaftsbereich schlichtweg anzuwenden, sondern das wirklich zu adaptieren hier auf die spezifische Fragestellung. Das war ganz wichtig. Und da war, glaube ich, die Erfahrung etwa von Herrn Terhart in der sogenannten „Terhart-Kommission", die ja nach ihm auch benannt worden ist. Also da haben wir alle, oder haben sehr viel Erfahrungen eingebracht und die sind gut zusammengekommen. Und da wurde nicht die eine Erfahrung als die eigentlich aber wichtigere verabsolutiert. Da war angenehm. 01:03:56-0

I: Also, man war insgesamt der Sache offen gegenüber gestanden. Versucht, aus allem möglichen etwas... 01:04:00-0

H: Also die „gemeinsame Lernbereitschaft". Ich fand, das war wirklich eine „professionelle Lerngemeinschaft", die da entstanden ist, wir sind da alle klüger wieder rausgekommen. 01:04:07-3

I: Darf ich das zitieren?! Das ist eine wunderbare... Professionelle Lerngemeinschaft! Was... 01:04:11-6

H: Professionelle Lerngemeinschaften sind in der Schulentwicklungsforschung und in der Theorie der Schulentwicklung ja ein Spezifikum. Und ich habe wirklich den Eindruck gehabt, dass das da gelungen ist. Eine professionelle Lerngemeinschaft. Aus sehr unterschiedlich besetzten Bereichen zustande zu bringen. 01:04:30-2

I: Ich denke, natürlich, solche Personen wie Professor Baumert, die auch in der Vergangenheit schon in mehreren solchen Kommissionen tätig waren und dann vielleicht auch auf der anderen Seite Mitglieder, die noch nicht diese Erfahrung mitgebracht haben. Konnte man da erkennen, dass die , die neu waren, oder neuer- oder kann man das sagen- im Laufe der Arbeit, dass ihre, ja nicht

Beiträge, sondern ihre Äußerungen sich da vielleicht verändert hatten. Durch die Tatsache, dass sie wirklich unheimlich schnell dazugelernt haben, auch durch diese Arbeit. 01:05:14-0 01:05:15-6

H: Also, ich glaube, dadurch, dass wir nicht immer im Plenum waren, sondern durch die Begehungen sind ja, ist ja die Kommunikationsdichte in diesen Untergruppen auch anders in Gang gekommen. Und danach verlaufen Plenumdiskussionen dann anders. Weil man bereits in diesen kleineren Gruppen auch die Erfahrung, die man dort als Mitglied dieser Gruppe gemacht hat, die gibt man dann ja ins Plenum weiter. Da geht es nicht nur um den Austausch persönlicher Meinungen, sondern da ist man ja auch immer ein Stück Brückenmitglied zwischen Untergruppe und Gesamtplenum und das, glaube ich, hat das Diskussionsverhalten ein Stück weit gleichmäßiger gemacht. Wenn man also, Tonbandaufnahmen von den ersten Sitzungen und von den späteren gemacht hätte, dann wäre sicherlich ein anderes, ein anderes Beteiligungsverhalten, wenn man das wie ein Soziogramm darstellen würde, hätte man da ein anderes Netzwerk abbilden können, als zu Anfang. Die Dominanz, die anfängliche Dominanz ist später in Richtung mehr Egalität, ohne bei vollkommen ausgependelter Beteiligung zu sein, zu landen. Aber es ist auf jeden Fall auf eine deutlich stärkere Ausgeglichenheit hin, hat sich das nach meiner Wahrnehmung, hat sich das hin entwickelt. 01:06:33-1

I: Gut. 01:06:37-7

H: Ansonsten muss man einen Mechanismus einfach noch erwähnen bei solchen Gruppen. Diejenigen, die auch schriftliche Beiträge formulieren, nehmen natürlich auch noch mal in spezifischer Weise Einfluss. Wer sich da sehr stark zurückhält, hat weniger Chancen seine Position zu vertreten und die traurige Tatsache, dass das eben leider auch viel Arbeit macht, muss man dann in Kauf nehmen. Und da hat sich natürlich auch nicht jedes Mitglied in gleicher Weise engagiert. Aber ich denke schon, dass darüber das wir ja, fallbezogen, wenn man so will, wir haben also zwei Berichterstatter für die einzelnen Standorte benannt, hatte jeder die Chancen, sich selber zu beteiligen und sich einzubringen und das ist zum Teil eben, waren das spezifischere Beiträge, die auch gezielt erbeten wurden, etwa von der Sonderpädagogik, der Berufspädagogik und so weiter. Die dann noch mal zusätzlich Positionen aufgegriffen haben. Aber ich fand, dass es eben, bei diesem Umfang des Berichtes ist das ja auch gar nicht anders machbar, dass sich da doch wirklich viele Leute für verantwortlich gefühlt haben und da wirklich einbringen konnten. Und dass es dann eine Endredaktion gab, mit der auch alle leben konnten und die alle noch mal kommentiert haben und wo dann letzte Sachen eingearbeitet wurden. Das fand ich einen ausgesprochen angenehmen Produktionsprozess. Habe ich auch schon andere erlebt. 01:08:07-6

I: Würden Sie sagen, es gab besondere Probleme, insgesamt bei der Arbeit. Irgendetwas, was speziell, also, eine Nuss zu knacken galt? Vielleicht lag es an der Institution, vielleicht lag es...ja,

ich weiß es nicht.... 01:08:24-2

H: Also, es gab eine gesetzte Beschränkung unseres Ermittlungsauftrages, die ich nicht unproblematisch fand. Und ich glaube, dass man gut daran tut als Kommission den, die Reichweite des Auftrages auch rechtzeitig sehr klar zu bestimmen. Also ich hätte mir gewünscht, dass diese Entscheidung, dass zwei Strukturentscheidungen offener gewesen wären, seitens der politischen Auftraggeber. Einerseits die Vorentscheidung, bei uns gibt es PHs und das soll auch so bleiben. Das fand ich unnötig. Wenn man sich eine solche Kommission leistet, dann sollte man die Diskussion auch soweit öffnen, dass man die Vor- und Nachteile noch einmal so deutlich abwägt, dass man sich da nicht vorher schon festlegt. Die Kommission hat ja nicht etwa gesagt, dass ist auch gut so, dass diese Struktur so ist, sondern wir haben zur Kenntnis genommen, unser Auftrag reicht soweit und weiter nicht. Und die zweite Vorentscheidung, wo ich auch fand, dass die politischen Auftraggeber, hätte ich mir an der Stelle offener gewünscht, da ging es um die Frage der Studiengangsgestaltung, also, die Einführung, die Umsetzung des Bologna-Prozesses. Da hätten wir uns eigentlich auch eine etwas stärkere Offenheit gewünscht. Das, was auf jeden Fall klar war, dass überall Modularisierungsüberlegungen schon in Gang waren, aber inwieweit das denn nun in Richtung B.A./ M.A.- Strukturen geht, also, da hätte ich mir von Vorneherein mehr Offenheit gewünscht. Also, es ist bei solchen Missionen, bei solchen Evaluationen, finde ich, auf der einen Seite wichtig, sich über die Eckpunkte klar zu sein, andererseits ging das also auch nicht, im Sinne von alles- möglich und alles- in- Frage- stellen, sondern es muss schon fokussiert sein, eine Fragestellung. Aber es sollten dann auch wirklich Entscheidungsalternativen aufgezeigt werden können. Wenn man die schon zu früh zu sehr einengt, dann erlahmt auch leicht das Interesse. Man muss gucken, wie bedeutungsvoll ist es denn eigentlich jetzt, was wir hier tun. 01:10:44-9

I: Ja, ja, sicher. Und es war nicht klar eben von Anfang an, dass man da an dieser Struktur in keinster Weise etwas machen dürfte... 01:10:56-2

H: Es deutete sich zu Anfang, haben wir da noch mal Klarheit hergestellt über die Frage, wie unabhängig ist jetzt diese Kommission in ihrer Meinungsäußerung und welche Offenheit besteht hinsichtlich bestimmter Empfehlungen überhaupt. Und als gesetztes Kriterium war also diese Nicht-Antastbarkeit dieser Doppelstruktur. Lehramtsausbildung gymnasialer Bereich bleibt an den Universitäten und die andere... Da kann man nachdenken über die verschiedensten Kooperationsmodelle, vielleicht auch hinsichtlich Fachhochschulen und PHs, aber eine Integrationslösung wird nicht gewünscht, wie in den anderen Bundesländern das jetzt auf Uni-Ebene erfolgt ist. Das wird zum Teil für unsere ausländischen Mitglieder etwas schwierig nachzuvollziehen, weil da wird das ja gerade nachgeholt, oder da wurden, wenn da Pädagogische Hochschulen eingerichtet wurden, war das immer der Schritt von der seminaristischen Ausbildung

und insofern war ihnen nicht so ohne weiteres leicht nachvollziehbar, warum denn ein einziges Bundesland jetzt hier auf dieser besonderen Struktur, erst einmal warum sie sich von Vorneherein darauf festlegen. Aber da hätte, dann hätte die Kommission die Arbeit nicht aufnehmen können, in der personellen Zusammensetzung. Das haben wir intensiv vorher noch mal diskutiert, uns über den Auftrag verständigt und dann haben wir beschlossen, dass so wie wir es im Vorwort dann festgehalten haben. So, das als die Ausgangspunkte zu nehmen. 01:12:43-6

I: Also, die eigene Rolle noch einmal erklären und zu sagen, zu dem sind wir im Stande, zu dem sind wir nicht im Stande. 01:12:49-4

H: Genau. 01:12:52-6

I: Weil das geht nicht mit unseren Vorstellungen überein, und damit dann mit dem Ministerium ein Verständnis, ein Übereinkommen zu bekommen. 01:13:03-9

H: Ganz genau. 01:13:06-1

Piepton 01:13:11-0

I: Ups 01:13:11-9

H: Das blinkt da vor sich hin? 01:13:14-5

I: Ja, nein, das ist okay. Das Remain, das ist okay. 01:13:22-0

H: Für die Zeitplanung wollte ich nur noch mal sagen, um viertel nach eins fängt meine Vorlesung an. 00:00:04-9

I: Nein, wir müssen...Jetzt sagt das immer noch forty-nine, jetzt verstehe ich das auch nicht. 00:00:10-1

Stimme im Hintergrund (SH redet mit einer anderen Person) 00:00:14-8

I: Den Rest kann ich ja aufschreiben. Ja, eine Frage zu den verschiedenen Bereichen, jetzt Sonderpädagogik oder empirische Päda... (Piepton) Ja, gut, dann mach es aus, den Rest kann ich ja aufschreiben. Will der nicht... 00:00:30-6

H: Vielleicht haben Sie die Batterie nicht richtig herum reingelegt 00:00:33-0

I: Habe ich, aber es liegt daran, dass das nur eine bestimmte... 00:00:36-8

bricht ab

APPENDIX H

Ulm Declaration

Ulmer Erklärung

zur berufsbezogenen Gymnasiallehrer-Ausbildung

an den Landesuniversitäten von Baden-Württemberg

Die Unterzeichner wenden sich mit dieser Ulmer Erklärung an den Landtag, an die Landesregierung und an die Öffentlichkeit, um die Aufmerksamkeit auf besorgniserregende Strukturentwicklungen in der baden-württembergischen Gymnasiallehrer-Ausbildung zu lenken, die für eine berufsbezogene Ausbildung, für deren Wissenschafts- und Forschungsfundierung in Pädagogik, Psychologie und Fachdidaktik sowie für die empirische Bildungsforschung im Lande katastrophale Folgen haben können.

Die Landesregierung von Baden-Württemberg hat beschlossen, die universitäre Ausbildung der Gymnasiallehrer erstmals in Richtung einer Berufsorientierung zu reformieren. Die neue Prüfungsordnung ist am 1.4.2001 in Kraft getreten. Die wesentlichen Neuerungen sind: Erhöhung des Studienanteils in Pädagogik/Schulpädagogik und Pädagogischer Psychologie sowie die Einführung verpflichtender fachdidaktischer Übungen und des Praxissemesters. Damit ist in Baden-Württemberg nach Jahrzehnten des Stillstands der Einstieg in eine moderne universitäre Gymnasiallehrer-Ausbildung gelungen.

Im Vorfeld dieser Neuerungen wurde immer wieder auf die erforderlichen Personal- und Sachmittel für die Verwirklichung dieser Reform hingewiesen; denn in Baden-Württemberg sind an den Universitäten insgesamt weniger Professuren für die Lehrerbildung vorgesehen als in vielen Bundesländern allein in einem einzigen Institut. Professuren für Fachdidaktik fehlen an den baden-württembergischen Universitäten bisher völlig!

Da das Wissenschaftsministerium für Kabinettsbeschlüsse mitgetragen hat, war davon auszugehen, dass den Universitäten die seit langem reservierten zusätzlichen Professorenstellen für Pädagogik bzw. Pädagogische Psychologie zugewiesen und für die erforderlichen Personal- und Sachmittel für die neu eingeführten Pflichtlehrveranstaltungen in Fachdidaktik sowie für die Mitwirkung beim Praxissemester in einem Ausbauplan bereitgestellt würden. Für eine Übergangsphase müssten Übergangslösungen für die Sicherstellung des Pflichtlehrangebots gefunden werden (Lehraufträge, Kooperation mit Studienseminaren und Pädagogischen Hochschulen):

In dieser Situation überrascht die Nachricht, dass die Universität Konstanz, die Pädagogische Hochschule Weingarten und das Studienseminar Rottweil einen Vertrag abgeschlossen haben, der vorsieht, dass die Universität sich auf die fachwissenschaftlich Lehrangebote zurückzieht und die pädagogischen, pädagogisch-psychologischen sowie die fachdidaktischen Anteile von der Pädagogischen Hochschule und die fachpraktischen Anteile vom Studienseminar erbracht werden. Die Universität Konstanz hat ihren einzigen Lehrstuhl für Pädagogik umgewidmet! – Einen analogen Vertrag hat kürzlich die Universität Karlsruhe mit der dortigen Pädagogischen Hochschule und dem dortigen

Studienseminar abgeschlossen.

In der Presse wird gefragt, ob diese „Modelle“ der Kostenersparnis Schule machen könnten. Wenn diese Regelung Schule macht, dann bedeutet dies nichts anderes als den Ausstieg aus der eben erst beschlossenen Ausbildungsreform. Dann ist absehbar, dass

- die jetzt noch an den Universitäten für die Lehrerbildung ausgewiesenen Professuren für Pädagogik und Pädagogische Psychologie bei eintretender Vakanz gestrichen und durch Lehraufträge ersetzt werden;

- diese Fachrichtungen damit aus den Universitäten verschwinden werden und sie dann nicht mehr zur Verfügung stehen werden für die Ausbildung für pädagogische Lehrberufe im Hauptfach, für pädagogische Anteile in berufsbezogenen BA/MA-Studiengängen, für die Fundierung der Hochschuldidaktik, für die Forschung und damit für die Förderung des wissenschaftlichen Nachwuchses, die Rekrutierung von Hochschullehrern sowie von Experten im außerwissenschaftlichen Bereich und in Landesbehörden;

- die dringend erforderlichen Investitionen in den Aufbau der Fachdidaktiken für Ausbildung, Forschung und Weiterbildung sowie in die Pädagogische Psychologie für die Aus-, Fort- und Weiterbildung von Schulpsychologen und Beratungslehrern, Erziehungs- und Bildungsberatern gar nicht erst erfolgen werden;

- die Gymnasiallehrer künftig an den baden-württembergischen Universitäten ohne eine grundständige berufswissenschaftliche Ausbildung und ohne wissenschaftsbasierte Fort- und Weiterbildung bleiben müssen, was der Dequalifizierung eines Berufsstandes gleichkommt und dass – entgegen den erklärten Zielen des für die Personalrekrutierung zuständigen Kultusministeriums – keine Kapazitäten für eine wissenschaftsbasierte Unterstützung der Professionalisierungsbestrebungen der Gymnasiallehrer mehr vorhanden sein werden;

- keine Ressourcen für die Begleitforschung zur Schulentwicklung und zu didaktischen Innovationen in den öffentlichen Schulen (Implementierung neuer Lehr-Lern-Formen, Lernen mit Neuen Medien usw.) mehr vorhanden sein werden;

- keine Expertise mehr entwickelt werden kann für die Sicherung der Bildungsqualität der Gymnasien, für Begleitforschung zur pädagogisch-didaktischen Effektivität von Investitionen (z.B. der Medienoffensiven des Landes und der Kommunen) und Lehrplanrevisionen sowie für forschungsbasierte Politikberatung.

Die Vorgänge in Konstanz und Karlsruhe stehen in krassem Gegensatz zu den jüngsten Empfehlungen der Kultusministerkonferenz und der Expertenempfehlungen in mehreren Bundesländern, die einhellig von der Verstärkung der berufswissenschaftlichen Anteile des universitären Lehramtsstudiums, besonders der Fachdidaktiken, ausgehen.

Die Landesregierung von Baden-Württemberg wird daher dringend gebeten, die in Konstanz und Karlsruhe erfolgten Vertragsabschlüsse in ihrer Geltung und Dauer zu begrenzen und das Wissenschaftsministerium anzuhalten, im Zusammenwirken mit dem

Kultusministerium und den Landesuniversitäten für eine sachgerechte Umsetzung des Kabinettsbeschlusses zur Neugestaltung der Gymnasiallehrer-Ausbildung Sorge zu tragen.
Prof. Dr. Peter Drewek, Universität Mannheim; Prof. Dr. Wolfgang Edelstein, Max-Planck-Institut für Bildungsforschung, Berlin; Prof. Dr. Helmut Fend, Universität Zürich; Prof. Dr. Martin Fromm, Universität Stuttgart; Prof. Dr. Ulrich Herrmann, Universität Ulm; Prof. Dr. Manfred Hofer, Universität Mannheim; Prof. Dr. Jürgen Oelkers, Universität Zürich; Prof. Dr. Heinz-Elmar Tenorth, Humboldt-Universität zu Berlin; Prof. Dr. Ewald Terhart, Universität Bochum; Prof. Dr. Rudolf Tippelt, Universität München

Die Ulmer Erklärung muss in einem Punkt korrigiert werden: An der Universität Karlsruhe blieb der Lehrstuhl erhalten und wurde mit Professor Rekus besetzt.

APPENDIX I

Schelten / Achtenhagen statement

Gründe gegen Kooperationsmodelle zwischen Pädagogischen Hochschulen und Fachhochschulen im Studium Höheres Lehramt an beruflichen Schulen

Gegen eine Kooperation zwischen Pädagogischen Hochschulen und Fachhochschulen im Studium Höheres Lehramt an beruflichen Schulen sprechen konzeptionelle und organisatorische Gründe.

Wesentliche konzeptionelle Gründe für ein universitäres Studium für das Höhere Lehramt an beruflichen Schulen und damit gegen Kooperationsmodelle zwischen Fachhochschulen und Pädagogischen Hochschulen sind in folgenden Punkten zu sehen:

1. Aufgrund immer kürzer werdender Innovationszyklen und Entstehung immer wieder neuer Berufe in kurzer Folge wachsen die Ansprüche an die fachliche Kompetenz der Lehrer an beruflichen Schulen immens. Allein ein wissenschaftliches Studium an der Universität auf höchstem theoretischen Niveau, das auf eine inhaltliche und denkmethodische Erschließungs-fähigkeit abzielt, kann dem Rechnung tragen. Es gehört zu den gesicherten Erkenntnissen, dass Transferfähigkeit und Flexibilität am ehesten durch eine vertiefte theoretische Ausbildung gefördert werden.

2. Bei einem Fachhochstudium steht pragmatisch die Vermittlung bereits entwickelter, regelgebundener Handlungsroutinen im Vordergrund. Ein Universitätsstudium legt mehr Wert auf Einsichts- und Problemlösungsstrukturen, die zur eigenständigen Entwicklung neuer Handlungsroutinen führen. Letzteres wird zunehmend für das Höhere Lehramt an beruflichen Schulen wichtig und spricht für das Studium ausschließlich an der Universität.

3. International gilt das duale System der Berufsbildung in Deutschland für die nichtakademische berufliche Bildung als sehr erfolgreich. Dieser Erfolg hängt wesentlich von der wissenschaftlichen Lehrerbildung für berufliche Schulen an den Universitäten ab. Andere Länder (z.B. Finnland) heben systematisch das Ausbildungsniveau für Berufsschullehrer an – und setzen dabei explizit das Ziel „universitäre Ausbildung“.

4. Das Studium für das Höhere Lehramt an beruflichen Schulen qualifiziert insbesondere für Berufsschulen, Berufsfachschulen, Technikerschulen, Fachoberschulen und berufliche Gymnasien. Es setzt in der beruflichen Fachrichtung, in dem zweiten Unterrichtsfach und in den Erziehungswissenschaften auf eine universitäre Bildung, bei der Forschung und Lehre zu einer vorwiegend wissenschaftsbezogenen Ausbildung verbunden werden. Dieses sichert ein hohes theoretisches Niveau.

5. Ein kompetentes didaktisches Reduzieren in der beruflichen Fachrichtung und dem zweiten Unterrichtsfach ist erst von einem hohen wissenschaftlichen theoretischen Niveau aus möglich.

6. Die Attraktivität eines Lehrerberufes für berufliche Schulen, der über eine jahrzehntelange Entwicklung erst mit der universitären Lehrerbildung sein Ansehen hat festigen können, wird bei einer Verlagerung seines Studiums an Fachhochschulen in Kooperation mit Pädagogischen Hochschulen erheblich abnehmen und leistungsstarke, universitätsorientierte Studierende von diesem Lehramtsstudium abhalten.

Es besteht keine Veranlassung, ein kooperatives Modell einer Lehrerbildung für das Höhere Lehramt an beruflichen Schule zwischen Fachhochschulen und Pädagogischen Hochschulen anzubahnen. An Stelle eine kooperativen Modells sollte auf ein additives Modell gesetzt werden: Fachhochschüler mit Diplomvorprüfung, besonders aber Fachhochschulabsolventen, gehen über geregelte Einstiegswege, wie bereits praktiziert, in das Lehramtsstudium für berufliche Schulen an der Universität über.

Aus organisatorischer Sicht spricht gegen eine Kooperation im Studium Höheres Lehramt an beruflichen Schulen zwischen Fachhochschulen und Pädagogischen Hochschulen das Folgende: Es werden vereinzelt Stellen für Berufspädagogik ohne weitere Ausstattung an Pädagogischen Hochschulen mit zu Teilen ausgeprägt geringer Forschungsorientierung eingerichtet. Es werden unterkritische Größen geschaffen, die kaum Forschung erbringen und eher Lehre betreiben können, die wiederum nicht mit selbst erbrachter Forschung wissenschaftsbezogen erfolgt.