

**The Assessment of Undergraduate Final Year Projects:
A study of academic professional judgment**

Suellen Butler Shay

Thesis Presented for the Degree of Doctor of Philosophy
In the Department of Education
University of Cape Town
2003

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

TABLE OF CONTENTS

Acknowledgments	i
Glossary	ii
Acronyms	iii
Abstract	iv
Chapter 1: Introduction	1
1.1 Focus of the study	1
1.2 Context for the study	3
1.3 The central problem and research questions	6
1.4 Overview of the thesis	8
Chapter 2: The Problem.....	10
2.1 A problem in practice	10
2.2 Educational Assessment: multiple agendas	12
2.3 Assessment as social practice	14
2.4 Conclusion	21
2.5 A problem in search of theory	23
Chapter 3: Theory.....	25
3.1 Introduction	25
3.2 The search for foundations	27
3.3 Beyond objectivism and relativism	29
3.3.1 Validity inquiry from a psychometric perspective	29
3.3.2 Validity inquiry from an interpretive perspective	36
3.4 A social theoretical perspective	43
3.4.1 The double reality of social life	43
3.4.2 Field and habitus	46
3.4.3 Validity inquiry from a social theoretical perspective	50
Chapter 4: Methodology.....	54
4.1 Introduction	54
4.2 Critical discourse analysis	55
4.2.1 A rationale for CDA	55
4.2.2 Fairclough's CDA	57
4.3 Methodological approach	64
4.3.1 The analysis of social structure	66
4.3.2 The analysis of conjuncture	67

4.3.3	The analysis of event	69
4.3.4	The analysis of text	70
4.4	Research design	73
4.4.1	Exploratory study	74
4.4.2	Main study	83
4.4.3	Summary	93
Chapter 5: Analysis.....		94
5.1	Introduction	94
5.2	Analysis of higher education field	98
5.2.1	The academic field	99
5.2.2	Academic capital, positions and rules	103
5.2.3	Academic practice: the logic of external examination	107
5.2.4	Conclusion	109
5.3	Humanities case study	110
5.3.1	Validating practices	110
5.3.2	Validating arguments	115
5.3.3	Conclusions	137
5.4	Engineering case study	139
5.4.1	Validating practices	139
5.4.2	Dimensions of disposition	144
5.4.3	Where differences play out	157
5.4.4	Summary	174
Chapter 6: Discussion and Conclusions.....		176
6.1	Introduction	176
6.2	Discussion	177
6.2.1	Academics' 'feel for the game'	178
6.2.2	Academic interpretive matrices	183
6.2.3	Academic stakes	185
6.2.4	The exercise of academic power	188
6.3	Conclusions	189
Chapter 7: Implications for practice.....		194
7.1	A focus on value	194
7.2	Implications for validation practice	196
7.2.1	Validation as a community process	196
7.2.2	Validation which prioritizes community values	198
7.2.3	Validation which respects difference	200
7.2.4	Validation which acknowledges power	202
Appendices.....		205
References.....		213

ACKNOWLEDGMENTS

I would like express my sincere appreciation to:

- Pamela Moss who intellectually inspired me throughout this study.
- Kay McCormick and Martin Hall, my supervisors, who in very different but significant ways strengthened my final product.
- Peggy Nightingale who has been a source of inspiration and encouragement.
- Those who participated in my research, particularly my key informants, who were both willing and able to critically reflect on their own practices.
- Carolyn Butler for her tireless proof-reading of the final product.

To begin, progress through and finally complete a PhD while at the same time raising three young children has been no easy task. For this I am deeply grateful to:

- Don Shay, my husband, who has been nothing but supportive of me all along the way.
- Ron and Carolyn Butler, my mother and father, who filled the gap over and over again.
- Aziza Nolan and Vanessa Choonoo, our child-minders, who did all the rest.
- Charissa, Danielle and Emily Shay, who each in their own way have kept me sane.

Others I cannot fail to mention who have provided me with essential physical, intellectual and emotional resources:

- Rob Moore, Robin Benny, Rosemary Kearny, Doreen Scott and Judy Baron.

I am also grateful to UCT's International Academic Programmes Office for facilitating a grant from the University of Michigan which enabled my stay there as a visiting scholar. I also acknowledge with gratitude a doctoral grant from the National Research Foundation. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the National Research Foundation or any other person or organization listed above.

And finally, to Sophia, the Goddess of *phronesis* to whom I owe it all.

GLOSSARY

Assessment task	What students are required to do, for example, an essay question, a project topic.
Assessors	Those who assess student performance, for example, academics and tutors.
Community of practice	A community of people who are involved in a sustained pursuit of a common social project.
Interpretive matrix	The complexity of contingencies which assessors bring to bear in their assessment of student performance.
Reliability	The consistency of assessment results.
Student performance	The actual product or multiple products which students submit for assessment purposes.
Validation	The process of evaluating the soundness of interpretations. Validation is a form of evaluation.
Validation practices	All the mechanisms which a community of practice has in place to ensure the reliability and validity of the assessment of student performance.
Validity	The soundness of the interpretations of student performance.
Validating argument	The actual discourse which academic staff use to interpret student performance and justify their interpretations to students, colleagues or the researcher.
Validity inquiry	The systematic investigation through research of validity issues in order to theorize the concept.
Validators	Those who validate assessment interpretations, for example external examiners.

ACRONYMS

CDA – Critical discourse analysis

DEM – Departmental examination meeting

EF – The engineering field of study

ES – The area of specialization within the field of study

HOD – Head of Department

MW – Marking workshop

OCDL – Oxford Concise Dictionary of Linguistics

SAU – South African University

ABSTRACT

The premise of this study is that the assessment of student performance is an interpretive process. This raises a fundamental validity question: on what basis do academic communities evaluate the soundness of their interpretations? The central problem which this study explores is how academic assessors validate their interpretations of student performance on complex tasks. I explore this problem by focusing specifically on the assessment of final year undergraduate projects through two case studies of disciplinary communities of practice, one in the Humanities faculty and the other in the Engineering faculty of a South African university.

Drawing on Bourdieu's theory of social practice on the one hand, and the methods of critical discourse analysis and ethnography on the other, I construct a theory and method of inquiry which illumines aspects of assessment as an interpretive process, aspects which have been obscured by traditional approaches to assessment. This analysis privileges the *context* of assessment, the inevitability of *difference* in assessment interpretations and the equally inevitable effects of *power*. My methodological approach identifies four elements which constitute social practice – social structure, conjuncture, event and text. These constitutive elements operationalize into a series of analytical stages which expose different aspects of social practice. My approach is consistent with Fairclough's method of critical discourse analysis, although I also include ethnographic methods.

My case study analysis was guided by a series of premises from which my key questions emerged. Firstly, academic communities of practice, by definition, share some common ground. Thus I inquired: what is the basis of this common ground? What explains the manner in which academic assessors, with minimal co-ordination or explicit articulation of criteria, are able to get on with the numerous classificatory acts which are required of them as members of a particular disciplinary field?

Secondly, despite this common ground it is not unusual to find multiple interpretations of the same instance of student performance. Thus I inquired: what explanation is there for these differences in interpretation? What can dissensus within a community of interpreters expose about the complexities of assessment as an interpretive process?

Thirdly, reaching consensus between assessors about what value to assign an instance of student performance is often difficult. The negotiation of interpretive differences can be a site of struggle, and this struggle implies that there are certain stakes in the resolution. Thus I inquired: in the context of collegial dialogue, what is at stake in the resolution of differences?

Finally I asked: if consensus is not a given, how then are differences resolved? Specifically, how is power exercised in the resolution of these differences?

The significant contribution of this study is to enrich our understanding of the complex interpretive environment out of which academic professional judgment emerges. By 'drilling down' into differences of interpretation, I illuminate the multiple contexts which constitute academics' judgments. These multiple contexts include: the macro-social conditions of the field and the ways in which these conditions legitimate particular classificatory systems; the meso-level context of disciplinary (and sub-disciplinary) communities of practice with their specific epistemological orientations; and finally the micro-level context of assessors' interpretive matrices, which are significantly constituted by assessors' disciplinary interests, professional experience and levels of involvement. My study argues that 'context matters' by exposing the multiple layers of contingency that shape assessors' interpretive acts in predictable and unpredictable ways.

Finally, the study unexpectedly shed light onto the patterns of rationality which underlie academic professional judgment – an alternative form of rationality which is neither 'objectivist' or 'subjectivist' but is something beyond either of these positions. Academic professional judgment is constituted by the tension between macro and micro

contexts and between subjective and objective modes of knowledge. This alternative rationality is pragmatic, variable, context-dependent and value-based – a form of rationality which has been greatly undervalued, and yet it is essential to an understanding of assessment as an interpretive process and to professional judgment more generally.

Chapter One: Introduction

1.1 Focus of the study

Educational assessment plays an increasingly dominant role in contemporary society. Over the past two decades assessment has come to be perceived by governments and other authorities as a powerful technology for both educational reform and quality control. The former refers to assessment as an instrument for improving the quality of teaching and learning. The latter refers to assessment as a mechanism for assuring the quality of educational provision to those outside the university. Thus at the beginning of the 21st century assessment finds itself at the nexus of complex and highly contested expectations for social and political change. This is a matter of concern for scholars and practitioners alike who question the extent to which assessment can fulfil these multiple, and some would argue contradictory, agendas. Foundational to their concerns is doubt about whether assessment can be used to measure and categorize individuals and institutions for a variety of political, social and economic purposes.

Out of these concerns arises a small but growing body of research which explores the social underbelly of educational assessment. It argues that assessment is best understood as a complex *social*¹ practice. There are two themes which are central to this work. The first theme is that assessment practices can only be understood within the social, cultural, and economic contexts in which they occur. The second theme is that the experiences, beliefs and expectations of both the assessed and the assessor constitute the meanings of assessment outcomes. Researchers in this area argue that one of the challenges for this sociological approach to assessment is to develop both theories and methodologies for exploring this complexity. It is to this body of research that this study contributes – to a deeper understanding of assessment as a social practice.

This study sets out to explore how academic assessors validate their interpretations of student performance² on complex tasks³. Complex tasks refer to those which allow

¹ I use italics for emphasis.

² I use the term 'student performance' to refer to the actual product or multiple products which students submit for assessment purposes.

³ I use the term 'task' or sometimes 'assessment task' to refer to what students are required to do, for

students substantial latitude in how they interpret and respond to the requirements. They require multiple skills and forms of knowledge, and the completed projects generally require expert judgment (Moss 1992). The premise of this study is that the assessment of complex tasks is an interpretive process. I use the term ‘assessment’⁴ to refer to the interpretation of student performance for the purposes of assigning some value to the performance. ‘Validity’ refers to the soundness of these interpretations (Nitko 2001). Since validity is not all or nothing but always a matter of degree (Messick 1989), establishing the validity of assessment interpretations is an on-going process. I use the term ‘validation’ to refer to the process of evaluating the soundness of interpretations. Communities of practice⁵ have a variety of mechanisms by which they evaluate the soundness of their interpretations, what I refer to as ‘validation practices’. The bases upon which an interpretation is considered to be sound will depend on the notions of science which these communities value; in other words, different scientific traditions will have different criteria for the evaluation of their interpretations. A discussion of different approaches to the validation of assessment interpretations is the central task of Chapter 3.

As assessment acquires an increasingly dominant profile within the societal landscape, validation is a matter of considerable consequence, not only for individual learners and educators, but for institutions and the wider society. I would argue that it is particularly relevant in the South African context given:

- Increasing diversity within the student population, diversity of cultural capital as well as educational preparedness, which results in an increasing variety of student responses to assessment tasks
- Anticipated diversity in staff composition which could have the effect of multiplying the potential number of interpretations of student performance

example, an essay question, a project topic.

⁴ In this thesis I use single quotation marks to signal terms or concepts which I wish to draw special attention to. I use double quotation marks for quotations. The exception is the use of single quotes for a quotation within a quotation.

⁵ ‘Community of practice’ is a term used by Wenger (1998) to refer to communities of people involved in the sustained pursuit of a common social project. I use the term specifically to refer to all the people, materials, technology and procedures which are involved in the assessment of final year projects.

- Major curriculum restructuring initiatives where fundamental notions about the purpose of undergraduate education are being reviewed and recast
- Increasing national and institutional pressure for accountability in relation to educational outcomes through, for example, quality audits
- Major resource cut-backs and increasingly large classes which might create an ethos of general disaffection among academic staff who may be tempted to resort to minimalist approaches

1.2 Context for the study

In order to pursue this interest I focus specifically on the assessment of final year projects within the context of two disciplinary communities of practice, one in the Humanities faculty and the other in the Engineering faculty of a South African university (henceforth referred to as SAU). SAU is a medium-sized, contact university, located in one of South Africa's large urban areas. It is one of South Africa's historically white, English-speaking institutions⁶.

It is a fairly prevalent practice in undergraduate programmes around the globe to require students to conduct a major project as a final culmination of their undergraduate learning experience. These are sometimes referred to as 'capstone' or 'senior' projects. I will use the term 'final year projects'. The management and assessment of these projects vary tremendously across contexts, and I will not attempt to summarize the differences. I focus instead on the particular institutional context of SAU and the two faculties where my case studies are located.

In these contexts project topics are proposed by academic members of staff. In the Humanities context these topics are often no more than an issue within a broad area of study, or they may take the form of an open-ended problem or question. In the Engineering context the topics take the form of a brief description (usually about a

⁶ 'Historically white' universities refers to universities which under apartheid were designed to provide higher education to South Africa's white population. These universities were further differentiated by language as English-speaking or Afrikaans-speaking.

paragraph in length) where staff members give relevant background information on the problem to be researched and the research objectives. They may also note particular courses which students should have had as prerequisites⁷. In addition to these brief descriptions, students in both contexts are given general guidelines on the final structure of their project reports.

These projects require the collection, analysis and synthesis of data from multiple sources, for example, literature, fieldwork, and laboratory. Students generally have several months to complete them. Depending on departmental practices and individual supervisor preferences, students may be required to work in groups or individually. The projects require a whole range of conceptual, theoretical and practical skills. The final products take the form of a written research report, but students may also be required to make an oral and visual presentation. They are typically assessed by a supervisor or a member of staff who has been involved in the research process to some degree. They are generally high-stakes since many academic programmes require a pass in the project for graduation. The primary purpose of the final year project is to assess individual students' achievements in relation to an undergraduate curriculum of study in order to determine whether they are competent to graduate. An additional purpose which is gaining momentum is the use of final year projects as a measure of accountability to state, institutional and professional bodies⁸. Thus final year projects are an example of high-stakes, non-standardized, complex assessment tasks.

Given these features, final year projects pose a particular set of challenges in terms of validity, that is, the degree to which assessment interpretations are sound and appropriate to their uses (Nitko 2001). I shall now elaborate briefly on some of these challenges as

⁷ In order to protect the anonymity of the departments involved in the study I cannot offer specific examples of project topics.

⁸ For example, for the interim registration of qualifications on the National Qualifications Framework (see 5.2.1 for more on NQFs), academic programmes were required to stipulate what forms of 'integrative assessment' they had in place. Another example: in 2000 the Engineering Council of South Africa's audit of SAU's Engineering Faculty required academic programmes to specify whether their assessment "verified whether each student satisfies the specified outcome" (Engineering Council of South Africa, Standards and Procedures System, Document PE-75, 18 January 2000, p. 2).

they manifest themselves in the communities of practice which are the focus of this study.

Firstly, there is generally a low degree of standardization of the topics which students are given. Project topics are set out by the supervisors. Thus the final year project topics which students select, or which they are assigned, are as diverse as the research interests of the academic staff who supervise them. This diversity of topics suggests that the range of conceptual, theoretical and practical skills which the projects require will vary across students. In addition there are a range of logistical issues (e.g. access to people and organizations, availability of materials and technology) which impact on the final shape of the project and what it requires of students.

Secondly, the actual skills or competences which final year projects are purported to assess are seldom explicitly specified. Academic staff who set the projects would have more or less cogent explanations for what they perceive the projects to be assessing, but there is seldom any systematic investigation by the department into the extent to which the projects are aligned against a specific content and skill domain. In South Africa recent calls for accountability from institutional, national, and professional quarters have stressed the importance of more explicit learning outcomes and clearer articulation of assessment methods and criteria. However, academic programmes have been slow to respond to these calls for explicitness, and in some cases have openly resisted them (I will return to this in Chapter 5).

Thirdly, a further challenge is that final year projects comprise multiple instances of performance which assessors can and do take into account, often unconsciously, in assessing students' abilities. Some are inferred from the research process, for example, the degree of effort that the student is perceived to be investing in the research process, the capacity of the student to overcome particular unanticipated dilemmas, the degree of independence that the student displays, and the extent to which feedback is considered and incorporated. Other abilities are inferred from the student's performance product, for

example, the fluency of an oral or written presentation, the technical sophistication of design, adherence to layout guidelines, and thoroughness of a literature review.

Fourthly, there is seldom a standardized scoring memorandum, that is, a set of criteria and weightings which all assessors apply in a consistent way. This is not to suggest that there are no criteria. It is more a case that they are implicit throughout the process of setting, supervising and assessing the projects, for example, in the kinds of topics that are set, in the feedback students receive, and in the debates which arise among assessors. Some would argue that given the great diversity in topics (noted above) it may be difficult to arrive at a uniform set of criteria and weightings which are appropriate to all projects.

Fifthly, the assessors are involved in the research process, sometimes in significant ways. Their supervision styles vary greatly. But even those who take a hands-off approach are involved in the design of the topic, the supervision of the research process, and crucially, they are involved in interpreting the performance in order to allocate some kind of mark. As members of the same community of practice they are likely to share some common understanding of what constitutes 'good' research. On the other hand, there are also likely to be personal and professional differences which lead to diverging interpretations.

Despite all these challenges the assessment of undergraduates continues year after year. and on the basis of these assessments decisions are made which significantly affect students' lives: decisions about who graduates, who is admitted to graduate studies, and who is employable. These decisions also affect, less directly but no less significantly, staff reputations and institutional resource allocations.

1.3 The central problem and research questions

The central problem which this study seeks to explore is how academic assessors validate their interpretations of student performance on complex tasks such as final year projects. The focus of the study is thus on academic validation practices, a term which, as

I noted earlier, refers to all the mechanisms which a community of practice puts into place to ensure the soundness of their interpretations of student performance. I explore this problem within the context of final year undergraduate projects in two different communities of practice, one in the Humanities faculty and the other in the Engineering faculty of SAU. With respect to these validation practices my initial point of interest was how academics reach consensus on what value to assign a particular instance of student performance. I was also interested in explanations for the phenomenon of multiple interpretations of the same performance and the difficulty which academics often face in reaching consensus about its value. These initial points of interest evolved into a series of premises and a set of questions which guided my case studies analysis:

- Firstly, academic communities of practice, by definition, share some common ground. What is the basis of this common ground? What explains the manner in which academic assessors, with minimal co-ordination or explicit articulation of criteria, are able to get on with the numerous classificatory acts which are required of them as members of a particular disciplinary field?
- Secondly, I was specifically interested in the phenomenon of multiple interpretations of the same instance of student performance. What explanation is there for these differences in interpretation? What can dissensus within a community of interpreters expose about the complexities of assessment as an interpretive process?
- Thirdly, reaching consensus between assessors about what value to assign an instance of student performance is often difficult. The negotiation of interpretive differences can be a site of struggle, and this struggle implies that there are certain stakes in the resolution. Thus I inquired, in the context of collegial dialogue, what is at stake in the resolution of differences?
- Finally I asked, if consensus is not a given, how then are differences resolved? Specifically, how is power exercised in the resolution of these differences?

1.3 Overview of the thesis

I now provide an overview of the remaining chapters.

Chapter 2: The problem

The second chapter articulates in more detail the problem which this study explores. I review some of the literature on assessment from socio-cultural and interpretivist approaches. These approaches to the study of assessment serve to problematize some of the central assumptions which undergird the foundations of psychometric assessment. In addition they call for alternative perspectives on assessment, perspectives which privilege assessment as a social practice. I argue that what is needed in particular are studies of assessment practices which illumine the role of context, difference and power in the interpretations of student performance.

Chapter 3: Theory

The central task of the third chapter is to construct a theoretical perspective for assessment as an interpretive process. In justifying my own perspective, I explore three different approaches to validity inquiry⁹: the psychometric, the interpretivist and the social theoretical. Drawing on Bourdieu's theory of social practice, I argue for the particular ways in which a social theoretical approach will illumine academic validation practices.

Chapter 4: Methodology

The fourth chapter accounts for the methodological approach I have taken to the gathering and analysis of data. My approach can be characterized as critical discourse analysis with the incorporation of some ethnographic methods. This approach draws substantially on the work of Norman Fairclough, but also takes account of some of the methodological weaknesses for which his work has been criticized.

⁹ The term 'validity inquiry' refers to the systematic investigation of validity issues in order to theorize the concept.

Chapter 5: Analysis

The fifth chapter has three parts. In part one, I sketch in broad strokes some of the social conditions of constraint and possibility which ensure the on-going production and re-production of academic validating practices. In Bourdieu's terms this is the analysis of the 'field'. The goal of this analysis is to sketch some of the contours of the field, to identify its forms of 'capital', and to expose the principles which inform academic classificatory acts. In parts two and three, I explore how individuals take up the positions constituted for them by the field. Each case begins with a description of the actual community with a particular focus on the final year projects and the procedures, events, rules, roles and technologies which enable members of these communities to implement final year projects from year to year. I then turn to a closer scrutiny of actual assessment discourse to explore the particular interpretive matrices which assessors bring to bear on the assessment of student performance.

Chapter 6: Discussion and Conclusion

The sixth chapter discusses the findings to the questions which guided the study. I conclude with the ways in which this study has contributed to a deeper understanding of assessment as a social practice. The first contribution is a theoretical and methodological approach to the study of assessment practices. The second contribution of this study is an illumination of the multiple contexts which constitute academics' judgments. This in turn sheds light on patterns of rationality which underlie academic professional judgment – a rationality which I argue is neither objectivist or subjectivist.

Chapter 7: Implications for practice

In the final chapter I argue that the nature of academic professional judgment and its underlying forms of rationality are not simply theoretical issues; they have implications for how we conduct our task as an academic community of interpreters. Thus I discuss four major implications of the insights gained from this study for the construction of validation practices.

Chapter 2: The Problem

Henceforth my dear philosophers, let us be on guard against the dangerous old conceptual fiction that posited a “pure, will-less, painless, timeless knowing subject”; let us guard against the snares of such contradictory concepts as “pure reason”, “absolute spirituality”, “knowledge in itself”: these always demand that we should think of an eye that is completely unimaginable, an eye turned in no particular direction, in which the active and interpreting forces, through which alone seeing becomes seeing *something*, are supposed to be lacking; these always demand of the eye an absurdity and a nonsense. There is *only* a perspective seeing, only a perspective “knowing”; and the more affects we allow to speak about one thing, the *more* eyes, different eyes, we use to observe one thing, the more complete will our “concept” of this thing, our “objectivity”, be.
(Nietzsche cited in Bourdieu 1990, p. 28)

2.1 A problem in practice

Some years ago I had a conversation with a tutor (I give her the name Fatima) which intrigued me. I use this conversation as a starting point for teasing out some of the strands which comprise the knot of the problem which this study seeks to untangle and explore. She was very upset about a decision taken by the convener of a course in which she was a tutor. It was standard practice in this undergraduate course, as in many Humanities undergraduate courses, for tutors to mark their own students’ essays, that is, those in their tutorial group. However, in breach of this tradition, a decision was taken that the final essays would be exchanged among the tutors, marked ‘blind’ by another tutor on the course. Fatima was extremely upset by this decision. She explained that it had been justified on the basis that exchange of essays would result in more objective marks. She challenged this assumption, however, arguing that this practice was potentially very “unfair” to students since their essays were being marked by someone who did not have a “history” with them. She had just that morning had a meeting with one of her students, one of her “really hard-working students”, who had been given a failing mark on her essay. The student was “very angry” and wanted Fatima to justify the mark. From Fatima’s perspective the convener’s decision had created a “huge problem” (field notes, 15/10/1997).

It is not my intention to judge whether the course convener’s decision to exchange essays was appropriate, or whether Fatima’s anger was justified. I offer the story rather as an opportunity to tease out in a very preliminary fashion some of the assumptions which underlie the convener’s decision and Fatima’s response to it, and to speculate about the common sense notions of knowledge which are brought to bear in the

assessment of student performance. One way of typifying these notions of knowledge often heard in assessment discussions is to label them either 'objective' or 'subjective', the former referring to the perspective of one who is distant, outside, and uninvolved, and the latter to one who is close, inside, partial or biased. One of the issues at stake in the cited scenario is the form of knowledge which can legitimately be brought to bear in the assessment of student performance. The implication of the course convener's decision is that a tutor's involvement with the student may result in marks which are not the truest reflection of the student's performance. Implied in Fatima's, and perhaps the student's, angry response is that it is the tutor who has a history with the student who can best assess the performance.

The intrigue sparked by this story was heightened by my experiences as an educational developer working with staff who were responsible for the assessment of student performance. It was common practice to organize a marking workshop where samples of student performance were discussed in order for staff to come to agreement on the marking criteria, as well as the actual marks. Staff members were encouraged to make their marking criteria explicit by designing marking memoranda which were applied to the samples of writing and debated in the context of the marking workshop. As workshop participants our assumption was that such criteria would increase consistency between assessors' marks. We found, however, that consistency was the exception rather than the norm, and consensus among markers not always easy to forge. I was interested in this phenomenon, in particular the dissensus and what explanations lie beneath it.

These instances of assessment practice illustrate the problem this study explores. Tangled up in this problem are issues about: assessors as interpreters of student performance, the kinds of evidence that can be drawn on to support interpretations of student performance, the status of these different kinds of evidence, how different kinds of evidence can lead to different interpretations, how differences of interpretation are arbitrated and how power is effected in these arbitration processes. In a nutshell, the problem which this study explores is the validation of the assessment of complex

performances, that is, the on-going process of evaluating the soundness of assessment interpretations. As assessment acquires an increasingly dominant profile within the societal landscape, this process is a matter of considerable consequence, not only for individual learners and educators but for institutions and the wider society.

In part one of this chapter I briefly sketch the shape of this growing profile, the increasing use of assessment as a technology for reform and control. In part two, on the basis of a small but growing area of research, I call attention to three features of assessment as social practice – difference, context and power – which have not been given sufficient attention in dominant approaches to assessment rooted in psychometrics¹. I conclude by arguing for the need to develop theoretical and methodological approaches which illumine these features as crucial to an understanding of assessment as a complex social practice.

2.2 Educational assessment: multiple agendas

Educational assessment plays an increasingly dominant role in contemporary society. Broadfoot (2000) notes that from its modest beginnings in the universities of the eighteenth century to the school systems of the nineteenth and twentieth century, education assessment has become “the unquestioned arbitrator of value” with respect to student achievement, institutional quality and national educational competitiveness (p. ix). The growth of interest in assessment is fuelled by two educational and profoundly political agendas: educational reform and quality assurance. Although these agendas are closely linked, the impetus of reform is inward, that is, the improvement or transformation of teaching and learning. In quality assurance the impetus is outward, that is, accounting for the quality of teaching and learning to those outside the university.

Assessment is increasingly understood as an essential technology for the achievement of both agendas. Assessment is perceived as central to educational reform, firstly as an instrument for documenting the *need* for change, and secondly as a *catalyst* for change (Linn 1993). During the 1990s a great deal of research provided powerful

¹ Psychometrics is the science of psychological and educational measurement.

evidence for the influence of assessment on what is taught and how it is taught, and what and how students learn. As Resnick and Resnick (1992) put it, “you get what you assess” (p. 59). On the basis of this founding premise, assessment literature has been dominated by advocacy for assessment’s capacity to transform education. For example, Resnick and Resnick argue for assessment as an “essential tool” in educational reform (p. 72). Brown and Knight (1994) argue that given that assessment “shapes” the curriculum and “embodies” the purposes of higher education, then “assessment reform is the most urgent priority confronting undergraduate education” (pp. 11-12). Biggs (1996 citing Elton and Laurillard) argues, “the quickest way to change student learning is to change the assessment system” (p. 5).

With reference to the quality assurance agenda, assessment has become one of the primary instruments in the service of a growing preoccupation with ‘performativity’ (Broadfoot 1999). Assessment is perceived to be a powerful way of ‘accounting’ for the quality of teaching, and ultimately the quality of the graduate, to a variety of ‘stakeholders’, both internal – the students, staff, managers – and the external – tax-payers, employers, and the state. Thus assessment is perceived not only as an instrument of reform, but as one of the primary technologies of reporting on that change (Broadfoot 1996, 1999). This has raised a great deal of optimism among both educators and policy-makers. In the United Kingdom Brown and Knight (1994) argue that assessment is “central” to meeting the pressure upon higher education institutions to demonstrate “exactly” what the state “gets” for its “enormous investment” (p. 12). The 1990s saw the rise of numerous national qualification frameworks (NQFs) in the United Kingdom, Australia, and South Africa which assume that assessment (specifically learning-outcomes based assessment) can provide a systematic means of “gathering evidence and making judgments about an individual’s performance in relation to registered national standards and qualifications” (South African Qualifications Authority, 2001, p. 16).

While there are those who see educational assessment as an imperfect but powerful tool for change, there are others who have grave concerns about the extent to which assessment can fulfil these multiple agendas (Broadfoot 1999, Davis 1998). Foundational

to these concerns is doubt as to whether and to what extent assessment is 'science'; whether, as natural science advocates (Flyvbjerg 2001), accurate and reliable measurements can be taken, which form the basis of context-independent generalizations; whether these generalizations can in turn produce explanatory theories of human behavior; whether on the basis of such theories (e.g. of educational achievement) individuals and institutions can be measured and categorized for a variety of political, social and economic purposes. Some of these voices of concern would argue that what is needed is a "new science of assessment" (Berlack 1992, p. 1), or a "paradigm shift" (Gipps 1995, p. 4), implying that the existing scientific paradigm (which they argue is rooted in psychometrics) is flawed, riddled with "conceptual mythologies" (Davis 1998, p.1) and thus inadequate for dealing with contemporary educational and political problems. Others such as Moss (1996) are less intent on a paradigm overthrow but would argue that the foundations of educational assessment need to be extended, the dialogue needs to be enlarged to include new perspectives on assessment which challenge the dominant discourse. What notion of science will most usefully inform our assessment practices is the central theme of the next chapter.

2.3 Assessment as social practice

I now turn to the work of authors who are approaching educational assessment from different theoretical perspectives. One such area of work is what has been referred to as socio-cultural studies of assessment (Filer 2000). This encompasses the work of authors such as Broadfoot, Gipps, Filer and Berlack. Their work stresses that assessment is an activity or practice which can only be understood within the social, cultural, economic and political contexts in which it occurs (Gipps 1999). The other area of work is that which approaches assessment from an interpretivist or hermeneutical approach. In addition to Moss' seminal work, I include the work of composition theorists, such as Broad and Huot, who explicitly adopt this interpretivist label, and others, such as Barritt et al. Hatch, Phelps, and Schwegler, whose work is broadly sympathetic to this approach but who do not label it as such. The interpretive approach focuses on how both the assessed and the assessors' experiences, beliefs, and expectations constitute meaning in the assessment event (Gipps 1999). What is common to both these areas of work is,

firstly, a problematization of some of the central assumptions which currently undergird assessment, assumptions which emerge from assessment's foundation in psychometrics, and secondly, proposals for alternative perspectives which privilege assessment as a *social practice*.

Berlack's work (1992, 2000) critiques the use of large-scale, mandated, standardized tests² in the United States as one of the chief instruments of school reform. Berlack (2000) argues that despite decades of criticism from both public and professional sources, this movement is at the height of its power. One of his chief concerns is how the use of these tests privileges the authority of those outside the school and classroom over that of teachers and principals and learners. This power is legitimated, in part, on the basis of tests as instruments of science authorizing those who govern to create "standards" to measure the "commodity" of "educational productivity" and make "objective" and "unbiased" statements about how little or much individuals, schools, states have in comparison with others (p. 190). This standardization means that, for example, a given score on a reading test achieved by a student in a school serving a solidly white middle and upper-middle class community is more or less equivalent in terms of what it aims to measure as the same score achieved by a student in an urban, lower-middle class area. The premise of standardization is that scores can somehow stand outside of history, outside of everyday social conditions. Berlack argues that this is untenable. "Test scores cannot transcend context...they are only meaningful in the context in which they were generated" (p. 193). In contrast to the psychometric paradigm, Berlack (1992) advocates what he calls the contextual paradigm of educational assessment. Counter to the view that there can be a single consensual meaning about what standardized tests claim to measure, a contextual paradigm takes plurality of perspectives and differences in values and beliefs as givens, and treats these differences as assets. The challenge however, is how to mediate these differences, how to search for common ground within these differences, how to give voice to all the participants in this dialogue.

² The term 'standardized' or 'norm-referenced' tests refer to tests which are constructed in a way that allows a standard score, grade equivalency, or percentile to be computed, which permits comparison of an individual's score, percentile or group mean to those of another individual or group (Berlack 1992).

Broadfoot's work (1996, 1999, Broadfoot and Pollard 2000) explores the social role of assessment in historical and contemporary society, how changes in the socio-economic context are reflected in changing assessment practices. In her more recent work (Broadfoot 1999, Broadfoot and Pollard 2000) she traces how the concerns of the British government (e.g. rising unemployment, declining budgets, lowering standards, rising international competition) have turned for resolution to the creation of more "accountable" education systems, ushering in a "new hegemony of performativity" (1999, p. 4). As noted earlier, large-scale, mandated, standardized tests are perceived to be the prime technology for this accountability regime. Broadfoot (1999) argues that underlying more than a decade of British educational policies is the notion of assessment as a measurement device, the results of which are intended to goad students, teachers, institutions, and the nation to try harder or face the threat of failure, resource cuts, and closure. Underlying this near "pathological belief" (p. 3) in assessment are numerous assumptions, including the belief that standards of quality can be objectively measured and that assessment is a neutral measuring instrument. Like Berlack, Broadfoot contests these notions of scientific objectivity. She argues, "At all stages the assessment process is a social one; no matter how technically sophisticated the techniques employed, it is human beings who design the test, human beings who take it and human beings who use the results" (p. 13).

In its aim of envisioning assessment as social practice, the work of authors mentioned above introduces a new assessment discourse, a sociological discourse (Filer 2000). Crucial features of this discourse are concerns for the *context* of assessment, the inevitability of *difference* in assessment interpretations, as well as the equally inevitable effects of *power*. It is not that these features themselves are new to assessment discourse (and I will return in the next chapter to a more in-depth discussion of these features from the perspective of psychometrics), but rather that from a social-cultural perspective they are profiled as significant; they are privileged as central to the task of validation. In other words, the task of justifying the soundness of assessment interpretations cannot be undertaken without considerations of context, difference and power.

In order to illustrate this privileging – its discourse, the kinds of questions it poses, the tensions which are thrown up, the compromises it chooses to live with – I turn to a long-standing debate in the United States between the two communities of assessment scholars and practitioners. On the one hand, the standardized testing community, comprising professional test-developers and researchers, is responsible for providing the educational system with reliable, informative and efficient data to inform, for example, decisions about who should be admitted to university and into what level of writing programme they should be placed. On the other hand, the composition community, comprising administrators and teachers, is responsible to provide sound writing courses, at the appropriate level which adequately serve the needs of the students in their future undergraduate career. Within this debate I focus in particular on the phenomenon of multiple interpretations of the same performance and the implications for validity, a matter which has greatly preoccupied the composition community (Yancey and Huot 1996). This debate also serves to locate the problem which my study explores closer to the site of interest, that is, higher education.

Up until the late 1970s the American educational testing profession had been using standardized, large-scale, indirect tests for the assessment of writing for purposes of selection and placement into the tertiary education system. ‘Indirect’ measures of writing refers to the notion that students’ writing ability is assessed not through actual writing but through, for example, multiple choice, short-answer, true-false type questions. Given the nature of these items, these tests produced highly reliable data, but as critics noted, at the cost of validity. In other words, *consistency* of scores was privileged over concerns about what the scores actually meant in relation to the constructs³ being assessed.

Not surprisingly, the American composition community was among the harshest critiques of these indirect tests. Consistency among markers is a necessary but not sufficient condition for validity, they argued. The only way to assess students’ writing is by assessing actual writing. The challenge however was to score ‘direct’ tests of writing

³ “Construct” in testing discourse refers to the psychological trait that the test is presumed to be measuring.

with the same degree of reliability⁴ achieved by indirect measures. Achieving consistency across markers in the case of complex tasks (such as essays) is difficult as illustrated by Diederich et al's classic study (1961, cited in Huot 1990a). Diederich et al analyzed 300 papers on two topics by first-year students across four American colleges. The 300 papers were scored on a 9-point scale by 53 readers. Ninety-four percent of the papers received at least 7 different scores, and no paper received less than 5 separate scores. Interrater reliability was .31, far below acceptable levels.⁵

The outcome of this dilemma for the composition community has been an obsession with the elimination of differences, with devising the technology of reliability, for example, the design of rubrics, multiple markings, training sessions, and a variety of statistical procedures to calculate interrater reliability (Broad 2000, Huot 1996). To a degree this technology has been successful, by psychometrics' terms. Huot (1996) argues that only at the point at which reliability could be guaranteed at statistically acceptable levels did the direct assessment of writing become psychometrically viable. Direct tests of writing are now standard practice for large-scale, standardized testing, including the growing use of portfolio assessment.

Far from the debate being resolved however, voices from within the composition community have critiqued this obsession with reliability (Huot 1996). They argue that it is a mistake to assume that a well-oiled reliability machine will necessarily satisfy the conditions for validity. The argument is for a different approach to the problem of difference, an approach which recognizes differences of interpretation as an inevitable part of the interpretive/hermeneutical task of assessing. Studies from this interpretivist perspective have sought rather to understand what these differences are about and how

⁴ "Reliability" refers to the consistency of assessment results (Nitko 2001) which in psychometrics is measured as a correlation coefficient with a range of 0 to 1 where 0 is a completely unreliable set of results and 1 is a perfectly reliable set of results. Interrater reliability refers to consistency of results between markers.

⁵ There have been two recent studies conducted in South Africa exploring variation across markers of the same task in the context of higher education. Miller (1996) argues that the results of his study challenge the assumption held by academics that there is broad agreement between them about what constitutes different levels of performance. Sutherland (1999) documents the challenges of achieving consensus where staff are involved collectively in moderation.

these differences can enhance our understanding of what assessment is (Broad 2000, Yancey and Huot 1996).

Huot (1990a) reviews a number of studies arising out of the field of composition studies in the United States which explore the influences on “raters’ judgment” of writing quality (p. 237). He concludes that the common theme which emerges from these studies is what he calls “rater expectation” (p. 255); what a marker *expects* to find in the student writing, expectations shaped by both personal and professional experience, has a powerful influence on the scores allocated. Phelps (1989), in a very insightful study, identifies four different “attitudes” about the text that readers bring to the assessment of student writing. These attitudes range from reading the text as “closed” (p. 49), where the reader treats the text as self-contained, to reading the text as “evolving” (p. 51) where the text acquires historicity and is only a sample excerpted from a stream of previous and subsequent texts. Phelps argues that these different attitudes powerfully influence readers’ evaluations. Other studies show that in spite of pre-established criteria, readers of student essays will bring their own knowledge, values and interpretive strategies to the assessment task (Barritt et al 1986, Schwegler 1991). Hatch et al (1993) argue that readers of student texts infer different writer personalities from texts and that these impressions have consequences on assessment decisions.

Broad (2000) explores differences through an in-depth qualitative study of one American university’s writing programme, a programme which he describes as experiencing a “crisis of standardization” (p. 213). Broad describes the crisis in this evaluative community as the outcome of two competing sets of values held by the staff on the programme: on one hand, the values of consistency and fairness, on the other, the values of diversity, complexity and context dependence. Compounding the crisis was the fact that the community lacked a language and a theoretical framework that would allow for the protection of both ideals. Broad suggests that this community’s commitment to holding these two competing sets of values in tension was what enabled the emergence of a new paradigm for their approach to assessment, a hermeneutical one. The lack of scoring rubrics, limited central monitoring and context bound knowledge that were

perceived as problems from a psychometric point of view became strengths from a hermeneutical point of view.

I include two other studies in this review, neither of which comfortably fits in the socio-cultural studies or the interpretivist approach. However, they explore one of the central concerns of my study which is how assessment discourse constitutes and is constituted by relations of power.

Kvale (1993) explores assessment as a site of contesting value systems through a case study of an examination committee in a Humanities faculty at a Scandinavian university. He notes how, in spite of sustained criticism from both inside and outside the university, the examination system remains deeply-entrenched. He suggests that critiques of examinations as invalid might be based on an invalid understanding of what is actually being assessed through the examinations. He argues that examinations have a dual function. They play, what he refers to as, an evaluative role; examinations evaluate students in relation to some agreed frame of reference. However Kvale argues that the more significant role of examinations is, what he refers to as, their meta-evaluative role. Their function in evaluating the disciplinary frame of reference itself. He argues that through their evaluative decisions examiners operationally define what is valid knowledge in a discipline (p. 220). In this meta-evaluative function assessment becomes the battle ground for disciplinary border disputes and students the "sacrificial lambs" (p. 229).

Starfield (2001) also explores difference and the effects of power in the negotiation of these differences within disciplinary communities. Specifically in the context of a Sociology marking workshop at a South African university. An analysis of marking workshop discourse offers a "small window" onto "the complex contexts in which student(s) texts are being produced and received" (p. 145). She contests the portrayal of disciplinary communities as homogenous and unconflictual, immune from the power plays of the wider social contexts in which they reside. Her analysis of the markers' discourse reveals "socially-situated individuals who constitute the discourse community,

with their varying amounts of academic capital (who) assess students' texts, not necessarily according to a pre-agreed upon formula" (ibid.). Her study exposes how "...in the process of talk, meaning is shaped and re-shaped", and how the outcome for students is powerfully constituted by "constantly shifting power relations" within the interpretive community (ibid.).

2.4 Conclusion

I now conclude by showing how these socio-cultural and interpretivist approaches have enlarged the assessment dialogue (Moss 1996); in other words, how the privileging of context, difference and power contributes to a deeper appreciation of the complexity of assessment as an interpretive process.

One of the issues at stake in this debate is how to make sense of difference. From a psychometric perspective, differences between assessors' interpretations of the same task are a reliability problem; low reliability or inconsistency is due to error and undermines validity. "Without reliability, there can be no validity" (Cronbach cited in Moss 1992, p. 6). In contrast, what socio-cultural and interpretivists studies of assessment are demonstrating, though in different ways, is that firstly, differences of interpretation are inevitable. This chapter began with a quote from Nietzsche who argues, "there is *only* a perspective seeing, only a perspective 'knowing'" (cited in Bourdieu 1990, p.28) (italics in original). Secondly, a point which composition theorists have vigorously argued is that eliminating difference is costly. It is a violation of what the interpretive process is all about (Barritt et al 1986). Thirdly, difference is a resource, an asset to assessment practice. Once again to quote from Nietzsche, "The *more* eyes, different eyes, we use to observe one thing, the more complete will our "concept" of this thing, our "objectivity", be" (cited in Bourdieu 1990, p. 28) (italics in the original).

Returning to Fatima's dilemma, a socio-cultural or interpretivist approach suggests that the differences that emerge between her and the other tutor in how they assess the student's essay, as well as the differences which emerge between marking workshop participants, are significant. As Broad (2000) discovered in exploring the crisis in his evaluative community, such differences may be signposts for underlying value

contestations. Broadfoot (1996) notes that assessment practices reflect the often conflicting values embodied in education systems. These conflicting values may, however, only become visible through crisis or paradigmatic change (Kvale 1993). One of the contributions of interpretivist and socio-cultural approaches to assessment is to show how, rather than being a threat to validity, differences of interpretation can be an asset. They provide an opportunity not only for an account of what we value but also, as importantly, for an account of our preconceptions and prejudices, those which are destructive as well as those which are productive of knowledge (Bernstein 1983, Broad 2000).

The second issue is that of context. One of Berlack's critiques of the use of standardized test scores in the United States (and similar critiques are made by Broadfoot and Gipps about their use in Britain) is the underlying assumption that these scores can have meaning outside of the contexts which produced them. Socio-cultural and interpretivist studies argue that assessment cannot rise above context. From the macro socio-political and economic contexts which Broadfoot (1996) explores to the micro contexts of the classroom (Filer 2000) context is built into the very fabric of interpretations. As Berlack (2000) argues, "context matters" (p. 193). Arguably one of the most important contextual features is the assessor him/herself. In a review of the field of educational assessment, Gipps (1995) argues, "Our mistake, with hindsight, has been to believe that learning can be assessed accurately and reliably ... Not only do we know that performance is context bound, affected by motivation and the assessment task, but more importantly (because it is hidden) performance is construed according to the perspectives and values of the assessor..." (p.12).

In describing her dilemma, Fatima astutely notes that a tutor's history of a student is valuable contextual knowledge which she alone can bring to the interpretation of her student's performance. But she fails to recognize that other assessors are not context-less. She is correct to challenge the status accorded to the 'objective' perspective at the cost of her own more 'subjective' view. There is no denying, however, that each tutor will bring different kinds of evidence, different kinds of knowledge to bear on the

process given the different contexts which they inhabit. The question is, how will these differences be resolved? Whose contexts will be privileged, and on what basis?

This leads to the final issue of power. Some of the harshest critiques against the standardized testing enterprise focus on the manner in which power is shifted to the top, away from stakeholders at the local level (e.g. teachers, school administrators) towards the state, the nation (Berlack 1992, 2000, Broadfoot 1999). And yet, the psychometric field has been, until recently, generally silent about issues of power. While there is an increasing acknowledgement of the consequences of assessment interpretations, I contend that psychometrics is not equipped theoretically or methodologically to evaluate the effects of power on these assessment interpretations (an argument to which I return in the next chapter). Kvale's (1993) study speculates about how disciplinary power struggles find effect in assessment. Starfield's (2001) analysis points to the effects of race, gender and academic status in assessment discourse. Though these studies and others have contributed to a greater realization of 'power effects' in assessment, there is still much that we have to learn about how power constitutes the interpretive process, in particular, the negotiation and resolution of difference.

2.5 A problem in search of theory

In giving shape to the problem which this study explores, I started off by briefly sketching how educational assessment finds itself at the nexus of complex and highly contested expectations for social and political change. In response, socio-cultural and interpretivist studies of assessment practice call for a critical scepticism of assessment as a technology for reform and control, and there is an increasing acknowledgment by researchers and practitioners that the assessment of student learning is "enormously complex, politically and conceptually" (Berlack 1992, p. 201). At the same time there is a recognition that the theory and methodology for exploring this complexity is undeveloped. Broadfoot (1999) argues that there is an urgent need for research from a more socio-cultural perspective which explores assessment as both social process and social product. At the end of her article "Enlarging the Dialogue in Educational Measurement", Moss (1996) calls for research within interpretive traditions to inform

assessment theory and practice by “providing theory, illustrations, and methodologies that illumine a complex social reality, thus enabling us to better understand the ways that assessment works within the local context” (p. 27).

I would add that what is needed in particular are studies of assessment practices which illumine difference, context and power in the interpretations of student performance. In the next two chapters I turn to the task of constructing a theoretical and methodological framework which will not only accommodate but be hospitable to these crucial features.

Chapter 3: Theory

Of all the oppositions that artificially divide social science, the most fundamental, and the most ruinous, is the one that is set up between subjectivism and objectivism. The very fact that this division constantly reappears in virtually the same form would suffice to indicate that the modes of knowledge which it distinguishes are equally indispensable to a science of the social world. (Bourdieu 1990, p. 25)

3.1 Introduction

The central problem which this study seeks to explore is how academic assessors validate their interpretations of student performance on complex tasks. Specifically I am interested in the phenomenon of multiple interpretations of the same performance. In the last chapter I explored in a preliminary way some aspects of this interpretive process from the perspective of different traditions. Implied in these alternative approaches is an intellectual tussle between different ways of approaching assessment with psychometrics on the one side and more socio-cultural and interpretivist approaches on the other. I pointed to some of the issues which might be at stake in this tussle: differences in the goals of inquiry, modes of knowledge, forms of evidence, and emphases in discourse. Some of the authors, in fact, seem to be suggesting that what are at stake are different understandings of science.

In this chapter I explore this intellectual tussle or contestation in more depth. I start (3.2) by briefly characterizing what Bernstein (1983) argues is the primary cultural and intellectual conflict of contemporary life – the opposition between objectivism and relativism. This opposition provides a basis from which to better understand the tussle between different approaches to assessment discussed in the previous chapter, and serves as well to foreshadow some of the themes which emerge from the case study analysis. I then turn to explore (3.3) two different approaches to validation, two ways of evaluating the soundness of assessment-based interpretations: psychometrics and interpretivism. With respect to these approaches I am interested in the following questions: From the perspective of these different traditions of scientific inquiry, what does validation mean? What are its goals of inquiry? What understanding of rationality does it assume? I offer a brief overview of these approaches.

In 3.3.1 I turn to validation within the psychometric tradition. Drawing principally on the seminal work of Samuel Messick and Lee Cronbach, I highlight some of the significant developments in the evolution of the concept of validity from its objectivist roots to its grounding in what Messick (1989) refers to as constructive-realism. My central argument is that although validity inquiry within that tradition has moved “beyond”¹ objectivism, it has not moved far enough beyond. It remains a social science aspiring to the ideals of the natural sciences where the goal of interpretation is the construction of generalizations which contribute to theory. While the study of regularities is an important and necessary task, there is a tendency in psychometrics (and Bernstein argues in all mainstream social science) towards reification, that is, taking its reality as given rather than investigating it as a historical and social artefact (Bernstein 1976, Moss et al 2002). Thus while the contributions of psychometrics to validity inquiry have been significant, an enriched understanding of assessment as an interpretive process will need to come from theoretical perspectives beyond those which have informed psychometrics.

I turn in 3.3.2 to validity inquiry from an interpretivist perspective. Moss (1996) challenges the psychometric community to enlarge the dialogue by exploring validity from the perspective of the interpretive tradition of science. In contrast to mainstream social science, out of which psychometrics has developed, interpretive science argues that there are intrinsic differences between the study of natural life and social life. In contrast to the objects of natural science (i.e. physical objects), the ‘objects’ of social life are reflexive: humans beings are constantly reflecting on what they do as part of what they do. Secondly, their meanings are intersubjective; they are not the property of individuals but the common property of a particular group. Thirdly, their meanings are discursive; they are constituted in and by language. I draw on Moss’ (1998b, Moss et al 1998) two approaches to assessment, the aggregative and the integrative, to illustrate the implications of these different epistemological perspectives for the evaluation of interpretations. Ultimately Moss (1996) is not arguing for one perspective over the other, the psychometric or the interpretive, but for a dialectic perspective which “can provide a

¹ “Beyond Objectivism and Relativism” is the title of Bernstein’s (1983) book.

more textured and productive view of the social phenomenon we seek to understand” (p. 22).

In 3.4 I argue that such a dialectic perspective can be found in Bourdieu’s theory of social practice, and I construct a social theoretical perspective which I believe is best equipped to illuminate the way in which context, difference and power are constitutive of assessment as an interpretive act². Firstly, it gives insight into assessment as a socially situated practice, meaning that validation practices only make sense within the logic of the ‘field’ in which they are located. Secondly, members of a community of practice, to the extent that they have been subjected to the same social conditions imposed by the ‘field’, share a ‘class habitus’, or a ‘professional vision’ (Goodwin 1994) which informs their classificatory acts. Thirdly, a social theoretical approach to assessment illumines the complexity of contingencies which influence assessment-based interpretations, contingencies which may result in an adjustment and even a divergence from tacitly agreed-upon points of reference. And finally a social theoretical perspective illuminates how language plays a crucial role in all these meaning-making, position-taking, power-effecting acts. I argue that all of these insights are important to a richer understanding of assessment as a social practice.

3.2 The search for foundations

Bernstein (1983) argues that there is an uneasiness which has spread through intellectual and cultural life which affects not only every discipline but every aspect of our lives. This malaise raises not only intellectual concerns, but the issue of some of the most perplexing questions concerning human beings: what we are, what we can know, what norms ought to bind us, and what are the grounds of our hope. This uneasiness is expressed by the opposition between objectivism and relativism, although there are other oppositions which point to the same underlying anxiety: rationality versus irrationality, objectivity versus subjectivity, realism versus antirealism. Bernstein’s thesis is that these oppositions are not the only viable alternatives, and that there is now an urgent call to

² I am greatly indebted to Pamela Moss who, from within the psychometric community, initiated and continues throughout her work to sustain a dialogue between psychometric, hermeneutical and social theoretical approaches to validity inquiry.

move beyond them. One of the central projects in this move is a re-examination of the very nature of human rationality. Although Bernstein's work was published 20 years ago, I believe that his arguments are still central to on-going debates in the philosophy of the social sciences, debates about what social science is and how it contributes to social life (see Bohman 1991, Flyvbjerg 2001). I now turn to a clarification of what Bernstein means by this opposition.

By 'objectivism' Bernstein (1983) means "the basic conviction that there is or must be some permanent, a-historical matrix or framework to which we can ultimately appeal in determining the nature of rationality, knowledge, truth, reality, goodness or rightness" (p. 8). Objectivists claim that there is or must be such a matrix and that the primary task of the philosopher is to discover what it is. Thus objectivists are in search of foundations, for some "fixed, permanent constraints" to which they can appeal, "which are secure and stable" (p. 19). In contrast, relativists not only deny the existence of such foundations, but argue that whatever philosophers take to be most fundamental – whether it is the concept of rationality, truth, reality, right, the good, or norms – must be understood as relative to a specific theoretical framework, paradigm, society, or culture. "For the relativist", Bernstein argues, "there is no substantive overarching framework or single metalanguage by which we can rationally adjudicate or univocally evaluate competing claims of alternative paradigms" (p. 8).

Therefore with reference to the issues of criteria or standards, objectivists make it their life quest to find *the* standards of rationality, standards which are universal, independent of history and society. On the other hand, relativists deny the existence of standards of rationality which are universal, that are not subject to historical or temporal change. We cannot escape, they argue, from the predicament of speaking of "our" standards and "their" standards of rationality, standards that may be "radically incommensurable" (Winch cited in Bernstein 1983, p. 27).

How do we account for the tenacious hold of this dichotomy on intellectual thought? Bernstein's explanation lies in what he refers to as the Cartesian anxiety. He traces the

roots of this opposition back to Descartes who, Bernstein argues, bequeathed to modern philosophy a set of problems, questions and metaphors which have been at the centre of philosophy ever since. This is the quest for some “fixed point, some stable rock upon which we can secure our lives against the vicissitudes that constantly threaten us” (Bernstein 1983, p. 18). The alternative to this “fixed point” is not just “radical epistemological scepticism but the dread of madness and chaos where nothing is fixed, where we can neither touch bottom nor support ourselves on the surface” (ibid.). The crux of the anxiety, Bernstein argues, is the manner in which Descartes leads us to the “grand and seductive either/or” (ibid.). “*Either* there is some support for our being, a fixed foundation for our knowledge, *or* we cannot escape the forces of darkness that envelop us with madness, with intellectual and moral chaos” (ibid.) (italics in original). It is this anxiety, Bernstein argues, which is the primary cultural conflict of contemporary life, contributing to a growing intellectual crisis.

Thus Bernstein’s task is to exorcize the Cartesian anxiety, and to move beyond these “misleading and distortive” (p. 19) dichotomies. He shows how postempiricist philosophies of science, notably the work of Thomas Kuhn, have called into question this whole framework of oppositional thinking, arguing instead for another alternative, a new image of science, a new concept of rationality.

I now turn to a discussion of validity inquiry from two different theoretical perspectives, each of which I argue is an attempt to move beyond objectivism.

3.3 Beyond objectivism and relativism

3.3.1 Validity inquiry from a psychometric perspective

As the cornerstone of psychological and educational measurement, validity has been the subject of sustained theorization since the mid-1950s when Cronbach and Mehl formulated the initial concept of construct validation (Cronbach and Mehl cited in Cronbach 1989). It is not my intention to review this volume of work³. My goal is to highlight some of the significant developments in the evolution of validity inquiry from

³ For such overviews see Kane 2001, Moss 1992, Shepard 1993.

the psychometric perspective. I draw principally on the work of Samuel Messick and Lee Cronbach, both leading contributors to the development of validity theory within psychometrics, whose work in reformulating validity emerged at a time when “storybook images of science” (Mitroff cited in Mishler 1990, p. 417) were being called into question. I close by arguing that, while these conceptual shifts have been significant, there are features of assessment as an interpretive process which the psychometric approach is not equipped to illuminate.

Berstein (1983) argues that one of the significant contributions of postempiricist philosophy and history of science is “the recovery of the hermeneutical dimension of science in both the natural and the social science” (p. 171). Consistent with this recovery, the concept of validity has shifted from being understood as a technical procedure to an understanding of validity as an interpretive process (Angoff 1988, Cronbach 1989, Messick 1989). Validity shifts from being a suite of types of validity (content validity, criterion validity, construct validity⁴) to a unified concept of validity, with content validity and criterion validity subsumed under construct validity⁵. All validation is construct validation (Cronbach cited in Messick 1989, p. 19), that is, all validation is about substantiating the meaning of scores in relation to the constructs the assessment purports to assess. Thus validity inquiry shifts from being largely a technical, quantitative operation (usually measured through correlations) to an on-going process of researching the strength of relationship between the test scores and the theory underlying the construct, a process which requires many lines of evidence, not all of them quantitative (Angoff 1988).

⁴ Content validity refers to the measure of alignment between the content which the test is assessing and the broader domain of content which it purports to represent. This ‘measure’ of validity requires some form of professional judgment. Criterion validity refers to the extent to which the test scores correlate with other independent measures of what the test is designed to assess and predict. This measure of validity usually takes the form of a correlation. In psychometrics, ‘construct’ refers to the psychological trait that the test is presumed to be measuring. Construct validity thus refers to the strength of the relationship between the interpretations of the test scores and the ‘construct’ it purports to be testing.

⁵ This unified concept of construct validity is only officially accepted in the 1985 version of the Standards for Educational and Psychological Testing (American Psychological Association, American Educational Research Association and National Council on Measurement in Education).

So while the term 'hermeneutical' itself is not used, validity (or increasingly the preferred term 'validation') is characterized as an interpretive process (see also Mishler 1990). "Validation ...is evaluation" (Cronbach 1988, p. 4). Messick (1989) defines validity as, "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (p. 13) (italics in original). He elaborates, "What is to be validated is not the test...but the inferences derived from the test scores...inferences about score meaning or interpretation and about the implications for action that the interpretation entails" (ibid.). It follows then that, "validity is a matter of degree, not all or none...validity is an evolving property and validation is a continuing process...because evidence is always incomplete, validation is essentially a matter of making the most reasonable case..." (ibid.).

Another significant development in the evolution of validity inquiry within psychometrics is the growing acknowledgment of the central role that values play in the interpretation of meaning, that there can be no permanent division between observation and theory (Bernstein 1983). In his seminal piece Messick (1989) grounds validity inquiry philosophically beyond objectivism and relativism in what he refers to as a constructivist-realist approach. It is realist, he argues, "because it assumes that traits and other causal entities exist outside the theorist's mind; it is *constructive*-realist because it assumes that these entities cannot be comprehended directly but must be viewed through constructions of that mind" (p. 29) (italics added). By acknowledging that constructs represent "our best, albeit imperfect and fallible" constructions of the mind (Messick 1998, p.35), Messick is highlighting the role that values and ideologies play in shaping those constructions. He argues, for example, that even construct labels are value-laden (e.g. the difference between the construct label 'stress' and the label 'challenge'). He cites theories of intelligence as an example of the ways in which theories and models are also value-laden. Thus in grounding validity inquiry philosophically, Messick's (1989) appeal to realism echoes the objectivist quest for a fixed point, something which "exists

independent of the theorist's mind" (p. 29)⁶; the appeal to constructivism is a recognition of the need to "embrace" some of relativism's basic points, namely, that "all observation and meaning is theory-laden and that all theory is value-laden" (p. 24).

Thus, validity inquiry, as Messick presents it, is a move beyond objectivism, founded on a "new realism"⁷ (p. 26). But how far beyond? Messick notes that "validation is scientific inquiry" (p. 14) but what kind of scientific inquiry? What are the goals of this inquiry? What scientific ideals drive this kind of validation work? What kinds of evidence are privileged? What are its greatest threats?

Moss (1996) argues that psychometric notions of validity have evolved out of mainstream or naturalist conceptions of social science⁸. While mainstream social scientists do not comprise a homogeneous group, in general they argue that the disciplines of social science differ in degree, not in kind from the well-established natural sciences (Bernstein 1976). Their success as scientists will be found in "emulating, modifying and adapting techniques that have proven successful in our scientific understanding of nature" (p. xv). The aim of the social sciences, from this naturalist perspective, is thus the same as that of natural sciences although even the most fervent believers are acutely aware of the challenges this poses for social science: collecting and refining data, discovering correlations, formulating testable empirical generalizations, hypotheses, and models, and most importantly contributing to the growth of testable and well-confirmed theories which explain phenomena and are derived from theoretical assumptions (Bernstein 1976).

⁶ From a 'constructivist-realist' stance Messick (1989, 1998) would argue that some constructs have real referents while others do not. "Just as on the realist side there may be traits operant in behaviour for which no construct had yet been formulated, on the constructive side there are useful constructs which have no counterpart in reality" (Messick 1998, p. 35).

⁷ Messick is here drawing on the work of Bhaskar (1979) and the philosophical tradition of critical realism.

⁸ Moss (1996) argues that the naturalist/interpretivist distinction serves a "useful purpose" in terms of highlighting some aspects of educational measurement, but she expresses caution about its over-emphasis. "Though the naturalist/interpretive cut foregrounds a particular set of issues – a particular set of commonalities within and differences across traditions – a different category scheme would rearrange allegiances among traditions" (p. 21).

Consistent with mainstream social science, Messick (1989) notes that the major goal of validity inquiry is the construction of generalizations:

“The major concern of validity, as of science more generally, is not to explain any single isolated event, behavior, or item response, because these almost certainly reflect a confounding of multiple determinants. Rather, the intent is to account for *consistency* in behaviors or item responses, which frequently reflects distinguishable determinants... We thus move from the level of discrete behaviors or isolated observations to the level of measurement. This is not to say that scores for individual items or discrete behaviors are not often of interest but, rather, that their meaning and dependability are fragile compared with response consistencies across items or replications” (p. 14) (italics in original).

One of the fundamental challenges to generalizability is the impact of context on score meaning. Messick (1989) notes, “Test responses are a function not only of the items, tasks, or stimulus conditions but of the *persons* responding and the *context* of measurement... a context which includes factors in the environmental background as well as the assessment setting (p. 14) (italics in the original). Thus a central validation concern is to establish whether the meaning of a measure is context-specific or whether it generalizes across contexts. Messick notes that although the impact of contextualization can be compelling, there are methods of “investigating or monitoring” the role of context through “multiple task formats” along with the “standardization of testing conditions” all of which will contribute to the goal of “convergent interpretations and comparability of scores across respondents and settings” (pp. 14-15).

Again consistent with a mainstream approach to social science, Messick (1989) argues that “although there are many ways of accumulating evidence to support a particular inference, these ways are essentially the methods of science” (p. 14). He

proposes a list of criteria, what he calls ‘aspects’ of validity or sources of evidence which need to be brought to bear on the validation of score meaning and uses⁹. They are:

- Content aspect of construct validity – This is evidence that the assessment tasks are relevant and representative with respect to a particular theory of the construct domain. This evidence usually takes the form of consensual professional judgment.
- Substantive aspect of construct validity – This is evidence that the assessment tasks are requiring the requisite thinking skills and processes specified by the theory of the construct domain. This evidence moves beyond professional judgment of content (as in the content aspect above) to evidence from actual responses that the skills being assessed are the ones specified by the domain.
- Structural aspect – This is evidence that the internal aspects of the assessment, for example, the structure of the test and the scoring rubric, are consistent with the theory of the construct domain.
- Generalizability aspect – This is evidence that the score interpretations are not limited to the sample of assessed tasks but generalizable across groups, settings, and tasks. This evidence is usually derived through factorial analyses.
- External aspect – This is evidence that the assessment interpretations are consistent with interpretations of other (independent) assessments of the same constructs. This evidence is usually derived through correlational analyses.
- Consequential aspect – This is evidence that there are no unanticipated negative side effects as a result of the assessment results’ interpretation and use.

Messick (1995) argues that “These six aspects of construct validity apply to all educational and psychological measurement, including performance assessments. Taken together, they provide a way of addressing the multiple and interrelated validity questions that need to be answered in justifying score interpretation and use” (p. 8). Again I suggest that this list of evidence types or criteria for validity, productive as they may be, echo the objectivist quest for explicit standards of rationality (Bernstein 1983).

⁹ I have drawn on Nitko’s (2001) adaption of Messick’s aspects of validity in order to translate Messick’s highly technical language into something more accessible to a non-measurement audience.

Thus for Messick at the heart of validity inquiry is a clearly specified theory of the construct domain. Validation is hypothesis testing, that is, evaluating interpretations against a theory of the construct domain. This requires the collection of particular kinds of evidence to support assessment-based interpretations or offer alternative interpretations. The kind of evidence privileged is that which supports or challenges the consistency of the interpretations. The move is from the particular to the general with a careful “monitoring” for the “intrusion of context” (Messick 1989, p. 15). This is essential if interpretations are to have any meaning outside the immediate contexts in which they are generated. Ultimately on the basis of these regularities, social scientists offer explanations which are empirically and theoretically sound. This, for Messick, is rationality. Interestingly, Cronbach (1989) appears somewhat less optimistic about the role of construct validation in relation to theory building. He argues that an “idealized strong (validation) program is most appropriate to a scientific perspective that reaches centuries into the future”¹⁰ (p. 163). He argues that “test interpretation almost never has a consolidated theory as its armature; mostly, we rely on crude theory-sketches. The loose assembly of concepts and implications used in typical test interpretations I shall call ‘a construction’ rather than ‘a theory’” (p. 152). Later, he adds, “Acceptance of constructions is inherently a community process” (p. 164), a process which he concedes is as much social as it is rational.

To summarize, my main goal has been to demonstrate the contribution of psychometric theory to validity inquiry, namely, its evolution from a largely pragmatic and empirical concept (Angoff 1988) to a concept more consistent with postempiricist notions of scientific inquiry. This is the recognition that validation, as are all forms of scientific inquiry, is an interpretive process in which the divide between fact and value is difficult to delineate. Nonetheless validity inquiry within this tradition remains committed to the aspirations of the natural sciences where the goal of interpretation is the construction of generalizations which support theory, although I noted that even within

¹⁰ This seems to reflect a mainstream view of social science as not yet ‘mature’ natural science.

psychometrics there is some scepticism about the role of construct validation in relation to theory development.

Bernstein (1976) argues that in the final analysis the issue is not the use of empirical and quantitative techniques to establish regularities and correlations. Social practices do exhibit regularities and correlations, and these must be understood. The question is, how do we understand their significance? He argues that the problem with mainstream social science is the tendency towards reification, “toward mistaking historically conditioned social and political patterns for an unchangeable brute reality which is simply ‘out there’ to be confronted” (p. 106). While this is a danger that both Messick and Cronbach are attuned to, Moss et al (2002) argue that psychometrics as a scientific discipline does run the risk, as do all scientific communities, of reifying its own perspectives; “It comes to see that part of reality which its methods are best adapted to study as the entire picture of the way things are” (p. 5). This reification may lead to blind spots. Some aspects of the interpretive process are illuminated while others remain backgrounded. The quest for the general potentially overlooks the richness to be found in the particular. The need for context-independent meanings may result in distortions of those very meanings. The privileging of consistency may obscure what underlies *inconsistency* in interpretations. Thus while there is no denying the significant contribution of psychometrics to the validity dialogue, a richer understanding of the interpretive process will need to come from theoretical perspectives other than those which have informed psychometrics.

3.3.2 Validity inquiry from an interpretive perspective

Moss’ main strategy for exposing some of psychometrics’ blind spots is to hold up an alternative perspective of science, that of the interpretive (or hermeneutical) tradition. With respect to the naturalist debate, the interpretive tradition rejects the notion that social life can be or should be studied in the same way as natural life. Their contribution to the debate has been to emphasize the intrinsic differences between the study of social life and the study of natural life, differences which, interpretivists would argue, make the scientific investigation of social life impossible (‘scientific’ as understood by the natural sciences). The interpretive tradition argues that the idea of a social science conceived of

as a natural science is based on conceptual confusion. But more importantly, Bernstein (1976) argues, interpretivists have shown that “the obsession with transforming social studies into natural science obscures, distorts, and suppresses the legitimacy of issues vital for theorizing about political and social life” (p. 59). This obsession distorts the ways in which social science can and *does* matter (Flyvbjerg 2001).

What are the differences between the study of ‘objects’ in social science and the ‘objects’ of the natural science? I turn to a discussion of three distinctive dimensions of social life and their implications for how we evaluate interpretations.

One of the distinctive dimensions of the study of social life is reflexivity, that is, that humans are self-interpreting beings; they are constantly reflecting on what they do as part of what they do. These self-reflections are of two types (what Giddens cited in Flyvbjerg 2001 refers to as the ‘double hermeneutic’). Firstly, there are the self-reflections of the people being researched. In contrast, the objects of natural science are not self-interpreting, in other words, they do not talk back (Flyvbjerg 2001). Secondly, there are the self-reflections of the researcher. These self-reflections, that is, beliefs that people have about themselves and others, are not, Bernstein (1976) argues, “simply subjective states in their minds; they are...*constitutive* of the actions, practices, and institutions that make up social and political life” (p. 61) (italics in original). In other words, the researched and the researchers’ self-reflections (the ‘subjective’) do not exist in a vacuum. They are in mutually constitutive relationship with the actions, practices and institutions (the ‘objective’). Context both determines and is determined by these self-understandings (Flyvbjerg 2001).

A significant implication of the reflexive dimension is that there is no disengaged observer. Bernstein (1976) argues that there are few mainstream social scientists who would deny the hold of subjective issues, such as beliefs, values and ideology, on interpretation. As noted above, given the developments in the philosophy of science, Messick (1989) accepts and even “embraces” the notion that all observation is value-laden (p. 24). But there are other implications which are more subtle and more profound.

The issue is the deeply entrenched perception of the subjective and the objective as dichotomies, a perception which Bernstein argues misrepresents human action. "Human action ... is internally related to the interpretations that are intrinsically constitutive of it" (1976, p. 62). The bottom line for social inquiry is, as Flyvbjerg (2001) puts it, "context counts" (p. 38). Thus this reflexive dimension of social life raises a scepticism about the quest for explanatory theories of social life based on context-independent generalizations. It calls to question an evaluative process which sees context as an intrusion or threat to validation. I return in the final section (3.4) to a more in-depth exploration of the constitutive relationship between the objective and the subjective, as this is one of the major themes of Bourdieu's theory of social practice.

A second distinctive dimension for the study of social life is what interpretivists refer to as 'intersubjectivity'. Intersubjectivity takes us to an even deeper appreciation of the importance of context for the evaluation of interpretations. As noted above, subjective and objective meanings are mutually constitutive of action. Taylor (1987), however, challenges the notion of 'subjectivity'. He argues that there are a "range of meanings open" to an interpretive community (p. 56), but these are not subjective meanings. In other words, "They are not the property of one or some individuals, but rather they are intersubjective meanings, which are constitutive of the social matrix in which individuals find themselves and act" (p. 57). Intersubjectivity consists of 'interpretive schemes' (Bernstein 1976) which are not private, but social; not subjective but intersubjective. They serve as a common point of reference for an interpretive community. However, Taylor (1987) argues, this common point of reference is not the same as consensus. Indeed it is because of intersubjective meaning that there can be and often is disagreement (or what Bernstein 1992 refers to as 'dissensus'). Intersubjectivity is only a condition for consensus, or put another way, consensus and dissensus are only possible where there is intersubjective meaning. This notion foreshadows Bourdieu's concept of 'habitus', and I return to it in the final section (3.4), and its implications for the evaluation of interpretations.

The third distinction between the study of social and natural life is that, unlike natural objects, social objects are constituted in language (Rabinow and Sullivan 1987). The significance of language in understanding social practice heralded the 'linguistic turn' in the social sciences, the recognition of the dialectic relationship between language and social life; language is not simply a reflection of social reality but constitutes and is constituted by it. Taylor (1987) notes how it is possible for the natural scientist to distinguish between, for example, the 'heavens' and discourse or theories about the 'heavens'. But it is not so simple for the social scientist to distinguish between, for example, 'voting practices' and the discourse used to describe these practices. The very choices of language by both voters and the researcher of voting practices constitutes the nature of the practice. Taylor (1987) concludes, "There is no simple one-way dependence here. We can speak of mutual dependence if we like, but really what this points up is the artificiality of the distinction between social reality and the language of description of that social reality. The language is constitutive of the reality, is essential to its being the kind of reality it is" (p. 54).

I return to this discursive dimension of social life in the next chapter. Consistent with critical discourse theory I argue that the analysis of discourse serves as one of the crucial points of access to social life. Suffice to say at this point that this argument has crucial implications for the evaluation of interpretations. Language is not simply a neutral container into which speakers pour their meaning. Critical discourse theory argues that the very choices of language are crucial to the construction of the interpretations themselves. Language is involved in "building tasks" (Gee 1999, p. 85): building belief systems, personal identities and interpersonal relations. While psychometrics gives central place to validation as an interpretive act, the role of language itself in constituting meaning-making acts is taken for granted.

The reflexive, intersubjective, and discursive dimensions of social life raise a number of implications for how we evaluate interpretations. Moss (1998b, Moss et al 1998) concretizes these differences by applying two different approaches to a practical

assessment problem (the assessment of portfolios for teacher certification¹¹). She refers to these approaches as the 'aggregative' and 'integrative'. While acknowledging the advances made by the psychometric community in the assessment and validation of complex performances, she critiques the conventional model for portfolio assessment. Typically the procedure consists of a panel of assessors who work independently, scoring one exercise at a time, blind to the candidate's performances on other exercises. In order to reach a decision, scores assigned by individual readers are then algorithmically combined and compared to a pre-determined cut-score. Moss (1998b) argues that while this is "sound and thoughtful" practice within the psychometric community, it is instructive to question the quality of this information and its consequences from another perspective (p. 140).

Moss (1998b, Moss et al 1998) uses the interpretive notion of the 'hermeneutical circle' to argue for an integrative and dialogic approach to assessment. Drawing on Gadamer (1987 cited in Moss 1998b, Moss et al 1998) she characterizes the circle as representing a dual dialectic: one between parts of the text and the whole, and one between the text and the reader's foreknowledge, preconceptions or "enabling" prejudices (Moss et al 1998, p. 141). Drawing on Bernstein (1983 cited in Moss, 1998b, Moss et al, 1998) she makes a crucial distinction between "blind" prejudices and "enabling" prejudices (Moss et al 1998, p. 142). Thus the interpretive task is not to remove preconceptions – this is not possible – but to "test them critically in the course of inquiry" (Bernstein cited in Moss et al 1998, p. 141). Moss (1994) argues, "A hermeneutic (or interpretive) approach to assessment would involve holistic, integrative interpretations of collected performances that seek to understand the whole in light of its parts, that privilege readers who are most knowledgeable about the context in which the assessment occurs, and that ground those interpretations not only in the textual and contextual evidence available, but also in a rational debate among the community of interpreters" (p. 7).

¹¹ The context of this work is the Interstate New Teacher Assessment and Support Consortium (INTASC) which is developing portfolio assessment to assist participating states in the US to make decisions about teacher licensure.

The quality of the interpretations which arise from this kind of assessment will be evaluated on the basis of a different set of criteria to those which the psychometric community privileges. (Recall Messick's 'aspects of validity' and the particular privilege afforded to reliability and generalizability discussed in 3.3.1). Some of the criteria which Moss (1994) mentions are: the assessor's extensive knowledge of the learning context; multiple and varied sources of evidence; an ethic of disciplined, collaborative inquiry that encourages challenge and revision to initial interpretations; and the transparency of the trail of evidence leading to the interpretations, which allows users to evaluate the conclusions for themselves. From an interpretivist perspective, the community of interpreters in critical dialogue is a crucial criterion for validation. If we accept critical dialogue as a criterion for validation, then Moss (1994) argues "the question is who participates in the dialogue and this is an issue of power" (p. 9). Within the psychometric paradigm assessment decisions are often made in which the voices of those "most knowledgeable about the context and most affected by the results" have been silenced (p. 10).

Moss (1998a) also argues that an interpretive approach to assessment requires a different kind of research agenda. She challenges the educational measurement community to broaden the set of evidence it draws on to include the "actual discourse that surrounds the products and practices of testing" (p. 7). Particularly generative is the interpretive perspective on dissensus (Moss and Schutz 2001), occasions when assessors fail to reach consensus. Rather than a problem to overcome, dissensus becomes another epistemic tool for greater insight into the validation process.

In summary, I return to my initial set of questions: From the perspective of these different traditions, psychometrics and interpretive: What is validation? What are its primary goals? What forms of rationality underlie it?

With respect to validation there appears to be general consensus concerning the central role of interpretation. Flyvbjerg (2001) argues that even the natural sciences have had to accept "the universality of hermeneutics" (p. 28), that even in the natural sciences

what constitutes facts, methods and theories is determined on the basis of a common interpretation of what constitutes scientific work. With respect to the primary goals of the interpretive process, again there seems to be some degree of convergence between the two traditions in that both seek to offer explanations for social life. There are, however, important differences between these traditions concerning what constitutes a legitimate explanation, in other words the basis upon which interpretations are evaluated. Consistent with mainstream social science, psychometrics prizes generalizability and reliability. The goal of interpretation in psychometrics is generalizability. These generalizations are based on the convergence of multiple sources of evidence; evidence which diverges or is inconsistent is a challenge to the validation process. By definition, generalizations are only meaningful to the extent that they are valid across individuals, settings and tasks; they must be context-independent. An interpretive approach to assessment, as proposed by Moss, does not deny the importance of generalizability and reliability but would challenge that they are necessary conditions for the validation process. The aim of interpretation, from the interpretive tradition is “not to uncover universals or laws” but to “explicate context” (Rabinow and Sullivan 1987, p. 14), an explication which takes full cognizance of the reflexive, intersubjective and discursive dimensions of social life. One of the crucial criteria for validation from this perspective is critical dialogue, a dialogue which acknowledges that context, difference and power are central to the formulation of assessment-based interpretations.

And regarding my final question, What forms of rationality characterize these two approaches? Flyvbjerg (2001) argues that underlying these two approaches are contrasting forms of rationality as ideal types. Underlying mainstream social science is a form of rationality prized by the natural sciences, one which privileges forms of inquiry which strive to be value-free as opposed to value-based, analytical as opposed to intuitive, instrumental as opposed to dialogical. Alternative forms of rationality (those characterized in opposition) have often been labelled ‘irrational’. Flyvbjerg’s central argument is that the kind of rationality prized by the natural sciences is inadequate for

comprehending the total spectrum of human activity¹². He calls for the explicit integration of properties which characterize high levels of expertise: context, judgment, practice, trial and error, experience, common sense, intuition and bodily sensation. I will return to this issue in Chapter 5 where I argue that this alternative rationality is crucial for an understanding of academic professional judgment.

While some theorists have argued that the notions of science which inform these two different types of rationality are incommensurable, others suggest that they need to be held together. Flyvbjerg (2001) notes that while these dualisms (general vs. particular, context-independent vs. context-dependent) “may facilitate thinking and writing, they inhibit understanding by implying a certain neatness that is rarely found in lived life ... rather than the ‘either-or,’ we should develop a non-dualistic and pluralistic ‘both-and’” (p. 49). Neither is Moss (1996) calling for one perspective over the other. She argues, “It is in the dialectic between the contextualized understanding of local meanings and the distancing analysis of regularities that the effects of symbolic power become most highly visible and open to critique” (p.22). Thus it is a dialectic perspective that “can provide a more textured and productive view of the social phenomena we seek to understand” (ibid.).

3.4 A social theoretical perspective

3.4.1 The double reality of social life

Such a dialectic perspective can be found in Bourdieu’s theory of social practice. In the introduction to *The Logic of Practice* (1990), Bourdieu notes, “Of all the oppositions that artificially divide social science, the most fundamental, and the most ruinous, is the one that is set up between subjectivism and objectivism. The very fact that this division constantly reappears in virtually the same form would suffice to indicate that the modes of knowledge which it distinguishes are equally indispensable to a science of the social world...” (p. 25). The goal, for Bourdieu, is to move beyond the antagonism between these two “modes of knowledge” while preserving the “epistemic virtues” from each of

¹² In constructing his argument Flyvbjerg is drawing heavily on the work of Dreyfus and Dreyfus (1986) who have developed a model of human learning processes which characterizes the kinds of rationality which distinguish between novices, competent performers and experts.

them (Wacquant in Bourdieu and Wacquant 1992, p. 7)¹³. In this final section, I turn to the task of constructing a social theoretical approach to validation inquiry which I argue will best illumine the questions of context, difference and power.

The social theoretical terrain is marked by heated debates between those who emphasize and give primacy to the search for underlying explanatory norms and rules and view action as a mere realization of these underlying structures (e.g. realists, structuralists), and those who focus on experiences and the local construction of individual or interpersonal events (e.g. constructivists, post-structuralists) (Wenger 1998). Muller (2000) characterizes this debate as the “dance of straw men” since neither position is tenable (p. 163). Constructivists cannot eliminate realism without being unresolvably self-contradictory. Realists, on the other hand, cannot eliminate constructivism without resorting to a form of argument that leads straight back to positivism.

Thus much of the contemporary preoccupation of social theory of the mid to late 1900s has been the search for a middle ground, the search for moorings in the aftermath of postmodernism. A common feature of this social theoretical terrain is the call for a ‘double view’ of social life which keeps not only structure and agency in view but their mutually constitutive relationship. It is this dual condition of social practice that is the very interest point for social science. For it is the resiliency, the permanency of structure that provides clues to patterns, to rules, to conditions which in part explain why we do what we do. But it is the relativeness, the perturbability, the contingency of agency that points to the possibility of change, to new ways of being and seeing which have not been considered before, to the very complexity of life in its unpredictability. Through this ‘double view’ lens, the focus of contemporary social theory is on ‘practice’, or more specifically *social* practice (Chouliaraki and Fairclough 1999). The focus is not on individuals, though the role of individuals within practice is crucial. Nor is it on society, though the macro social conditions which enable practice are relevant. The focus of

¹³ While both Bernstein and Bourdieu are critiquing similar sets of antinomies which they argue distort social scientific inquiry, their critiques are contributing to different debates. Bernstein (1976) defines ‘objectivism’ as a doctrine of foundationalism and his critique is aimed at empirical theory on the one hand, and postmodernism on the other. Bourdieu’s (1996) definition of ‘objectivism’ appears similar to Bernstein’s but his critique is directed at structuralism and post-structuralism.

social theory is on relationships: relationships between the acts of agents within practices and relationships between networks of practice and broader social conditions which constitute these networks, with a particular interest in their co-determination.

One of the fundamental goals of Bourdieu's work, as noted above, is to straddle some of the "deep-seated antinomies" which have held social science inquiry captive, antinomies between subject and object, structure and agency, subjectivist and objectivist notions of knowledge (Wacquant in Bourdieu and Wacquant 1992, p. 3). The double reality of social life, Bourdieu would argue, requires a "double reading" (p. 7). The first reading (from the "objectivist point of view") perceives society as a social physics, that is, an objective structure which can be "grasped from the outside, whose articulation can be materially observed, measured and mapped independently of the representations of those who live in it" (p. 8). The role of the researcher as "external observer" is to "ascertain the objective regularities (that agents) obey" (ibid.). The chief danger of this point of view is that "lacking a principle of generation of those regularities" (ibid.), the regularity slips from model to reality (what Bernstein 1976 refers to a reification as noted in 3.3.1). Bourdieu (1996) characterizes this distinction as one of rule as 'fit' (i.e. a rule that fits the observed regularities in a purely descriptive way) and rule as 'guide' (i.e. a rule that governs, directs and orients behaviour). "Pushed to its limits, objectivism cannot but produce an ersatz subject, and portray individuals or groups as the passive supports of forces that mechanically work out their independent logic" (Wacquant in Bourdieu and Wacquant 1992, p. 8).

The second reading (the "subjectivist point of view") perceives society as social phenomenology, that is, "Society appears as the emergent product of the decisions, actions, and cognitions of conscious, alert individuals to whom the world is given as immediately familiar and meaningful" (p. 9). From this perspective, pride of place is given to agency and subjective meaning. The danger of this perspective is that it "cannot go beyond a description of what specifically characterizes 'lived' experience of the social world, that is, apprehension of the world as self-evident, as 'taken for granted'. This is because it excludes any inquiry as to its own conditions of possibility" (Bourdieu 1990, p.

25). With specific reference to academic forms of evaluation as an example, Bourdieu (1996) notes, “While we should bear in mind, in opposition to a certain mechanistic view of action, that social agents construct social reality, both individually and collectively, we must take care not to forget ... that they have not constructed the categories that they implement in this construction” (p. 29).

Thus Bourdieu argues that in order to understand social life we need to transcend the false antinomies of objectivism and subjectivism; we need to apply a double reading to social life, one that “capitalizes on the epistemic virtues of each reading while skirting the vices of both” (Wacquant in Bourdieu and Wacquant 1992, p. 7).

3.4.2 Field and Habitus

For Bourdieu social practice is best accounted for in the mutually formative relationship between structure and agency. “The stuff of social reality, of action, no less than structure... lies in relations” (Wacquant in Bourdieu and Wacquant 1992, p. 15). ‘Field’, as Bourdieu uses the concept, is a “network or configuration of objective relations between positions” (Bourdieu in Bourdieu and Wacquant 1992, p. 97). He (1996) likens a social field to heavenly bodies within the same gravitational field which produce effects on each other from afar, where the boundary of the field is the limit of the effect. The distribution of capital, in other words social, economic and cultural resources, constitutes ‘positions’ within the field. But a field is not simply a set of “‘empty places’...but a *space of play* which exists as such only to the extent that players enter into it who believe in it and actively pursue the prizes it offers” (Wacquant in Bourdieu and Wacquant 1992, p. 19) (*italics in original*). This “space of play” is always in a state of tension as fields are sites of struggle, both power struggles within the field and struggles over power to define the field (Postone et al 1993).

If field is constituted by positions, habitus is ‘disposition’. Habitus consists of a set of objective relations (constituted by the field) ‘deposited’ within individual bodies, ‘dispositions’ which “function at every moment as a matrix of perceptions, appreciations and actions and make possible the achievement of infinitely diversified tasks” (Bourdieu

cited in Bourdieu and Wacquant, 1992, p. 18). Habitus is a “structuring mechanism” operating within individuals, though it is not “strictly individual”, granting individuals “practical mastery” within the field which they inhabit (Wacquant in Bourdieu and Wacquant 1992, p. 18). The relationship between field and habitus is precisely the means by which Bourdieu’s theory of practice transcends the traditional antinomies of the objective and the subjective, of structure and agency. For habitus is the embodiment of field within individuals, the “internalization of externality” (Bourdieu 1996, p. 29); it is “socialized subjectivity” (Bourdieu in Bourdieu and Wacquant 1992, p. 126).

Given the homologous relationship between field and habitus, habitus only generates dispositions which are compatible with the field. In other words, habitus tends to generate “reasonable”, “common sense” behaviour; it excludes “extravagances” and all that has been negatively sanctioned because it is incompatible with the field (Bourdieu 1990, p. 56). Habitus explains the ‘practical sense’ which individuals have as members of a particular field. Bourdieu, borrowing from the world of sport, refers to this practical sense as a ‘feel for the game’. “Because native membership in a field implies a feel for the game... everything that takes place in it seems *sensible*: full of sense and objectively directed in a judicious direction” (p. 66) (*italics in original*). Thus without explicit rules, codes or criteria, armed with this practical sense, individuals know how to get on with the myriad of classificatory tasks required of them as members of a particular field. Bourdieu relates this practical sense or practical knowledge to Aristotle’s concept of *phronesis*, sometimes translated as wisdom, a form of judgment which is pragmatic, variable, context-dependent, based on practical value-rationality (Flyvbjerg 2001). This practical sense or ‘feel for the game’ explains how an individual does what he or she “has to do” without “posing it explicitly as a goal, below the level of calculation and even consciousness, beneath discourse and representation” (Bourdieu in Bourdieu and Wacquant 1992, p. 128).

Bourdieu’s concept of habitus accounts not only for the particular classificatory scheme which an individual adopts as a member of a particular field, but extends to an account of a “unity of style” (Bourdieu 1998) which characterizes a group of individuals.

It accounts for a *collective* 'feel' or 'taste' positively sanctioned by individuals who are products of the same objective conditions of the field. Bourdieu (1990) refers to this *class habitus* as a "subjective but non-individual system of internalized structures, *common* schemes of perception, conception and action" (p. 60) (italics mine). It is class habitus that enables practices to be "harmonized" without any "conscious reference to a norm", without any "direct interaction", without any "explicit co-ordination" (p. 58). "The corrections and adjustments (that) agents themselves consciously carry out presuppose mastery of a common code" (p. 59).

I turn briefly to look at some of Bourdieu's (1988, 1996) empirical work which focuses on the class habitus of *homo academicus*, in particular the classificatory schemata which underlie professorial judgments. Consistent with all sociology, the goal of Bourdieu's work is to "uncover the most deeply buried structures" (1996, p. 1) of the social world, that is, the conditions for the production and reproduction of practice. In terms of academic assessment practices, Bourdieu's interest is to illuminate the "principles of vision and division" (p. 265) which lie behind academic forms of classification. What lies behind categories of professorial judgment of student performance? What are the organizing principles?

Bourdieu explores 'professorial judgment' within the context of the elitist French tertiary system in the late 1960's. He finds a strong correlation between professors' judgments (measured by clusters of adjectives used to evaluate student performance, e.g. 'fine', 'intelligent' at one end of the spectrum vs. 'servile', 'vulgar'), the marks which students receive (measured by average mark for the year) and the students' social origin (measured by professional origin of the father). What he finds is that the higher the student is in the social hierarchy, the higher are their marks, and the higher the frequency of commendatory judgments. Thus the most culturally rich students escape the most negative comments and average the highest marks. The least culturally rich escape the most insulting comments, although they receive the lowest average marks. The prime targets of negative judgment are the middle class, although they average middle range marks. Bourdieu's (1996) study exposes professorial judgment as a "socially constituted

preference system” (p. 2); academic classifications are, he argues, a guise for social classification. “Working as ideology in a state of practice, producing logical effects which are inseparable from political effects, the academic taxonomy entails an implicit definition of excellence which, by constituting as excellent the qualities possessed by those who are socially dominant, consecrates their manner of being and their lifestyle” (Bourdieu 1988, p. 204).

Bourdieu’s (1996) study shows how academic systems of classification are constituted both objectively by social structures and subjectively by mental structures; they are the “internalization of externality” (p. 29). Crucial to their effect is that they are never explicitly codified. They are subconscious, acquired through practical mastery. “As objective structures which have become mental structures through a process of apprenticeship ... academic taxonomies classify according to the logic of the structures whose product they are” (Bourdieu 1988, p. 207). Ultimately these systems of classification serve the interests of the field, that is, they serve to produce and re-produce the “space of possibilities” (Bourdieu 1988, p. 220) within the field. Bourdieu’s study challenges academic classifiers’ “perception of the norm” and the “propensity to assess performances in a dehistoricized and dehistoricizing language that is nevertheless laden with social connotations and assumptions, and therefore tailor-made for converting them into essences” (Bourdieu 1996, p. 11)¹⁴.

¹⁴ Bourdieu’s studies have been critiqued on the basis of their geographical and historical specificity. His critics have argued that his findings are based on a very particular and unusual higher education system, and that these data are *outdated*. Bourdieu counters these critiques (Bourdieu in Wacquant and Bourdieu 1992, p. 75-78). In terms of its geographical specificity, Bourdieu argues that the opposition between the general and the particular is a false antinomy; the goal of analysis is to “grasp the particularity within the generality and generality within the particularity” (p. 75). “The deepest logic of the social world can be grasped only if one plunges into the particularity of an empirical reality, historically located and dated...” (1998, p. 2). It is the uniqueness, he argues, of the French system that “makes it a highly propitious terrain for uncovering some of the universal laws that tendentially regulate the functioning of all fields” (Bourdieu in Wacquant and Bourdieu 1992, p. 75). In terms of historical specificity, Bourdieu argues that one of the goals of analysis is to uncover “trans-historical invariants or sets of relations between structures that persist within a clearly circumscribed but relatively long historical period” (p. 78). Proof that his work has to some extent achieved this goal can be found, Bourdieu argues, in the numerous studies which have been conducted more recently in a variety of contexts (e.g. the United States, Sweden and Japan) which suggest that the mechanisms of class production which Bourdieu identifies are still powerfully in operation.

To summarize thus far: the task of this chapter is to construct a theoretical perspective that will illuminate assessment as an interpretive act, with a particular interest in the role of difference, context and power. I argued that the first move, that of psychometrics, has not gone far enough beyond its objectivist roots, committed as it remains to the ideals of natural science. The second move, the interpretive tradition, offers a significant contribution to social inquiry with its particular emphasis on the reflexive, intersubjective and discursive dimensions of social life. The third move, a social theoretical perspective as theorized by Bourdieu, sets the interpretive act within the context of practice. I now turn to the task of clarifying the implications of Bourdieu's theory of social practice for my approach to the exploration of academic validation practices.

3.4.3 Validity inquiry from a social theoretical perspective

I propose four ways in which a social theoretical approach will illumine validity inquiry:

Firstly, consistent with the socio-cultural and interpretivist approaches to assessment reviewed in Chapter 2, as well as Moss' work reviewed in this chapter, a social theoretical perspective on assessment firmly endorses that 'context matters'. Academic assessment and validating practices only make sense within the logic of the field, within the logic of the academic profession and its disciplinary sub-fields. For it is the institutional and professional field that determines the epistemic 'principles of vision and division' which shape and inform the legitimacy of assessment-based interpretations: what is perceived to be legitimate knowledge, the legitimate criteria for assessing this knowledge, the legitimate assessors of this knowledge and the legitimate validators of these assessments. That there is contestation over these principles is a result of different positions within the field, positions which are defined by the uneven distribution of capital across institutions and individuals. Thus different kinds of knowledge are valid in different kinds of higher education institutions, and who can legitimately set criteria and assess knowledge will vary across institutions depending on their position in the field. Thus locating validation practices within the field requires an analysis of the forms of capital which matter in this field, the position of the institution in relation to these forms

of capital, and finally the epistemic rules which determine legitimacy and make sense of this 'game'.

While much of Bourdieu's interest lies with the analysis of macro-fields (e.g. the fields of education, art and economics), other social theorists have been more interested in local actions and interactions which serve to reproduce and transform structures (Wenger 1998, Engeström 1999). Acknowledging the importance of the field, their focus is more local; it is on 'communities of practice'. Communities of practice are not simply any social configuration but social groupings which, Wenger (1998) argues, cohere because members of these communities are mutually engaged in action, negotiating a joint enterprise and developing a shared repertoire. These communities of practice are themselves part of constellations of interconnected practices within broader institutions.

The concept of community of practice serves as a useful analytical tool for the analysis of academic validating practices; it emphasizes the importance of conducting validity inquiry, cognizant of how these validating practices are profoundly shaped by, not only the field but the particular communities (in this case disciplinary or specialist communities) within which they are located. While the academic field determines the broad epistemic rules which shape academic validating practices, the specialist communities of practice profoundly shape the particular epistemological orientation which academics adopt in their validation work. From the perspective of communities of practice we also get a close-up view of how capital is distributed unevenly among individuals and the effects of this on personal and professional identities, as well as interpersonal relations. Thus a social theoretical perspective illumines not only that 'context matters' but how 'power effects'.

Secondly, Bourdieu's concept of habitus, and specifically *class* habitus, suggests that by definition communities of practice to a large extent share a common set of "principles of vision and division" (Bourdieu 1996, p. 265), a common classificatory framework. This means that as members of the same field academics share a practical sense, a 'feel for the game'. With very few explicit codes, rules or criteria, they know how to get on

with the myriad of classificatory tasks required of them as members of the field. In reference to this practical sense Goodwin (1994) uses the term 'professional vision' to encapsulate the "socially organized ways of seeing and understanding...that are answerable to the distinctive interests of a particular social group" (p. 606). Members of a community, to the extent that they have been subjected to the same social conditions imposed by the field, share a relatively common "perceptual framework" (p. 616). Thus the ability of academics to interpret student performance as meaningful, is not a "transparent, psychological process but instead a socially situated activity" (p. 606). This vision is "lodged not in the individual mind but instead within a community of competent practitioners" (ibid.). Out of this collective, on-going practice of producing and re-producing these interpretations emerges structure, that is, the conditions or constraints which become internalized again into new acts of construction. This professional vision or what Taylor (1987) refers to as intersubjectivity (see also Goodwin 1994) is not synonymous with consensus; it is merely a condition for consensus. Only those with common points of reference can hold one another accountable for, as well as contest, the legitimacy of these reference points (Taylor 1987). Thus an exploration of academic validating practices from a social theoretical perspective will privilege an understanding of academic professional vision as central to the validation process.

Thirdly, I have argued that Bourdieu's concept of habitus is useful for explaining the common frame of reference which constitutes the assessment and validation practices of academic communities. At the same time the notion of positions within the field defined by differential access to capital is crucial for understanding contestation, that is, the reality that members of the same community can and do disagree; that *different* perspectives are as much a feature of communities of practice as consensus is. Thus I can speculate that while two members of a particular academic community of practice might share a common professional vision of what constitutes research in their field, different epistemological stances within that disciplinary field might result in different interpretations, and the effects of power might result in some interpretations being marginalized and others privileged. However, differentiation of capital may not be the only explanation for why differences of interpretation emerge. Each interpretive act

involves a whole complexity of contingencies which influence the final outcome, for example, the particular role the interpreter (assessor) is playing, the functions of the interpretation (assessment), and the object of interpretation (what the assessor has in view). I refer to this complexity of contingencies as an 'interpretive matrix' where 'matrix' is understood as a "context or environment" or "womb" (Oxford Dictionary), "something within which something else originates or develops" (Webster Dictionary). Thus a crucial part of understanding academics' validation practices comes from exploring how the local exigencies of an interpretive act may result in an adjustment or even a divergence from tacitly agreed-upon points of reference or criteria.

Fourthly, a social theoretical perspective illuminates how language plays a crucial role in all these meaning-making, position-taking, power-effecting acts. Though Bourdieu's empirical work focuses on language and he refers to the "magical power" (1988, p. 208) of language in the production of social practice, critical discourse analysis (CDA) is best placed to offer a theory of the relationship between social practice and language. CDA calls our attention to the discursive constitution of social practice: social practice is not merely expressed or reflected in language, but language actually constitutes social practice – more specifically, the constitution of belief systems, identities and interpersonal relations. Thus from a critical discourse perspective, actual assessment discourse becomes a powerful point of access into assessment as social practice. Discourse analysis as a means of access to social practice becomes the central theme of the next chapter to which I now turn.

Chapter 4: Methodology

What counts, in reality, is the rigour of the construction of the object.
(Bourdieu in Bourdieu and Wacquant 1992, p. 220)

4.1 Introduction

In the last chapter I argued that social theory calls for a ‘double view’ of social life, that is, a view that keeps not only both structure and agency in view but their mutually constitutive relationship. Thus the focus of the analytical gaze is not on individuals, nor on society, but on social practice; for *practice* is the point of connection between “objective relations and structures” on the one hand, and the “practical dispositions of subjects” on the other (Chouliariki and Fairclough 1999, p. 30). I also noted in the last chapter that one of the distinctive features of the study of social life, in contrast to the study of natural life, is that it is in part discursive. “Language is constitutive of reality, is essential to its being the kind of reality it is” (Taylor 1987, p. 54). I concluded the chapter by arguing that, given this constitutive relationship between language and social life, the analysis of discourse may be a particularly powerful point of access into assessment as social practice. Even though discourse is only one element of social life, most interaction substantively and centrally involves discourse; it makes sense, therefore, to focus on discourse to gain insight into social interaction (Chouliariki and Fairclough 1999). While Bourdieu acknowledges the constitutive role of language, it is to critical discourse analysis (CDA) that I turn for a more fully theorized relationship between language and social practice.

The titles which demarcate the content of this chapter and the previous one would suggest that there is a clear-cut distinction between ‘theory’ and ‘methodology’. However, this is a distinction which Bourdieu considers to be artificial. He argues that methodological decisions cannot be disentangled from our theoretical premises (Bourdieu and Wacquant 1992). Consistent with Bourdieu, Chouliariki and Fairclough (1999) argue that CDA is both theory and method: a method for analyzing social practices which make practical the theoretical constructions of discourse in social life. This analysis in turn contributes to the development and elaboration of these theoretical constructions. Given the acknowledged artificiality of the distinction between theory and methodology,

I have chosen to address the theoretical underpinnings of CDA in this chapter as a prelude to my discussion of its centrality in my methodological approach.

Wacquant (Bourdieu and Wacquant 1992) argues that one of Bourdieu's primary contributions to the task of social science is not theory per se, but a sociological method, a set of tools and procedures for the "construction of the object" (p. 5). "What counts, in reality", argues Bourdieu, "is the rigour of the construction of the object" (Bourdieu in Bourdieu and Wacquant 1992, p. 221). Thus my task in this chapter is to account for how I have constructed my object. I start (4.2) with the unfinished business of Chapter 3, that is theorizing the relationship between discourse and social life, and for this I draw on CDA, in particular the work of Norman Fairclough. I then turn (4.3) to describe my own methodological approach which I argue is a slight adaptation of Fairclough's. And finally, in the last section (4.4), I report on my exploratory study and the research design of the main study.

4.2 Critical discourse analysis

4.2.1 A rationale for CDA

Discourse analysis is a tool used by a variety of disciplines, for a variety of purposes (Titscher et al 2000). Gee (forthcoming) makes some useful distinctions between the tasks of critical and non-critical approaches to discourse analysis. All discourse analysis, he argues, performs two tasks: the form-function task and the language-context task. The form-function task is the analysis of how specific forms in a language are used as tools to carry out certain functions, to express certain meanings; for example, in English the dependent clause functions to express information that is often taken-for-granted or assumed. The language-context task is the analysis of how particular language forms take on particular 'situated meanings' in specific contexts¹. Gee argues that most contemporary approaches to discourse analysis assume a reflexive view of the relationship between language and context, that is, that an utterance influences what we

¹ For Gee (forthcoming) context refers to "an ever-widening set of factors that accompany language in use. These include the material setting, the people present (what they know and believe), the language that comes before and after a given utterance, the social relationships of the people involved, and their ethnic, gendered and sexual identities, as well as cultural, historical and institutional factors".

take the context to be and context influences what we take the utterance to mean. These two tasks, the form-function and language-context, are fundamental to any form of discourse analysis, but they do not make discourse analysis 'critical'. This requires an additional task.

Some forms of discourse analysis add a third task: they explain how form-function correlations (task 1) and language-context interactions (task 2) are associated with social practices (task 3). 'Non-critical' approaches to discourse analysis part with 'critical' ones in that the former would treat social practices solely in terms of patterns of social interaction, for example, how people use language to 'pull off' a job interview. Critical discourse analysis goes further and sees social practices in terms of their implications for things like status, solidarity, social goods and power (in Bourdieu's terms, the distribution of capital and its effect on power relations). For example, in relation to a job interview, a critical approach would be interested in how language functions in a job interview as a gate-keeping device allowing access to some people and denying access to others. A critical approach would understand that social practices are inherently and inextricably political, in the sense that they involve social relationships where social goods such as status and solidarity are at stake. Since language is part and parcel of social practice, it follows that language is itself inherently and inextricably political (Gee, forthcoming).

Thus critical discourse analysts share a common theory about the relationship between language and social practice, and a common set of theoretical assumptions about the inherently political nature of language. Given its political nature, they would also agree that one of the goals of discourse analysis is to critique and challenge dominant institutional practices. Indeed they would understand discourse analysis to be a political act in itself, "an intervention in the apparently natural flow of talk and text in institutional life that attempts to 'interrupt' everyday common sense" (Silverman and Torode cited in Luke 1995-96, p. 12).

4.2.2. Fairclough's CDA

While critical discourse analysts might share a common set of theoretical assumptions, there is no homogeneous method of CDA (Blommaert and Bulcaen 2000, Titscher et al 2000, van Dijk 1993). Branches of CDA tend to be identified with individual scholars in the field who have adopted different methodological approaches. Norman Fairclough is one of CDA's leading figures; his *Language and Power* (1989) is considered to be the landmark publication for the "start" of CDA (Blommaert and Bulcaen 2000, p. 454), and a substantial volume of work has followed. One of Fairclough's particular accomplishments has been to make CDA a platform for on-going dialogue between social theory and language theory (Slembrouck 2001). Indeed Chouliariki and Fairclough (1999) would argue that CDA, rather than providing a single theory or blueprint for analysis, is a shifting synthesis of other theories². This dialogue between social theory and discourse theory is particularly generative for my project.

I start by defining Fairclough's use of the term 'discourse' (4.2.2.1). Secondly, I present his three-dimensional theory of discourse (4.2.2.2), and finally explain how this theory of discourse becomes operationalized into a methodological framework for the analysis of discourse (4.2.2.3).

4.2.2.1. Definition of Discourse

The term 'discourse' has a variety of meanings. The term is used in its more narrow linguistic sense, in a more sociological – what Pennycook (1994) refers to as a 'Foucaultian' – sense, or some sense which lies in between (as I will argue Fairclough's does).

In its more narrow linguistic sense, discourse refers to samples of naturally occurring spoken or written language where the unit of analysis is above the sentence level, for example, a conversation or an interview. Discourse analysis in this tradition (referred to

² Though one of the criticisms of Fairclough's work (Blommaert and Bulcaen 2000) is that it is only in dialogue with certain linguistic theories, namely, systemic functional linguistics, and has failed to take cognizance of critical developments in other language-oriented sub-disciplines (e.g. linguistic anthropology).

earlier as non-critical) is typically interested in language use, that is, the relationship between language and the contexts in which it is used. For example, discourse might refer to particular kinds of language used in different settings, such as classroom discourse, medical discourse. The analysis would be interested in how these various contexts affect the use of language but as Pennycook (1994) points out the notion of context is limited, referring to immediate surroundings, speakers' intentions, background knowledge, or conversational rules.

More social theoretical understandings of discourse are credited to Foucault. Foucault's understanding of discourse does not refer to language or uses of language but to ways of organizing meaning, that are often, though not exclusively realized through language (Pennycook 1994). For Foucault, discourse analysis is not equated with linguistic analysis, or discourse with language. Discourse analysis is concerned rather with specifying socio-historically variable discursive formations, that is, systems of rules which make it possible for certain statements to be made and others not, at particular times and within particular institutional locations (Fairclough 1992). For example, in contrast to its narrow linguistic sense, for Foucault medical discourse refers to the dominant ways of thinking (i.e. systems of power/knowledge) in the medical profession and how these constitute "objects of knowledge, social subjects and forms of 'self', social relationships, and conceptual frameworks" (Fairclough 1992, p. 39). Given Foucault's pervasive influence on social theory, the term 'discourse' is often used loosely and interchangeably with the notion of practice.

While Pennycook (1994) argues that these different understandings of discourse are incommensurable³, critical discourse analysis draws from both of these distinctive but related meanings of discourse. For example, Gee (1996, 1999) distinguishes between 'Discourse' and 'discourse' where 'Discourse' (big D) refers to the more Foucaultian understanding of social practice and 'discourse' (small d) refers to its more linguistic

³ Pennycook (1994) argues that these various understandings of discourse (the traditional linguistic understanding, the CDA understanding, the Foucauldian understanding) are incommensurable on epistemological grounds. For example, fundamentally different notions of ideology and its relationship to inequality underlie these different notions of discourse.

meaning of language use. Gee (1999) argues that these two senses are “melled integrally” together (p. 7). Fairclough’s (1992) ‘discourse’ refers to “language use as a form of social practice” (p. 63). Like Gee, his use of discourse is clearly rooted in the linguistic notion of discourse as ‘language use’, but his addition of “as a form of social practice” signals a departure from its narrow linguistic sense to a notion of discourse as practice. Discourse is “not just a way of representing the world, but of signifying the world, constituting and constructing the world in meaning” (p. 64).

4.2.2.2 Theory of Discourse

I now turn to Fairclough’s (1992) theoretical elaboration of discourse as “language as social practice” (p. 63). By bringing together the more linguistic notions of discourse with the more social-theoretical notions of discourse, the outcome for Fairclough is a three-dimensional concept of discourse and a corresponding framework for discourse analysis. He makes the distinction between discourse-as-text (i.e. product), discourse-as-discursive-practice (i.e. processes of text production and interpretation), and discourse in its more Foucaultian sense of discourse-as-social practice. Discourse encompasses all three dimensions so that “any discursive event is...simultaneously a piece of text, an instance of discursive practice and an instance of social practice” (p. 4). This three-dimensional understanding of discourse is theoretically consistent with the notion of the interconnectedness of language and social life (Fairclough 1989, 1992).

I now turn to a discussion of each of these different theoretical dimensions of discourse.

Discourse-as-text

As illustrated by his diagram (see diagram 1), text is central to Fairclough’s theory of discourse. Fairclough draws on a Hallidayan understanding of ‘text’, that is, language doing something (Halliday and Hasan 1985). Although Halliday argues that text has both product and process dimensions, Fairclough (1989) uses text to refer specifically to the product of the process of text production. A text is analyzed for “traces” of its production and “cues” to its interpretation (Fairclough 1989, p. 24).

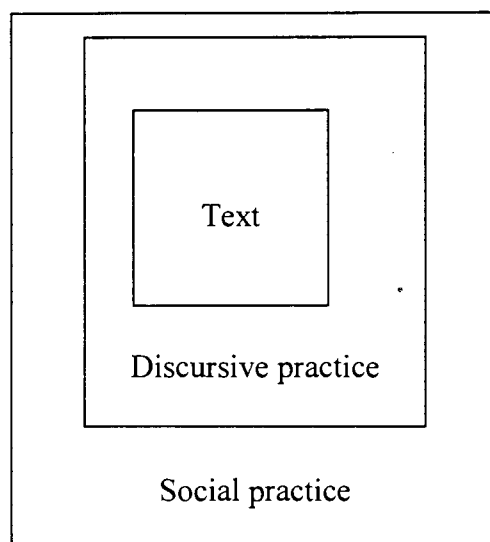


Diagram 1: Fairclough's (1992) three-dimensional conception of discourse

Consistent with a functional theory of language (Halliday 1994), an analysis of discourse-as-text begins with the recognition of function as the fundamental principle of language. This is not simply to say that language is used for a variety of different purposes, but more significantly that function is built into language; it is a "fundamental property of language itself" (Halliday and Hasan 1985, p. 17). Drawing on Halliday, Fairclough (1992, 1993) specifies three functions of language or three ways in which language is constitutive; it is constitutive of social identity, social relations, and systems of knowledge and belief. The identity function refers to the way in which social identities are set up in discourse. The relational function refers to how social relations between discourse participants are enacted and negotiated. The ideational function refers to the ways in which texts signify the world and its processes, entities and relations. As Gee (forthcoming) puts it, the ideational function is the way in which texts constitute what we take the world to be and the way we take things in the world to be related to each other. Thus the aim of analysis of discourse-as-text is to understand how speakers'/writers' language choices amount to "choices about how to signify (and

construct) social identities, social relationships and knowledge and belief” (Fairclough 1992, p. 76).

Analysis of these language choices pays close attention to the formal features of texts. Fairclough (1992) organizes these formal features under four broad categories: vocabulary, grammar, cohesion and text structure. Vocabulary deals mainly with individual words, grammar deals with words combined into clauses and sentences, cohesion deals with how clauses and sentences link together and text structure deals with large-scale organizational properties of texts. Thus, for example, textual analysis would explore the ways a particular exchange structure (i.e. the pattern of conversational interaction) constitutes particular ‘subject positions’, and how these positions may privilege a particular perspective and marginalize others. As another example, textual analysis would attempt to expose how the modality of a particular utterance (i.e. the degree of certainty with which something is said) signals ways in which speakers are negotiating relations with other speakers.

Discourse-as-discursive-practice

The second dimension of Fairclough’s theory of discourse is ‘discourse-as-discursive-practice’. By this Fairclough (1992) means the processes by which texts are produced, distributed and consumed, processes which vary significantly depending on the type of text and its social context. The complexity of discursive practice is compounded by the multitude of positions which producers and interpreters occupy which influence the text product and how it is interpreted. Part of this positionality has to do with what Fairclough (1989) refers to as, ‘members’ resources’, that is, “what people have in their heads and draw upon, including their knowledge of language, representations of the natural and social worlds they inhabit, values, beliefs and assumptions” (p. 24). These resources constitute in effect a ‘mental map’ of the social order. This mental map enables a reading of a situation which foregrounds certain elements and backgrounds others; it relates elements to each other in certain ways. In addition to members’ resources, positions are also constituted or constrained by the social practice that the textual processes are a part of, since these practices determine what elements of members’

resources are drawn upon, and how they are drawn upon, whether in conservative or transformative ways. Thus the aim of analysis of discourse-as-discursive-practice is to understand how processes of text production and interpretation are constrained by members' resources, as well as by the broader social conditions of practice (Fairclough 1992).

While this discourse-as-discursive practice dimension is a central feature of Fairclough's theory, the extent to which CDA is able to access these processes of text production/distribution/consumption has come under critique. Blommaert and Bulcaen (2000) argue that the treatment of context is arguably the biggest methodological weaknesses of CDA. One of the particular methodological problems is the "*framing of discourse in particular selections of contexts*, the relevance of which is established by the researcher but is not made into an object of investigation" (Blommaert 2001, p. 15) (italics in original)⁴. To some extent this is unavoidable; analysts will always be bringing particular presuppositions and assumptions about what is relevant to their analysis. But Blommaert and Bulcaen (2000) argue that this is a particularly pressing problem in CDA where analysts project their own political biases and prejudices onto their data and analyze them accordingly.

In relation to the analysis of textual processes, another critique directed at CDA is the extent to which language users' own interpretations are taken into account. As Slembrouck (2001) argues, "CDA has been mostly text-based and at best speculative in its claims about how actual participants are likely to interpret texts, exchanges and moves within talk" (p. 43). The problem with CDA is that it relies solely on extrapolation from text: there are no other explicit methods for getting access to the producers' and consumers' interpretations of their own texts⁵. Language users' own interpretations are assigned to them by proxy (Widdowson 1998). The resulting risk is that CDA theorizes the language-user out of sight (Slembrouck 2001). Thus, despite claims to the contrary,

⁴ This is a criticism also raised by Verschuere 2001 and Widdowson 1998, 2000.

⁵ This is a point that Fairclough (1992) acknowledges: "How people interpret texts in various social circumstances is a question requiring separate investigation" (p. 86).

Blommaert (2001) argues that CDA's ultimate ambition still lies in the interpretation of text, rather than the interpretation of social practice through text.

Acknowledging these critiques of CDA's methodological weaknesses, I would argue that in Fairclough's later work (Chouliariki and Fairclough 1999) there is an acknowledgement of these limitations, as well as some attempt to address them. Chouliariki and Fairclough (1999) argue that critical discourse analysis should be seen as only one aspect of research into social practices working together with other social scientific methods. They specifically mention ethnography as a method which can "establish precisely the sort of knowledge that CDA often extrapolates from text" (p. 62). There are also some subtle methodological shifts in what they refer to as, a "new version" (p. 59) of their analytical framework which I believe begin to address some of the problems identified with CDA and context. These concerns, namely, the methodological limitations of CDA and in particular its failure to incorporate into its analysis language users' own understandings, are central to the rationale for my own methodological approach which I discuss in 4.3.

Discourse-as-social-practice

Fairclough (1989, 1992) argues that no account of discourse is complete without an account of how the processes of production and interpretation (the focus of the second dimension) are socially determined. Thus Fairclough's third dimension, discourse-as-social-practice, emphasizes the social conditions of production and interpretation, that is, the way in which discursive practices are constrained and made possible by 'social order'. Social order for Fairclough relates to three levels of social organization: the immediate social environment in which the discourse happens, the level of the social institution, and the level of society as a whole. To the extent that much of social life is discursive, this social order is, at least in part, a discursive order or an order of discourse. Fairclough uses the term 'orders of discourse' (which he borrows from Foucault) to refer to the totality of discursive practices within an institution or society and the relationships between them. This social or discursive ordering shapes members' resources, that is, the mental map which people bring to the text production and interpretation, which in turn

shapes the way in which texts are produced and interpreted. Thus consistent with Bourdieu's analysis of the field, the aim of analysis of discourse-as-social-practice is to expose the social conditions of production and interpretation which constitute social practice.

4.2.2.3 Fairclough's framework for analysis

Chouliariki and Fairclough (1999) set themselves the task of strengthening the theoretical grounds of CDA, with a particular interest in the social theoretical underpinnings of CDA. The outcome is that, while Fairclough's three-dimensional theory of discourse remains consistent, they offer a new version of the analytical framework. They argue that the rationale for this revised framework is a reflection of CDA as a method which evolves to accommodate new theorizations of the relationship between social life and language. Of particular interest are the three stages of analysis which they propose need to be conducted – the analysis of the conjuncture, the analysis of the particular practice or practices, and the analysis of the “discourse proper” (p. 63). (I elaborate on these stages below.) These stages serve as the basis of my own methodological approach with some adaptations. I now turn to a description of my approach.

4.3 My methodological approach

Consistent with Chouliariki and Fairclough's framework, my own approach acknowledges the analysis of “discourse proper” (ibid.) as a crucial stage of analysis for exposing the ideational and relational fabric of social practice, as well as identifying traces and cues to the processes of interpretation and production. However, I argue that explanations of social life are constructed on the basis of different kinds of texts. There are what I refer to as the focal texts; in my study this is the actual dialogue between colleagues in the moderation meetings. In addition to these focal texts, there are other texts. There are the participants' own interpretations of those focal texts, and there are the participants' interpretations of the researcher's analysis. All of these texts shape the researcher's explanations. The limitation of CDA, as noted earlier, is that it privileges the analysis of the focal texts and neglects the others. Thus while it theoretically accounts for

the *processes* of interpretations and production, it is methodologically limited in its access to these processes. These require the use of other kinds of analysis and other methods of data collection. Thus in my approach I have firstly de-centred the analysis of focal texts as privileged by CDA. Instead I conceptualize social practice as my 'object' of interpretation. Secondly, I argue that the interpretation of social practice requires a variety of methods, including discourse analysis. As noted above, Chouliariki and Fairclough (1999) concede that CDA should be seen as only one aspect of research into social practice "working together with other social scientific methods, particularly ethnography" (p. 61). My own methodological approach is thus best characterized as critical discourse analysis with the incorporation of some ethnographic methods.

My approach to the "construction of the object" (Bourdieu in Bourdieu and Wacquant 1992, p.220) is the identification of four constitutive elements of social practice which were foreshadowed in Chapter 3. I argued that social practice is constituted by social structure, that is, particular macro-social conditions of constraint and possibility. It is also constituted by more micro-level networks of practices within which it is located, what Chouliariki and Fairclough (1999) refer to as conjuncture. Social practice is also constituted by a particular instance of practice or a particular event or occasion. And finally, social practice is constituted by text; the very choices of language that are made by speakers (consciously and unconsciously) serve to constitute and are constituted themselves by social practice.

These multiple constitutive elements of social practice – social structure, conjuncture, event, text – operationalize into a methodological approach, a series of stages of analysis, each stage exposing different aspects of the social practice of interest, in my case academic validation practices. In order to illumine these constitutive elements, I draw on a set of 'tools' which expose not only the constitutive elements of social practice but the relations between them. Bourdieu notes that his theoretical tools (e.g. field, habitus) are "pense-bête" (i.e. memory-joggers) which remind the analyst to "think relationally" (Wacquant in Bourdieu and Wacquant 1992, p. 16). Thus, the concept of field 'works' when it exposes the constituting macro-social conditions of

practice. The concept of community of practice ‘works’ when it exposes the more immediate social networks which constitute academics’ social and professional identities. Genre⁶, as an analytical tool, ‘works’ when it exposes the way particular social conventions privilege certain perspectives over others. Facework⁷, as an analytical tool, ‘works’ when it exposes the way in which the text functions ideationally and relationally in the maintenance or transformation of relations and practices. Drawing on both Bourdieu and Fairclough, I briefly re-cap these constitutive elements of social practice. I note in particular what aspects of practice are exposed, what analytical tools enable this exposure, and the key questions which will be used to interrogate the practice. I will describe the specific methods of data collection in the discussion of the research design (4.4).

4.3.1 The analysis of social structure

The analysis of social structure seeks to account for the conditions which both constrain and make possible the social practice, conditions which ensure its on-going production and re-production. In Fairclough’s terms this requires a description of the social order, the immediate, institutional and societal contexts within which the social practice is located, as well as the orders of discourse which constitute this social order. In Bourdieu’s terms this requires a description of the field, a network of positions in social space defined by the distribution of capital.

Drawing on secondary data, I seek to explore how academic validation practices are constituted by a particular institutional identity, that is, how universities give rise to particular conditions of constraint and possibility for academic professional judgment, and how these conditions shape the kinds of validation systems which are considered legitimate within such institutions.

⁶ Genre refers to “a relatively stable set of conventions that is associated with, and partly enacts, a socially ratified type of activity” (Fairclough 1992, p. 126).

⁷ Facework refers to the various strategies which people use in dialogue to manage and minimize the threat to their own face and that of others.

Some of the key questions that I use to interrogate academic validating practices include:

- What are the forms of capital which define institutional and individual positioning within the field of higher education?
- Given these forms of capital, what are the “principles of vision and division” (Bourdieu 1996, p. 265) which legitimate academic validation practices?
- In what ways do struggles within and on the boundaries of the field challenge the legitimacy of these principles, and hence practices?

I now turn to the analysis of conjuncture. Chouliariki and Fairclough (1999) note that there is no clear cut-off point between social structure and conjuncture. The task of these two types of analysis is to specify the overall frame within which the practice is located.

4.3.2 The analysis of conjuncture

The analysis of social practice as a conjuncture of events seeks to understand academics' validation practices, not as an isolated act or event, but rather as a chain of events within a network of practices. Chouliariki and Fairclough (1999) define conjunctures as “relatively durable assemblies of people, materials, technologies and therefore practices around specific social projects” (p. 22). Within this “relatively durable” network of practices which constitute the field, a conjuncture represents a particular “path” (p. 61) or chain of events associated with the social practice of analytical interest. The aim of analyzing conjuncture is thus to account for the series of events which comprise the practice, as well as the people, materials, procedures and rules. The focus is on the habitual nature of the practice, for example, the annual routines which comprise the assessment of final year projects in any given department. The particular vantage point of conjuncture is that it allows the analyst to “trace through time not just individual events but conjuncturally linked series of events in both sustaining and transforming (re-articulating) practices” (p. 22).

For the analysis of conjuncture I draw on Wenger's (1998) concept of community of practice but retain the particular emphasis of conjuncture, that is, a community of people engaged in a specific social project over a designated period of time. I use the term 'community of practice' to refer to all the people, materials, technologies and procedures which are involved in the assessment of the final year projects. The membership of the communities of practice which are the focus of my study include staff members who are involved in the assessment of the final year project, as well as individuals outside of institutional boundaries, for example, the external examiners. It excludes students from its membership, since in the cases which I am studying students are not involved in the assessment of their own projects.

Thus the analysis of conjuncture seeks to describe the community in terms of its members and their roles, status, and seniority within the community. Then I search for explanatory connections between the community's 'ways of seeing' and its validation practices.

Some of the key questions in the analysis of the conjuncture include:

- Who are the members of these communities of practice? What are their roles, status, and seniority within the community of practice?
- What events, technologies and procedures make up the final year project process? How have these evolved over time?
- What global, national, and institutional events do members of this community note as significant in shaping their practices? How do they themselves interpret the significance of these events?
- What are some of the features of members' 'interpretive matrices'⁸ which they bring to bear on the interpretation of student performance?

⁸ As I noted in 3.4.3, this refers to the interpretive context or environment which assessors bring to bear on their assessment of student performance. This concept is developed in 5.3.2.2 and 5.4.2.

4.3.3 The analysis of event

The analysis of social practice as constituted by a particular event or occasion shifts the gaze away from the more structural perspective on social practice (the analytical focus of social structure and conjuncture) and zooms in on social interaction. Chouliariki and Fairclough (1999) stress the importance of both the structural dimension of social practice and its interactional dimension. They note,

It is easy for a critical social science oriented to the abstract structures and social relations...to miss the richness and complexity of social interaction. And yet it is of crucial importance for any dialectically conceived critical theory to grasp the complex qualities of social interaction. For not only is any social structure dependent upon its ongoing instantiation in social interaction, and not only is it in social interaction that structures are problematized and contested, but social interaction is also the nursery for new social forms and themes of all sorts (Volosinov cited in Chouliariki and Fairclough 1999, p. 38).

The particular instance of social interaction through which I choose to view academic validating practices is a moderation event. This is the occasion when staff come together to discuss the marks which have been allocated to student performance. I perceive these events to be particularly generative sites for exploring academic validation practices as academics have to (to some extent) make explicit to other colleagues their inferences and justifications for these inferences.

I draw on the analytical tool of genre for the analysis of event. Genre, a term borrowed from literary theory, is a “relatively stable set of conventions that is associated with, and partly enacts, a socially ratified type of activity” (Fairclough 1992, p. 126). I would argue that marking workshops and departmental examination meetings are genres of academic social practice. Though genres are characterized by a particular thematic content, a particular style, a particular compositional structure (Bakhtin cited in Chouliariki and Fairclough 1999, p. 49), they serve to delimit a range of options rather than specifying a single rigid pattern. The analytical value of genre is in exposing the way that conventional aspects position participants within the social interaction. Kress (1989)

argues that genres construct positions or roles which the participants in the genres occupy. Through genre, the position “encodes and presents (the participant with) a set of possibilities ...of being a certain kind of social agent with all that entails” (p. 40). Of course participants do not have to comply with the demands of the position constructed; they can resist. How individuals actually take up these positions is what the analyst seeks to explore through the analysis of text, that is, the analysis of actual interaction/dialogue (which I discuss in 4.3.4).

Some of the key questions that I draw on for the analysis of the event include:

- What is going on in this event? What is the meeting about? What has preceded it? What follows? How did it come about? Who is attending it? Why are they attending? Where is it happening? What materials are available?
- What is the dominant discourse of the meeting and how does this serve to foreground particular agendas and background others?
- What is the exchange structure of the meeting and how does this position participants in particular ways?
- What common prior experiences are participants drawing on and what effect do these have on how individuals are positioned?
- How does the presence of a researcher and tape-recorder position participants?

4.3.4 The analysis of text

To analyze social practice as text is to zoom in even closer to the analysis of social interaction, and I specifically focus on the analysis of what I refer to as validating arguments, that is, the actual discourse which academic staff use to interpret student performance and justify their interpretations, whether to students, colleagues or the researcher. I am interested in the ideational and relational functions of the text⁹. With respect to the ideational function I am interested in the kinds of evidence which participants privilege in their justifications, and what this suggests about the interpretive matrices (defined in 3.4.3) which they bring to bear on the student performance. With

⁹ I combine Fairclough’s social identity and interpersonal functions into one function which I refer to as relational. This is consistent with Halliday’s original meta-functions.

respect to the relational function I am interested in 'facework'. Linguists define 'face' as the public self-image that every person wants to claim for him/herself. There are two related aspects of face. Negative face is people's general wish for freedom from intrusion and imposition. Positive face is people's general wish to be well-thought of (Brown and Levinson 1987). "All linguistic action involves face-threat of some kind" (Green cited in Kasper 1997, p. 378) and thus facework refers to the various strategies which people use in dialogue to manage and minimize the threat to their own face and that of others (Ting-Toomey 1994).

I explore this ideational and relational functioning through two types of analysis: analysis of the topic and analysis of modality. Topic analysis attempts to get at what speakers are talking about. Fairclough (1992) argues that the way people chain topics together can offer insight into the "common sense structuring of their lifeworld" (p. 155). The aim of the analysis is to determine the overt topic of the dialogue, which is identified by the noun phrase in prominent position and all the referents to it. Of greater significance, however, are the topic shifts, and what underlying ideational work these shifts point to. Thus an analysis of the topic chain exposes the ideational functioning of the text; it exposes the kinds of evidence which assessors value and the contestation over these values.

The analysis of modality identifies a variety of textual features which signal the degree of commitment or certainty that a speaker has for their utterance, features such as modal auxiliary verbs (e.g. must, might, can, should), modal adverbs (e.g. probably, definitely), hedges (e.g. sort of, like), and intonation patterns. Fairclough (1992) notes, however, that it is often difficult to disentangle the commitment that speakers have for their utterances from the commitment or solidarity they have with other interactants. Thus an utterance expressed with low modality could signal a lack of commitment for the utterance, or it might be a softening of what might otherwise be considered a face threatening act (Bloor and Bloor 1995; Brown and Levinson 1987). For example, a staff member may have a strong opinion about a particular instance of student performance, but will express that opinion in such a way as to not offend another colleague. An

analysis of modality exposes the manner in which the ideational and relational functions of texts “are woven together” (Halliday and Hassan 1985, p. 23), and work simultaneously in the construction of meaning.

Some of the key questions that I draw on for my analysis of text include:

- How do the assessors in their different roles (e.g. whether supervisor, second marker, external examiner) constitute a particular validating argument, based on what kinds of evidence?
- What do speakers’ utterances reveal about the degree of certainty they feel for their interpretations?
- What do speakers’ utterances reveal about the ways in which they seek to preserve or challenge collegial relations?

In closing I note some general points about this methodological approach. Firstly, the stages of analysis do not proceed in any particular order, but rather they are iterative. Each stage of analysis will illumine aspects of social practice which from the perspective of another constitutive element are obscured. For example, the analysis of text will point to traces of structural conditions, and the analysis of social structure will point to potentially salient features for text analysis. Secondly, each stage combines both description and explanatory analysis. For example, the analysis of social structure at one level is a description of the immediate, institutional and societal context. But the analysis must move beyond description to explanations for the connections between the macro-level conditions and the more micro-level practices, for example, institutional identity and the validation practices within the institution. Thirdly, while all stages of analysis are important to an explanation of the social practice, the analyst may prioritize some stages over others. In my own analysis I have privileged the analysis of conjuncture, event and text over the analysis of structure because I feel that there has been very little research of assessment practices at the micro, interactional level.

Fourthly, I note some of the criteria for the validation of my own study. I noted earlier the multiple texts that contribute to my explanations of the social practice, namely,

my interpretations of the focal texts, the research participants' interpretations of those texts, and their interpretations of my analysis in progress. Implicit in these multiple layers of interpretation are three principles which contribute to the validation of my interpretations. These principles are adapted from Gee (1999). They are linguistic detail, coverage, and convergence. With respect to linguistic detail, Gee (1999) argues that the validity of discourse analysis rests on the extent to which the communicative functions being uncovered in the analysis are linked to grammatical devices that "manifestly can and do serve these functions, according to the judgments of 'native speakers' of the social languages involved and the analyses of linguists" (p. 95). Thus part of the validity of my story rests on the strength of my argument for the particular links between the features of text (e.g. modality) and the social function this modality serves (e.g. solidarity). With respect to coverage, my approach attempts to explore the social practice using multiple stages of analysis and multiple methods of data collection. With respect to convergence, my approach has built-in feedback loops which provide opportunities for interpretations to be challenged (mine as well as those of the research participants). While convergence of interpretations is important for validation, divergences do not necessarily falsify those interpretations; they may rather signal the need for deeper analysis. For example, in the Engineering case a discrepancy between what people said in interview and what emerged from the analysis of the validating arguments led me to explore more deeply the ways in which the genre of the meeting itself was constituting particular outcomes.

4.4. Research Design

As noted earlier, my research interest arose out of the context of my own practice as an educational developer working with groups of staff responsible for the assessment of student performance. My particular interest was why assessors, in the context of marking workshops, disagreed. It made sense to me therefore to return to these sites of practice to explore them. This return as a researcher (as opposed to an educational developer) had two implications for me. Firstly, I needed throughout the process of the research to consciously 'objectify' my position as researcher. What were my own interests in this project? In what ways would my identity (as a young, white, American woman; an academic of relatively junior rank compared to my interviewees; my institutional position

within the Academic Development Programme¹⁰) influence my inquiry? In what ways would my very presence (whether as participant or observer) influence the data? I address these specific issues in the report on my exploratory study. Secondly, I needed to make a break with common sense. As an educational developer my goal had been to assist staff to reach consensus about what they value in student performance. Disagreement was perceived as a problem to be fixed, and common sense educational practice dictated that explicit assessment criteria were part of the solution to this problem. As a researcher however I needed to shift from an understanding of disagreement as a problem to be 'fixed' to an understanding of disagreement as a problem that needed to be understood and explained. These understandings and explanations could then serve as the basis of discussion between me and others invested in these practices. Did the problem require changes in existing practices, and if so, what might some of the challenges be in bringing about those changes?

I thus decided to return to the site of practice in an exploratory study which I conducted in 1997. The insights gained in this exploratory study then informed my main study conducted at SAU. The main study consisted of two case studies, the first conducted in 1998-1999 and the second case study conducted in 2000-2001. In 4.4.1 I report on the exploratory study, and the rest of the chapter (4.4.2-4.4.3) describes the design of the main study.

4.4.1 Exploratory study

4.4.1.1 Research design and rationale

I went into my exploratory study with some initial hunches about why assessors did not always come to consensus about what marks to award students' performance in the context of marking workshops. I speculated that causes for these differences might in part be explained by assessors looking for different qualities in student writing. I also speculated that even when assessors did have explicit criteria (in the form of a marking memorandum) they might be applying the criteria differently, and that in addition to the

¹⁰ The Academic Development Programme is a department which offers a wide range of services to students from educationally disadvantaged backgrounds and works closely with faculties and other support services in areas such as curriculum design, teaching methodology, assessment of student performance and course evaluation.

explicit criteria there might be other implicit criteria which assessors use to assess. Thus at this exploratory stage I was interested in what can marking workshops reveal to us about:

- The particular qualities which academics are looking for in student writing
- How the explicit criteria are used to assess student performance
- Additional implicit criteria (or other factors) which influence the assessment

In 1997 I conducted an exploratory study in the context of two undergraduate courses in my own institution, a first year Humanities course and a second year Science course. There were a number of factors that made these courses appropriate sites for an exploratory study. Firstly, they both had a tradition of marking workshops. Secondly, the courses were situated in different disciplines, the one considered a 'hard' science and the other a 'soft' one. I felt it was important to explore the commonly-held perception that in contrast to the social sciences, the assessment of writing in the 'hard' sciences is more 'objective', that is, given that the range of appropriate responses is more narrow in scientific writing, the assessment is less likely to be influenced by assessors' biases. Thirdly, my selection was motivated by the fact that I was already involved in these courses as an educational developer, though to a lesser extent in the Humanities course than in the Science course. Given the sensitivities around assessment issues, I felt that it was important at this stage to explore in sites where I had already to some degree established a relationship of trust with the staff involved in these courses.

These last factors foreshadow two important methodological issues which I explore in the next section. Firstly, how the relationship between researcher and the researched influences access to what kinds of data, and secondly, the role of the researcher as perceived by the researched. Pertinent to my situation was the potentially conflicting role between researcher and educational developer.

4.4.1.2 Data-collection and methods of analysis

Although I had a good working relationship with the course convenor of the Humanities course, I had only had one previous formal encounter with the rest of the staff

involved in the course. In conversation with the course convenor he mentioned that they had conducted a marking workshop the previous year and had found it a useful exercise which he intended to repeat. When I asked if I could attend the workshop he insisted that I convene the marking workshop; he argued that this would “lend credibility to the exercise” (field notes, 14/5/97). However, when I requested permission to tape-record the workshop, he denied my request. He noted that there were sensitivities in the group and that some of the participants might feel “policed” (field-notes, 14/5/97). He suggested that once I was better known to the group, tape-recording would be acceptable. I therefore requested that a colleague attend the workshop with me in order to take detailed notes of the discussion. In addition to the written record of the workshop, I also collected all workshop documentation, the assessment task and the writing samples. The convenor’s denying my request to tape-record is an example of the kind of sensitivities which surround assessment practices and thus supports my argument for the influence of my presence on the proceedings of these moderation meetings, and thus on the data.

In contrast to the Humanities course, I had worked closely with the Science course staff for a period of three years¹¹. I had been involved in facilitating a number of similar marking workshops and the tradition of such workshops was part of their established moderation process. This particular workshop was convened by two of the members of the course staff, and I was given permission to attend the workshop as an observer and to tape record the discussion. In addition to the audio-recording, I also collected all workshop documentation. In both sites, but particularly in the Science course, I also had access to contextual information due to my previous involvement in these departments (e.g. information about who the participants were and what roles they played within the course).

The aim of the exploration at this stage was less about finding answers to my questions than to confirm whether the questions were worth asking. It also served as an opportunity to explore theory that would help me make sense of the complexity which was emerging, as well as to assist me in developing a methodological approach. The

¹¹ I was involved in assisting this course to develop a writing project within their undergraduate curriculum.

analysis was informed by my early wrestling with critical discourse analysis as both a theory and a method for exploring social practice, with a particular interest on how genre analysis and the analysis of intertextuality¹² might expose other factors which were influencing assessors' interpretations of student performance.

4.4.1.3 Analysis

Both marking workshops took place in May 1997 within the same week. Though the genre of writing required of the students differed in each course – the Humanities' course required students to write an essay and the Science course required a report – both types of writing could be described as instances of complex tasks, in other words, tasks that permit students substantial latitude of interpretation and response; tasks which reflect the integration of multiple skills and knowledge; tasks which require expert judgment for assessment (Moss 1992).

The broad aims of each of the workshops were similar. In each case the stated aim of the workshop was to provide an opportunity for markers to build consensus around the marking criteria and their application to samples of student performance prior to the marking of their own reports/essays. The manner in which the workshop was organized was also similar. In each case two essays/reports were selected to illustrate different levels of performance (i.e. relatively 'good' and 'poor'). All forms of student identity were removed from the samples of writing. Workshop participants were requested (either in the workshop as in the case of Science, or prior to the workshop as in the case of Humanities) to allocate marks to the samples and be prepared to justify their marks to other workshop participants.

Both workshops used explicit assessment criteria although they took different forms. In the Science marking workshop a marking memorandum had been prepared beforehand by the writing project co-ordinator and staff was asked to use the memorandum in assessing the samples of writing. The assessment criteria (listed in the memorandum) had

¹² 'Intertextuality' is the property texts have of being full of snatches of other texts, which may be explicitly or implicitly demarcated (Fairclough 1992).

been given to the students when the topics were handed out and thus were not negotiable for the staff. In contrast, in the Humanities workshop the assessors marked the samples of writing without a commonly agreed upon marking memorandum. During the preparation for the Humanities workshop, the course convener informed me that the course did not have any standard set of criteria and he was reluctant to propose any in advance of the workshop. He thus started the workshop by proposing some criteria which were written on the board and discussed by the rest of the workshop participants.

In terms of the workshop participants, the Science marking workshop involved five members of staff (plus me); all were full-time members of the academic staff from the department, though at different levels of seniority. The senior professor was the course convener and did some teaching on the course. One of the other more junior members of staff was the writing project coordinator and had primary responsibility for conceptualizing and administering the writing assignment. In consultation with the others, she had designed the topic and drawn up the marking criteria. Two of the others also contributed to teaching on the course. One was not involved in the course at all¹³. They were all specialists in the broad disciplinary area which the topic was assessing, but two of the participants were not familiar with the specialization which was being assessed. None of the workshop participants, except the writing project coordinator who had selected the samples, knew the students who had produced the reports under discussion.

The Humanities marking workshop involved eight members of staff (plus me and another colleague). The course convener was a full-time member of the academic staff, as was one other participant, and they both taught on the course. The course convener was also responsible for designing the essay topic. All other participants were tutors, that is, postgraduate students responsible for running tutorials¹⁴. Though all the participants

¹³ This department had a policy of sharing the marking load of this writing intensive course across the department so that staff was recruited to mark who had no other involvement in the course.

¹⁴ In the South African higher education system, 'academic' typically refers to full-time, permanent members of staff, whereas 'tutors' refers to post-graduate students who have part-time, temporary appointments. Typically the academic staff is responsible for the lectures and the tutors are responsible for the tutorials.

were involved in the course, the academic staff had less direct contact with the students except in large lecture mode teaching, and typically they did not mark. The tutors, however, had direct contact with their own tutorial group/s, and as a general rule assessed their own students' essays.

Consistent with my experience of other marking workshops, both workshops produced a range of marks for each instance of student performance. In the Science marking workshop the ranges of marks for the two projects were 49% - 76% and 83% - 90%. The range of marks for the first project is noteworthy given the perception that science writing is more 'objective' and therefore less likely to produce a range of different responses from markers. In the Humanities workshop, the ranges of marks for the two essays were a 'fail' to a low 60 and a 'fail' to a high 60¹⁵. Despite the wide range of marks for the first essay, the assessors in the Science workshop were able to eventually come to an agreed-upon mark. In the Humanities workshop the assessors could only agree to pass both of the essays but no resolution was reached in relation to whether the essays should be awarded a high or low 50.

4.4.1.4 Discussion

The goal of this study was to explore what marking workshops might reveal about:

- The particular qualities which academics are looking for in student writing
- How the explicit criteria are used to assess students' performance
- Additional implicit criteria (or other factors) which influence the assessment

In terms of the particular qualities academics look for in student writing, both workshops were characterized by very little contestation over the qualities or criteria for 'good' writing. One interpretation might be that there is broad agreement, for example, in the Science workshop, that conceptual understanding (i.e. understanding of the key concepts being assessed) was the most valued quality. Or that in the Humanities workshop there was general agreement that analysis or analytical ability was most important (i.e. the ability to perform the comparative analysis that was required by the

¹⁵ Assessors were not required to assign precise marks.

task). This interpretation needs to be treated with caution. For a variety of reasons workshop participants may have felt that to dispute the criteria was pointless, since, for example, in the Science workshop, the marking criteria had been set ahead of time. Thus while there was no contestation about the criteria, this lack of contestation cannot necessarily be read as agreement. A more valid picture of the criteria which academics value would need to be constructed from a wider data set: interviews with markers, observation of their interactions with students and their writing, and markers' written feedback.

In terms of how the criteria were being used to assess student performance, what was clear from both of the workshops is that despite the explicit criteria, there was variation in how they were interpreted and applied. For example, in the Science workshop there was quite a bit of debate about what conceptual understanding means and what constitutes evidence for conceptual understanding. In the Humanities workshop there was a similar discussion about what an explanation might mean and the course convener was asked to clarify.

In the Science workshop an interesting issue emerged which shed some light on the use of the explicit marking criteria. Having assessed the sample of writing using the marking memorandum, two of the markers admitted that they felt uncomfortable with the mark that had resulted. It was too high. They had therefore adjusted their marks downward to fit their overall impression or 'gut-feel' about the report. This would suggest that in addition to or instead of the explicit marking criteria, these markers bring intuitive judgments to the assessment of student performance, intuitions constituted by, among other things, years of experience as professionals in a particular field of study.

With respect to what additional implicit criteria (or other factors) are influencing the assessment, a number of factors came up in the discussion which appeared to be influencing the markers' assessments. In both workshops markers made inferences about the degree of "effort" which the students had put into the writing. One marker in the Humanities workshop argued that, despite the fact that the essay "missed the point", she

would not fail it because the student had “worked too hard” (field notes, 21/5/97). Since the samples were anonymous these were inferences from the text and not inferences based on direct observations of the student performance. Another factor which came up in both workshops is how markers draw on information from the assessment context to influence their assessment. For example, one of the markers argued that he would be less generous in his mark allocation on the basis of his perception that the students had received assistance with the reports in the form of feedback on drafts. Yet another factor which emerged as an influence was students’ performances relative to each other. One marker in the Humanities workshop argued that he found it very difficult to assess these two samples in isolation from other writing samples.

The exploratory study also suggests that markers’ assessments are influenced by different interpretive matrices which markers hold. These matrices seem to be in part constituted by different resources which assessors have access to, including different kinds of knowledge, for example, expertise in the area of specialization, direct experience of the learning and teaching context, and involvement with the students.

In the Humanities workshop, the course convener was responsible for designing the topic which was in his area of specialization. In contrast to the tutors (one of whom admitted that she had never read in this area), he had published extensively in this subject. He was granted authority by the tutors to clarify what ‘explanation’ meant in the context of this topic, and to provide general guidelines about what tutors could expect from a ‘good’ explanation. In contrast to the tutors however, he had had little direct contact with the students and the immediate learning environment of the tutorial, nor had he marked any of the essays. Thus when it came to assigning a mark, it was the tutors, not the course convener, who haggled with conviction about what could actually be expected from the students in relation to the difficulty of the topic, the challenge of the readings, and the level of assistance the students had received. There was a similar dynamic in the Science workshop, where again the course convener had designed the topic around his area of specialization. He argued vociferously for a particular interpretation of the key concepts which students had to tackle. However, two of the

other workshop participants had consulted with about half the students on drafts. Thus to their interpretation of the criteria and application to the samples, they brought knowledge of the kinds of problems which students were experiencing with the reports. It seemed in both workshops that underlying assessors' different interpretations was access to different kinds of knowledge. These different forms of knowledge seemed to both empower and disempower individuals within the context of the marking workshop; some assessors were silent, while others argued with conviction.

4.4.1.5 Questions for further exploration

This study revealed a number of areas for further exploration. These include a multitude of factors which shape assessors' 'interpretive matrices':

- The relationship of the marker to the assigned task – Has the marker been involved in designing the task? What is the relationship between the disciplinary expertise of the marker and that covered by the task?
- The marker's relationship to the student being assessed – Does the marker know the student whose writing is being assessed? How invested is the marker in the student's performance?
- Interpersonal relations between the markers – Who are the markers? What individual resources do they bring to the marking exercise? What are the power relations between them?
- The degree of shared understanding between the markers – Do the markers hold similar views about learning, teaching and the role of assessment? Are there deep ideological differences between markers in relation to their discipline/s?
- Departmental differences – How are power and authority distributed within different departmental roles?
- Disciplinary differences – How do epistemological approaches within a discipline impact on markers' approaches to assessment?
- Institutional culture – How do the norms, values and interests of the institution impact on markers' assessment judgments?

4.4.1.6 Methodological insights from exploratory study

This preliminary study gave me the opportunity to grapple with some of the methodological complexities of researching academic validation practices. Although the marking workshop discourse revealed traces of other texts which influence markers' assessment, they remained only traces and could not be substantiated on the basis of the marking workshop data alone. There were also inevitably silences, for example, common experiences and prevailing beliefs that these collegial communities drew on which were not apparent to me. I concluded that an interpretation of 'text in context' would necessitate a far more ethnographic style of research. An analysis of a marking workshop, while affording me a view of the contestation, did not give me insight into the motivations and interests which were informing the various markers' individual assessments. For a deeper analysis of this contested terrain, I needed to extend my inquiry beyond just the marking workshop to include the broader context in which this event was situated, for example, the community of practice, the networks of practices and the field. This would necessitate also expanding my methods of data collection to include a range of qualitative methods that would enable me to more fully develop the context: more participant observation, interviews with not only those who participated in the moderation events but others invested in the assessment practices, and observation of other sites of assessment practice outside the marking workshop.

As a result of this exploratory study I become less interested in the criteria, either explicit or implicit, which academic assessors brought to bear on their interpretations of student performance; instead I became interested in assessors' interpretive matrices, that is, the whole complex of contingencies which influence their interpretations. My study shifted from the search for criteria to an exploration of the environment or matrix out of which interpretations emerge.

4.4.2 Main study

4.4.2.1 Institutional profile

SAU, the university where my main study was located, can be characterized as a medium-sized, contact university. It is a historically white, English-speaking institution,

located in one of South Africa's large urban areas (more about SAU in 5.2). It is one of South Africa's leading universities in terms of both the quality of teaching and learning and its research output. Within this institution my case studies focus on two courses: a Humanities final year undergraduate research methodology course (henceforth referred to as Hum300) and an Engineering final year thesis course (henceforth referred to as Eng400). I now turn to a description of these case studies.

4.4.2.2 Humanities Case study

4.4.2.2.1 Rationale for the case selection

There were several reasons for my selection of the Hum300 site for further exploration of the issues which arose from the preliminary study. Firstly, I felt that the study of final year undergraduate projects was an important assessment site. As a final culminating performance for the undergraduate qualification, this particular assessment opportunity seemed most likely to generate dialogue about the kinds of graduate performances valued by this community. The stakes are relatively high as decisions are being made which affect students' prospects for graduation, as well as their admission to postgraduate studies. In addition students are conducting research projects in which the members of staff have, to varying degrees, vested interests.

Secondly, the assessment of the final year projects involves most of the members of the department. This is crucial since I was interested not only in individuals' validation practices, but also individuals as members of a community. Thirdly, in addition to diversity in terms of academic rank, race, gender and ethnicity, this particular community also reflected an interesting diversity in terms of experience. In the year of my study there were two new, young members of staff. I felt that new staff as apprentices to academic life are likely to be catalysts for dialogue, raising issues which have become naturalized and fossilized forms of practice for more experienced members of staff. Fourthly, I discovered once I commenced the research that two years prior to my study the department had gone through a painful and protracted inquiry into accusations of racism which were directly related to assessment issues. From a researcher's point of view, this event served as a 'moment of crisis' (Fairclough 1992), an event which serves to disrupt

and de-naturalize practices. Finally, there was an openness and interest to my study by both the head of department (HOD) and the course convenor. When I approached the HOD about the study he acknowledged that, “the course had been running a long time and could benefit from some reflexivity” (field notes, 13/2/98).

4.4.2.2.2 Data Collection

While the staff was supportive, the HOD did however express some wariness about my study. He noted that it would likely “open a psychological can of worms” and suggested jokingly that it might be necessary to place an “embargo” on my thesis as far as students were concerned (field notes, 13/2/98). Despite these sensitivities he did not feel it was necessary for my project to be approved at the level of the Faculty Ethics committee¹⁶ since I was not “using” students (field notes, 13/2/98). He requested that I get permission from the Hum300 team to conduct the research; once permission was granted individual members of staff could choose whether or not they wanted to participate. He also warned me that staff would be under great pressure during the assessment period and this might make it difficult for me to collect the kind data I wanted. I took this as a concern on his part that busy academics might not appreciate extra demands at highly pressured times of the year. This same issue came up when I negotiated access to the Engineering site. I was thus, throughout my study, extremely sensitive to methods of data collection which might be perceived by staff as intrusive or a waste of their time.

My methodological approach (discussed above) required various stages of analysis, that is, the analysis of social structure, conjuncture, event and text, each of which entailed different methods of data collection. For the analysis of social structure I relied extensively on secondary sources, for example, the work of Burton Clark, a leading higher education researcher. For the analysis of conjuncture, I relied on participant

¹⁶ During the entire time of my study I struggled to establish what the procedure was for having the ethics of one’s research approved at SAU. When I started my research, SAU’s ethics committees were faculty-based; they were then centralized as a university-wide committee, and then again de-centralized to faculty level. The procedure was never clear to me even by the date of my last inquiry (11/01). Thus, with the guidance of my supervisors, I followed a procedure whereby I got permission from the HOD of each department, agreed in each case to protect the anonymity of my participants to the extent that I could, and to make available to all participants in the research a draft of the analysis for their feedback.

observation, interviews and documentation. For the analysis of the event and text I used a combination of interviews and textual analysis for the former and textual analysis for the later. I briefly touch on each of these methods in more detail.

With respect to participant observation, in my first year of data collection it was important that I experience as much of the course as possible in order to understand the immediate context in which the assessment of final year projects took place. I attended all course staff meetings, all the main lectures, two small group sessions and the final departmental marks meeting; I observed and tape-recorded the marking workshop, and engaged in numerous informal discussions with members of this community.

My interviews were semi-structured; I had a set of questions which I had prepared in advance but in the interview I allowed the conversation to flow in a guided but not too structured manner. Through experience I learned to relax and allow interviewees to go off on 'tangents' which often became some of the richest sources of data.

Out of a total of eleven members of staff involved in the course across the two years of my study, I was able to interview eight of them, most of them twice. The course convener in 1998 became one of my key informants, and although he was not involved in the course in 1999 (due to sabbatical leave) he continued to play a valuable role in giving me feedback during the early stages of data analysis. In the first year of the study, I conducted interviews with the course convenor and the HOD – both of whom had been involved with Hum300 for over a decade – to gain general background information about the course, its purposes, and its role within the department, the faculty and the institution. My interviews with the option convenors¹⁷ focused on general background issues (e.g. their career histories, their research specializations, how long they had been teaching on the course) and what their feedback revealed about the kinds of qualities they were looking for in student performance. I also interviewed a member of staff who had been at

¹⁷ 'Option conveners' refers to the academic staff who were responsible for groups of students. An 'option' was an area of interest which students selected for their research projects.

the centre of the departmental inquiry into alleged racist assessment practices¹⁸. I was never able to interview the external examiners despite several attempts.

In the second year of the study (1999) I focused in the interviews on the kinds of processes that staff go through to derive their mark, and the kinds of decisions which they face in the actual assessment process. These interviews were conducted right after the staff had finished the marking of the research proposals. I asked them to bring to the interview a couple of proposals which had raised interesting assessment issues/questions/dilemmas for them. These 'stories' became the focus of the interview. I found 'stories' about particular assessment cases to be a particularly effective means of gaining insight into assessors' interpretive processes.

The marking workshop comprised one of my central sites of data collection. Out of the total of nine staff involved in the course in 1998, seven attended the marking workshop (plus a visiting scholar and me). Although Hum300 did not have a tradition of marking workshops, staff members were familiar with them; many staff had participated in marking workshops at the junior undergraduate levels where tutors are involved in the assessing of essays. At my suggestion, the course convener agreed to organize such a workshop, although he admitted to me later that he had been hesitant because it would be "dangerous" as it would open up all kinds of "cans of worms" (Dan 99-1).

With respect to documentation, I collected:

- File of correspondence regarding the 1996 postgraduate dispute
- Course handouts
- Photocopies of a sample of research reports with markers' feedback to students
- Course evaluations
- Photocopies of examination papers and answers
- Reports from external examiners on the examination
- Spreadsheets of marks for final marks' meeting

¹⁸ This case was fully investigated and the member of staff involved was completely cleared of the accusations.

- Documents discussing the future of the course within a re-structured undergraduate programme

4.4.2.2.3 Data analysis

My analysis of the interview data was on-going and iterative throughout the research process as I forced myself to write-up throughout the initial stages of data-collection. Once I had conducted the Engineering case study and had a clearer sense of what I was looking for, my analysis of the Humanities interviews became more systematic (I discuss this in 4.4.3.3). I used NVivo¹⁹ for initial coding of the interviews for major issues. Out of these issues I looked for emerging patterns, and these emerging patterns become the major strands of my argument. For the analysis of the focal text (i.e. an excerpt from the marking workshop) I did a topic analysis on assessors' validating arguments.

My analysis process had various feedback loops. In 1999 I gave the Humanities course convener an earlier version of the analysis. His main critique was that my 'reading' of the marking workshop had missed some of the complexity of the social interaction. He was specifically concerned that my analysis of the marking workshop had failed to take account of the social dynamics of the meeting, in particular the issue of collegiality. While staff may profoundly disagree with another colleague's assessment, for a variety of different reasons it is not in their interest to publicly challenge their interpretations. This feedback sensitized me in a very real way to the dialectical relationship between discourse and interpersonal relations, and while these collegial relations per se were not the focus of my study, I could not ignore the effects of power and hope to do justice to interpreting the event. In 2002 I sent the HOD and the course convener a penultimate draft of the analysis. Once again their critical feedback served to validate and strengthen my argument.

¹⁹ NVivo is a qualitative data analysis software which enables the researcher to link a set of documents (e.g. interview transcripts) with an index system. An index system is a set of hierarchically linked nodes (e.g. concepts/issues/key words). The software allows the researcher to electronically pull out data from the interviews which have been coded under the same node.

4.4.2.3 Engineering case study

4.4.2.3.1 Rationale for the case selection

Toward the end of the Humanities study, I felt that it would be really important to explore the issues which had emerged from this case in the context of another faculty in the same institution as my first case study. In my search for another case study, I was informed about some particular problems that an Engineering department had experienced during its end of year examination meeting, problems which echoed some of the issues which had arisen in the Humanities marking workshop. I decided to explore the possibility of this site for my second case study. I concluded on the basis of an exploratory conversation with a member of staff from this department that this would be an ideal second study. There were some important similarities and differences between the two cases.

Like the Humanities department, the Engineering department also has a culminating research project in their final undergraduate year. The project is also high stakes; it is a requirement for graduation and thus the consequences of failure are serious. Again as in Humanities, the supervision and assessment of the final year project is the responsibility of the whole community, not simply a few individuals. This community also represented an interesting range of diversity in terms of academic status (from lecturer to professorial level), disciplinary specialisation, as well as nationality, race and gender. As in Humanities, several young members of staff had recently joined the department.

In terms of their differences, as in the preliminary study, I felt it was important to conduct research in a contrasting 'hard' science discipline. This enabled me to explore how the different approaches to knowledge production influence assessment practices. Like Humanities, the Engineering community did not have a tradition of marking workshops, nor were they an established part of the Engineering departmental culture. Given the highly specialized nature of the research, academics within one area of specialization would consider it inappropriate to assess projects from another area. All members of the department did however come together to discuss student performance at a departmental examination meeting. At this event all members of the department came

together with the external examiners to discuss discrepancies between marks which had been allocated to the projects by a first, second and sometimes third marker. This seemed an appropriate event for the exploration of validation practices.

4.4.2.3.2 Data collection

I met with the HOD in early 2000 to request his permission to study the assessment of final year projects. He was very receptive to my study, but noted two conditions. Like the Humanities HOD, one major concern was that members of staff should not feel imposed upon; to prevent this he asked that I request for volunteers. Secondly, he stressed the importance of confidentiality so that no member of staff's reputation would be adversely affected by my study. I was able to ensure him that all information would be treated confidentially; we also discussed some of the difficulties of assuring complete anonymity.

My data collection was divided into three stages in Engineering. The first stage was prior to the assessment of the final year project (August 2000 to mid-October 2000), the second stage was the actual final year project assessment process (mid-October 2000 to mid-November 2000), and the third stage was following the assessment (mid-November 2000 to November 2001).

During the first stage of data collection I conducted two rounds of interviews. The first round (with the course convener, the current HOD, and the ex-HOD who was responsible for setting up the final year project process) was primarily to establish the appropriateness of this site. In addition to getting background information about the course and its context, I was particularly concerned about the extent to which the department examination meeting would be the appropriate discursive event. Having determined that it would be, I conducted a second round of interviews with the wider staff. Initially I thought I would sample from the staff group in order to narrow down the number of interviews. Both the convener and the HOD advised me against this. The convener argued that it would be difficult to get a take on the department's practices in any representative way since each person operated individually. Therefore I sent out an

interview request to all staff and ended up interviewing approximately three-quarters of them. In addition to general background information (e.g. career history, involvement with the course, disciplinary specialization, their own particular supervision and assessment process), I probed for their views concerning some of the contentious issues which had arisen in the previous year's departmental examination meeting. In particular I was interested in what they believed to be some of the reasons for the wide discrepancies which sometimes arose between first and second markers.

During the second stage of analysis, I conducted the third round of interviews (with a slightly smaller group of ten) immediately after they had completed their marking of the projects and before the departmental examination meeting. It was important for me to capture their assessment processes while they were very much 'live' issues. As in the case of Humanities, I asked them to note projects which were particularly challenging or interesting. In the interview I probed those experiences, asking them to articulate their perspective as first marker, second marker, and the kinds of qualities they looked for and what the mark meant. I also observed all the major, formal assessment events which happened during that week: the Open Day Oral presentations, the consensus meetings, and the departmental examination meeting. (I describe these events in more detail in 5.4.1).

In the final stage of analysis, after the departmental examination meeting, I conducted follow-up interviews with the ex-HOD and the faculty's educational development officer, who had also attended the examination meeting. I was also able to interview one of the external examiners. I collected all the e-mail discussion in the aftermath of the examination meeting and attended a departmental staff meeting to discuss a proposal concerning the role of external examiners in the future. I then followed the progress of the committee set up to review the final year project assessment procedures through interviews with the educational development officer and the course convener. Finally I attended and tape recorded the 2001 departmental examination meeting in order to note consistencies or inconsistencies between the discursive patterns of the meetings. Unfortunately that recording was not successful for technical reasons.

4.4.2.3.3 Data analysis

As with Humanities, the analysis of the data consisted of two major forms: an analysis of the major themes which emerged from the interviews and discourse analysis of excerpts of the departmental examination meeting. Again the process of collecting and analyzing was iterative. During the first round of interviews I probed interviewees' own theories or 'reasons' for why discrepancies arose between marks. These 'reasons' became my initial set of nodes (in NVivo) which I used to sift through my data. The 'reasons' which appeared most consistently across interviews were the ones which I pursued in further interviews. One frequently noted 'reason' was the different 'perspective' between the first marker (supervisor) and the second marker. This became a dominant theme which I pursued in a more focused way in the second round of interviews. I then created a 'tree of nodes' with 'assessor perspective' as the 'parent node' and 'supervisor', 'second marker', 'third marker', 'external examiner' as 'child nodes'. Through 'memos' I was able to keep a running commentary on assessors' perspectives, and how this impacted on their interpretations. These analyses-in-progress in their very early, raw forms were tested in follow-up interviews so that part of my interpretive task was to feed back to my interviewees their own interpretations. I was then able to take this analytical frame to the analysis of the Humanities data.

The themes which arose from the content analysis served to focus my textual analysis. For example, I explored whether the validity arguments which arose in assessors' interview accounts were consistent or inconsistent with what the textual analysis revealed. In this way the content analysis and the textual analysis served as a form of triangulation for my findings. As in the Humanities analysis, I also explored academic validating arguments for the kinds of evidence which supported different interpretations. Since I was particularly interested in the facework of this meeting, this kind of analysis required a slightly more detailed transcription than the Humanities analysis, for example, attention to modal expressions, pauses, emphasis, intonation and simultaneous speech.

4.4.3 Summary

The task of this chapter has been to account for the 'construction of my object' which from Bourdieu's perspective is the most crucial research operation (Bourdieu and Wacquant 1992). My methodological approach identifies four elements which constitute social practice – social structure, conjuncture, event and text. These constitutive elements operationalize into a series of analytical stages which expose different aspects of social practice. In order to illumine these constitutive elements, I proposed a set of theoretical tools which expose not only the elements but the relations between them. I argued that this methodology is broadly consistent with Fairclough's method of CDA, with some adaptations which address some of the critiques which have been made of Fairclough's work. It draws on ethnographic methods to give access to other texts which are crucial to explanations of social practice, in particular language users' interpretations of their own texts. In the final section I reported on my exploratory study and then described the design of the main study.

I now turn in the next chapter to the analysis of the case study data.

Chapter 5: Analysis

For judgments and arguments of magic to be valid, they must have a principle that eludes examination... These principles, without which the judgments and arguments are not believed possible, are what philosophy calls categories. Constantly present in language, without necessarily being explicit, they exist rather in the form of guiding practices of consciousness, which are themselves unconscious.
(Mauss cited in Bourdieu 1996, p. 7)

5.1 Introduction

I noted in Chapter 4 that one of the initial questions which guided my preliminary study was: what particular qualities are academics looking for in student writing? I concluded that my preliminary study had not enabled me to explore this question in sufficient depth. While both marking workshops had explicit marking criteria (although the Humanities' workshop criteria emerged *post hoc*), it was difficult in the context of the marking workshops to pursue the extent to which there was broad agreement among the staff around these criteria. I also reported on two assessors in the Science marking workshop, who, after applying the explicit criteria, had adjusted the overall mark to fit their impression or 'gut-feel' about the report. I argued that this suggests that in addition to or instead of the explicit marking criteria, assessors bring intuitive judgments to the assessment of student performance, intuitions constituted by, among other things, years of experience as professionals in a particular field of study.

Eliciting from academic staff the criteria which they bring to bear in the assessment of student performance proved to be one of the most challenging tasks I faced in the main study. I recall one particularly difficult interview near the beginning of my study. The Hum300 course convener and I sat with two exam scripts where essentially the same question (about research methodology) had been answered by students at two different levels; one was a final year undergraduate student and the other was a first year postgraduate student. The convener was clear that, even though the question was the same, he would be looking for "qualitatively different" kinds of responses from these two students. But during the course of the interview, I, who had hoped to get clear statements about these qualitative differences, was to be disappointed. The convener's responses defied neat statements of criteria. He finally admitted the difficulty he experienced in articulating what was for him essentially "the concept of a good scholar", which he

argued is highly “intuitive”. He compared his task as an assessor to those who judge dance or sport. When they recognize talent or skill in the performers, he asked, “How do they know that? They know. They know when somebody’s really skilful. There’s no dispute.” Pointing to his marks, he added, “I think that’s what’s going on here” (Dan 98-2)¹.

What I discovered as I interviewed nearly thirty academic assessors from different ends of the disciplinary spectrum was that the appeal to intuition as the basis for academic judgment was not uncommon. From both the Humanities and the Engineering staff there seemed to be a great deal of ambivalence towards explicit marking criteria. In each of the communities of practice, there were members of staff who did use a marking memorandum very systematically, working through each of the criterion and allocating a certain percentage of points out of the total. They were the exceptions, however. Most of the staff expressed varying degrees of resistance towards explicit marking criteria. One member of staff in Humanities felt that criteria were important as signals to students of what was valued. He admitted, however, when it came to marking, “I’m not going to tie myself to (these criteria)². I can’t” (Jos 99-1). In Engineering where there was an official marking memorandum, some “disregarded it completely” (Car 00-1). One noted, “It means nothing to me” (Joh 00-1). Another admitted that he had tried to use it but did not find it particularly helpful (Ric 00-1). Another objected to the false compartmentalization which marking memoranda imposed; “I can’t do it,” he noted emphatically. “I don’t have a mind that splits into seven bits” (Rob 00-2). While some of the resistance was towards the specific official marking memorandum, my sense was that much of the resistance was to the very notion that final year projects could be assessed by the allocation of points against a fixed set of criteria.

¹ References to interview data consist of the first three letters of the interviewee’s pseudonym, the last two digits of the year of the interview and the number of the interview. Thus, ‘Dan 98-2’ refers to the second interview conducted with Dan in 1998. Other relevant information for each interviewee is provided in Appendix 1 for the Humanities case study and Appendix 3 for the Engineering case study.

² Where I have added a word or phrase to clarify the meaning of a quotation from an interview, I have put these additions in parentheses.

A common theme that emerged from both communities of practice was that assessment was something more akin to ‘taste’ (Geo 00-2)³. One member of staff noted, “A thesis is not something which you could judge quantitatively...it is a qualitative judgment” (Ken 00-1). Another member of the Engineering staff commented, “I don’t know if there is a recipe. I think in many ways it is your stamp of approval. It’s more like (judging) the arts...I think there is a tendency to say, ‘mmm, I really like this’, whereas another marker...because they are prejudiced against the argument, would mark it down a little bit” (Car 00-1). One member of staff assessing for the first time recalled, “The impression that we were given as new supervisors, is that, it’s left to academic feeling, that you decide how you want to mark” (Kur 00-2). Another commented on the difficulty he had in using the marking memorandum, “I tried to use it (the memorandum) (but) what I do now is I mark in a more intuitive way. I am not sure if intuitive is the right word, but I read it and I decide”. He noted that academics become “imbued” with a “form of thinking” that enables them to “put (theses) into categories” (Ric 00-1). Commenting on the difficulty of making such criteria explicit, a member of staff noted, “In time you find ways of assessing something that would take a lot of articulation to get it all out” (Joh 00-1).

Thus it would appear that experienced academics operate with powerful intuitive categories, and the rules or criteria which define these categories resist explicit definition. Bourdieu (Bourdieu and Wacquant 1992) distinguishes between the rules or criteria of, for example, a game of soccer and the rules which inform classificatory acts of social life, for example, assessment. The former are explicit, the latter are not. He argues that the latter may not even be conscious. In *State Nobility* Bourdieu starts a chapter entitled “Academic forms of classification” by citing Mauss on categories: “Constantly present in language, without necessarily being explicit, they exist rather in the form of guiding practices of consciousness, which are themselves unconscious” (1996, p. 7). Therefore, early in my study, I concluded that if I was to understand the logic of academic validating practices, I needed to move beyond the search for ‘rules’, that is, the search for explicit or

³ Piper’s (1994) survey of university examiners’ practice in the U.K. supports this same theme. He found “a heavy reliance on intuitive global judgments (or individually developed marking schemes) and a tacit understanding of marking standards” (p. 77).

even implicit criteria which inform academic judgment. With reference to these kinds of rules, Bourdieu (1990) notes, “(They) have a particularly small part to play in the determination of practice” (p. 145). Rather my task was to expose the “objectivity of the subjective” (p. 135), or as Postone et al (1993) explain it, “to capture the practical mastery that people have of their social situation, while grounding that mastery itself socially” (p. 4).

The aim of the first part of my analysis (5.2) is to develop, largely on the basis of secondary literature, some understanding of the social conditions of constraint and possibility which ensure the on-going production and re-production of academic validating practices. In Bourdieu’s terms this is an exploration of the ‘game’ (Bourdieu in Bourdieu and Wacquant 1992, p. 98) where the goal of the analysis is to sketch the contours of the field, to identify some of the stakes in the competition (i.e. the various forms of capital), and to expose some of the rules of fair or legitimate play which inform academics’ classificatory acts, what Bourdieu refers to as “the legitimate principles of legitimation” (Bourdieu 1996 p. 265).

The analysis is guided by the following questions:

- What are some of the distinguishing features of the higher education field?
- What are some of the particular forms of capital which define institutional and individual positions within the field?
- What are the epistemic rules shaping academic validating practices?

I conclude this section by illustrating the “logic of practice” (Bourdieu 1990) through the external examination system, one of the traditional and deeply entrenched validating practices of the academic field.

I then turn in the remainder of the chapter (5.3 and 5.4) to my primary interest which is how individuals ‘play the game’, that is, how they take-up (contest, negotiate, construct) the positions defined by the field. I explore this take-up in the context of two different academic communities, one situated in a Humanities faculty and the other in an Engineering faculty. Each case begins with an exploration of the actual community with a

particular focus on its specific “joint enterprise” (Wenger 1998, p. 77), the assessment of final year projects and the procedures, events, rules, roles and technologies which enable members to carry off the project assessment from year to year. Then I turn to a closer scrutiny of actual assessment discourse to explore the particular complexity of contingencies (what I call the interpretive matrix) which assessors bring to bear in the assessment of student performance. I am interested not only in the differences of interpretation which emerge as different interpretive matrices are brought to bear, but what is at stake in the negotiation of these differences and how these negotiations are resolved.

This analysis is guided by the following questions:

- Within the particular assessor roles to which academics are allocated, what are some of the dimensions of the interpretive matrices which they bring to bear on student performance?
- Do these interpretive matrices result in different interpretations?
- If so, what is at stake in the negotiation of these differences?

I now turn to an analysis of the higher education field.

5.2 Analysis of the higher education field

I have argued in Chapters 3 and 4 for a social theoretical and methodological approach to the study of academic validating practices. Here I briefly re-cap some of the main theoretical points which frame this analysis. I argued that academic validating practices only make sense within the logic of the field, that is, the field of higher education and the academic profession. The field of higher education is a configuration of positions which are defined by the distribution of capital. Thus institutions and individuals within the field are differentiated from one another on the basis of their particular accumulation of capital, capital which converts to a form of power. It is also the field which determines, what Bourdieu (1996) refers to as, the “legitimate principles of legitimation” (p. 265), that is, principles which shape and legitimate classificatory acts. Given the contestation over legitimacy, fields are always characterized by power

struggles, power struggles within the field and power struggles to define the field (Postone et al 1993).

I start by noting two particular features of the higher education field (5.2.1). I then turn to identify the particular forms of capital which count in this field and the manner in which this capital defines institutional and individual positions within the field as well as the epistemic rules which legitimate validating practices (5.2.2). In the next section (5.2.3) I explore the logic of the external examination system. Within each section I relate the theoretical analysis to the South African higher education system and focus on SAU in particular.

5.2.1 The academic field

One of the notable features of higher education as a field of institutions is that it is highly differentiated (Clark 1983, 1987a, 1987b). One of Clark's (1987a, 1987b) central arguments is that there is no single experience of the academic profession. The academic profession is profoundly shaped by national, institutional and disciplinary contexts. For Bourdieu the primary principle of this differentiation is the distribution of capital, that is, social, cultural and economic resources. These differentiated forms of capital form the basis of a hierarchy within the field which becomes most evident within a national system of institutions, with research universities at one end of the spectrum and community and vocational colleges at the other (Clark 1987a). One of the implications of this highly differentiated system is that the epistemic principles which inform academic validating practices are not homogeneous across the field; what is perceived to be legitimate knowledge, what are the legitimate criteria for this knowledge, and who can legitimately assess and validate this knowledge will vary across institutions and disciplines according to their position in the field⁴. This indeed becomes one of the sources of power struggle within the field.

⁴ Piper (1994) found interesting differences between the responses of academics in the university sector and those in the polytechnic and college sector to the question of who could be legitimately appointed as external examiner for their subjects. Very few academics in either sector felt that external examiners from outside their disciplinary group could legitimately examine their subjects (e.g. a psychologist moderating an architectural project). But in response to whether an external examiner from a different but related

In addition to institutional differentiation, there is also disciplinary differentiation. While the academic field determines the broad epistemic rules which shape validating practices, the specialist communities of practice also profoundly shape the particular epistemological orientation which academics adopt in their validation work (as the Engineering case study illustrates). Becher's study (1989, Becher and Trowler 2001) of 'academic tribes' argues that any study of academic social life needs to give sustained attention to how epistemological orientation shapes academic culture, that is, what academics believe, how they feel, how they behave and, I would add, how they assess.

A second notable feature of the higher education field is that, given significant shifts in the global political and economic landscape at the end of the 20th century, higher education is a field which is currently in the midst of turbulent change, so much so that Clark (1998) suggests that "no one knows with any degree of confidence what the 21st century holds in store for universities" (p. xiv). Some of these changes include: growth in both numbers and diversity of students demanding a tertiary education; a rapidly expanding professional labour market with increasing demands on universities for competent graduates; greater expectations of accountability from government but decreasing state subsidy; massive explosion and fragmentation of knowledge which places strain on the range of disciplinary and interdisciplinary fields which any university can offer (Becher and Trowler 2001, Clark 1998). This suggests that the epistemic principles of legitimacy informing validating practices will feel the winds of change and be pulled by the strain of competing discourses, and traditional forms of validation will be challenged by new demands. These pulls are power struggles to define the contours of the field. Again, depending on institutional and disciplinary positioning, validating practices will respond to these competing discourses in different ways.

The South African higher education system emerged in 1994 from its years of apartheid isolation to face two formidable challenges. On the one hand was the challenge

discipline could legitimately moderate their subject (e.g. a chemist moderating another physical science), 18% from the polytechnic sector agreed in contrast to only 6% of university academics.

of transforming what was pre-1994 a racially divided and un-coordinated system comprised of 36 higher education institutions controlled by eight different government departments, a system socially engineered to marginalize South Africa's black (specifically African) population and to privilege its white minority population (Bunting 2001)⁵. Thus post-1994 the emphasis of South Africa's initial phase of higher education policy was on equity and access, particularly for constituencies (students and staff) who had been disadvantaged under apartheid. One of the results of these social-political changes (the effects of which were felt long before 1994⁶) was an influx into the higher education system of many students who previously had been excluded on racial grounds. The impact of this influx was most strongly felt by the historically white institutions (initially the English-speaking, but also later the Afrikaans-speaking ones) where the diversity in levels of educational preparedness between students was most acute. Students who had benefited from the highest standard of schooling in the country sat side by side with peers who had been schooled under the worst educational, economic and political conditions. This influx of previously excluded students was felt most acutely by Humanities faculties. Alternative admissions policies made university entrance possible for students who under traditional admissions criteria would not have been admitted to higher education. However, these students were seldom admitted to the faculty of their first choice (usually the professional degrees) given these faculties' more stringent admissions requirements. Thus Humanities programmes in historically white universities in particular felt the strain of a rapidly changing student population. As I indicate in the case study analysis, this is a social condition of the field which strongly influences validating practices, particularly in the Humanities department.

In addition to its own local social-political transformation process, the other challenge for the higher education system in South Africa has been to 'find its feet' in a global transformation process. Becher and Trowler (2001) refer to this as higher education in a post-industrial environment, a period characterized by turbulent changes.

⁵ In this study, unless otherwise specified, the racial designation 'black' includes groups who under apartheid were designated 'African', 'coloured', and 'Indian'. 'White' refers to those of European descent.

⁶ As early as the 1980s historically white English-speaking universities in defiance of the apartheid government were admitting students classified as 'African', 'coloured' and 'Indian'.

Some of these changes have already been noted, for example, massification, marketization, managerialism (see Becher and Trowler 2001). One of the significant changes in higher education globally has been an altered relationship between universities, the state and wider society. In contrast to the past where higher education has enjoyed tremendous autonomy, the current climate is characterized by less trust on the part of the state and the wider public that higher education is fulfilling its role in society. Thus one of the features of this current period is increasing calls for higher education to be more accountable to the state and to the wider society for all aspects of its practice, namely, its research, teaching, administration and community outreach.

In relation to the quality of its teaching, I argued in Chapter 2 that assessment has been perceived by policy makers to be an essential technology for both the reform of curriculum and its quality assurance. I noted the emergence in the 1990s of national qualification frameworks (NQFs) in the United Kingdom, Australia, New Zealand and South Africa. Some of the key characteristics of NQFs include a national qualifications authority, a specified number of levels or standards of learning, definitions of qualifications by level, learning outcomes, assessment criteria and accreditation of prior learning (D'Andrea et al 2001). The purposes of these NQFs vary across contexts. In South Africa the purported purpose has been to address fundamental inequalities in provision of and access to educational opportunities. While the tertiary sector has been in principle supportive of the broad political agenda underlying the NQF, there has been tremendous resistance⁷ to the implementation of the framework, particularly by the university sector⁸.

These local and global challenges pose a set of competing discourses for the South African higher education system where on the one hand, the transformation agenda at

⁷ See D'Andrea et al (2001) and Kraak and Mohamed (2001) for a discussion of university resistance to NQFs. D'Andrea et al (2001) argue that university resistance to NQFs closely resembles the resistance that the academic community has towards many forms of new 'managerialism' in higher education, which are perceived to threaten traditional academic values. Kraak and Mohamed (2001) note that one of the primary reasons for universities' resistance to the NQF is the perception that NQFs privilege vocational qualifications over the formative emphasis of university qualifications.

⁸ This resistance by universities is not unique to South Africa. Kraak and Mohamed (2001) note that New Zealand's NQF, after which South Africa's policy is modelled, has been unable to get buy-in from its university sector which has to date remained independent of the framework.

national level prizes issues of equity, access and re-dress of both institutions and individuals. On the other hand, in order to survive in a globally competitive system, the global agenda prizes quality, efficiency, and excellence. In the face of these potentially competing challenges, Maassen and Cloete (2002, p. 32) argue that higher education national policy-makers are likely to privilege issues that 'fit' globalization discourse, such as efficiency, effectiveness and competition. Specific national issues, for example in South Africa such issues as institutional and individual redress, are more likely to be marginalized. Thus it could be argued that South African higher education institutions, at the level of policy and practice, are negotiating their way through these competing discourses. I shall point to traces of these competing discourses in the case study analysis.

5.2.2. Academic capital, positions and rules

Bourdieu argues that, "People are at once founded and legitimized to enter the field by their possessing a definite configuration of properties. One of the goals of research is to identify what these active properties are...that is, these forms of specific capital" (Bourdieu in Bourdieu and Wacquant 1992, p. 107). In this section, I explore some of these forms of capital or values of the academic profession, and how these values, in turn, define institutional and individual positions within the field, as well as how these values shape the epistemic principles of vision and division which constitute academic validating practices, specifically academic professional judgment.

One of the core values, if not the core value of the academic profession is the production of knowledge. On the basis of his interviews with academics across a wide variety of institutional types (from top-ranking universities to community colleges) Clark (1987a) notes that "knowledge – creating it, caring for it, teaching or otherwise transmitting it – proved to be the nearest thing to common ideological ground" (p. 130). When academics think of their service to society, the provision of knowledge is front and centre (Clark 1987a). This pursuit of knowledge constitutes the "inner logic" of academic practice, its built-in gyroscope, what preserves its identity (Ashby cited in Clark 1987a, p. 268).

Accumulation of knowledge (in the form of research) thus is one of the primary bases of institutional and individual differentiation within the field of higher education. Institutions which have accumulated large amounts of specialized knowledge are well-positioned within the “powerful hierarchy of prestige” (Clark 1987a, p. 58). The more highly specialized the knowledge, the higher the qualification, the more the prestige; this prestigious position is occupied by what Clark (1987a) refers to as research universities (although even within this prestigious category there are many differentiations). Conversely, the less specialized the research, the lower the institution sits in the prestige hierarchy, for example, technikons and vocational training colleges. Individual academics’ positioning within the field is also inseparably intertwined with their role as disciplinary specialists (Becher and Trowler 2001, Clark 1987a and Henkel 2000). Not only is disciplinary expertise the primary basis of professional identity but this form of capital converts to a form of power. Disciplinary expertise is thus also one of the primary bases of academic authority (Clark 1987a).

Another core value of the academic profession, which is intrinsically linked to disciplinary expertise, is professional autonomy. Again this value is notably strong in research universities; the higher the institution sits in the prestige hierarchy the more tenaciously the rights of self-rule are guarded. Both Bourdieu (1988) and Clark (1987a) emphasize the historical link between specialized knowledge and academic autonomy. Bourdieu (1988) notes how the autonomy of the university field grew constantly through the nineteenth century. The “professor of higher education” evolves from being a “dignitary appointed by political authorities and committed to politics” to becoming a “select and specialized teacher...animated by a specifically academic ideal” (p. 37). With reference to the rise of ‘academic guilds’ in the United States, Clark (1987a) also links rising academic autonomy with academics’ “gathering control over knowledge brought about by research-driven specialization” (p. 16).

The notion of academic autonomy or self-governance has two levels of application. The first is that of individual professional autonomy, that is, the right of an academic to determine the precise nature of their work, for example, what they research, what they

teach, and how they assess. The second is the autonomy of the academic community to rule itself with relatively few impositions from outside, especially the state. A profession which prizes professional autonomy, based on specialist expertise, will elect governance structures which best protect these values. These structures have been traditionally “bottom-heavy” (Clark 1987b, p. 381), often referred to as collegial. The higher the standing of the university the more “authority rolls around loosely” (p. 152). Ambiguity in lines of authority is seen as functional. The effect of this form of governance is that academics wield a fair amount of personal control over their research and teaching with, as Clark argues, “their judgments...only minimally circumscribed by bureaucratic rules or collegial norms that would foreclose individual discretion” (ibid.).

Therefore, these forms of capital, knowledge and autonomy serve to explain the hierarchy of prestige (or the “space of play”, Bourdieu in Bourdieu and Wacquant 1992, p. 102) within the field of higher education, that is, that on the basis of their accumulation of these forms of capital, institutions and individuals are positioned within the field either advantageously or disadvantageously. Not only does capital define positioning but it defines the principles which shape and legitimate classificatory acts within the field. The logic of academic professional judgment or ‘taste’ is rooted in knowledge or specialist expertise. The question then follows: if specialist expertise is the legitimate basis of taste, then who will judge these expert tastes? Given the field’s investment in autonomy, the validity of professional taste can only be assessed by other specialists, by peers. Only someone with comparable expertise can legitimately assess and validate the assessments of the academic as subject specialist (Becher and Trowler 2001). These principles of legitimacy, namely knowledge-based expertise and peer-validation, serve to explain the logic of academic validation practices for institutions which are well-positioned or aspire to be well-positioned within the field of higher education. They also explain resistance from the field to any perceived threats to these traditional validation practices.

In relation to this distribution of capital, where is SAU positioned? Despite an era (post-1994) of higher education policies aimed at transforming the system, the pyramid of institutional differentiation (and hence prestige) inherent in the apartheid design has

remained relatively intact, with the historically white English and Afrikaans universities controlling a vast proportion of economic and cultural capital. SAU is among these favourably positioned institutions. It is considered to be one of the leading universities in South Africa, with some of the highest accumulation of cultural capital (measured in terms of research output, highly rated scientists, top-ranking undergraduate and postgraduate students) and economic capital (measured in terms of widely diversified funding base). From this position of power, Cloete and Maassen (2002) argue that institutions like SAU have been able to resist reforms that run counter to their institutional culture. They argue that while a few compromises have been made in terms of de-racializing the student body and management profile, these institutions have to a large extent remained as they were before (p. 468).

This is well-illustrated by SAU's responses to increased demands from the state for accountability for its educational programmes. I noted earlier the widespread resistance from universities to the implementation of NQFs. During the early phase of my study, SAU experienced in 1996-1998 an institutional programme re-structuring exercise, and then in 2000 a concerted attempt on the part of the State to impose external quality measures through an interim registration of qualifications on the NQF. Both exercises required academic programmes to be more explicit about programme purposes, learning outcomes, forms of assessment, and assessment criteria. Consistent with the international experience, the general response from SAU academics was not receptive. I would argue that the contestation between universities and the NQF is a struggle over the "legitimate principles of legitimation" (Bourdieu 1996, p. 265); NQFs are perceived by universities as a threat to their traditional forms of academic capital. They also fundamentally challenge the epistemic rules which govern academic validating practices, namely, that final authority rests within the disciplinary sub-field and can only be validated by other members of this field. The issue of who has final authority is a significant theme in the case study analyses which follow.

I now turn to brief discussion of this logic in practice.

5.2.3 Academic practice: the logic of external examination system

These principles of legitimacy serve to explain the logic of many academic classificatory acts (e.g. editorial boards, selection and promotion committees) and specifically the curious phenomenon that for most of their history universities have guaranteed the quality of their own work. In relation to the quality of their degrees, one of the primary validation mechanisms used in the British higher education system and many of its ex-colonies (including South Africa) is the external examination system. While often the subject of academic scepticism, the external examination system remains deeply entrenched as one of the central validation mechanisms. Piper (1994) argues that the convention of external examiners is still commonly regarded as the principal guarantee that British university degrees are of a “consistent and high standard” (p. 1).

From the early history of British universities through to the present, Piper (1994) argues that the most commonly recognized functions of the external examiners are firstly, to judge whether the standard of award is on par with other awards in the country, and secondly to ensure that students are treated fairly (p. 107). Although there has been no systematic study of the external examination system in South Africa⁹, there is some evidence from my study that there is, as in the United Kingdom, the perception of a strong link between external examiners and the maintenance of standards. SAU’s policy manual endorses this link: “The university has a responsibility to the professions in particular and to the public at large to ensure that its students attain the required professional and academic standards. These can best be maintained through the services of examiners external to the university” (University policy manual 5.4.1.1). One of my interviewees in Engineering argued that the role of externals is strongly associated with the question of standards. There is a “popular opinion” that what maintains standards is the presence of an “independent arbiter from the outside”; that “anything internal is open to corruption” (Ala 00-2).

⁹ Since the rise of accountability culture in South Africa there has been some critical reflection on the role of external examiners, particularly in relation to their potential role in programme review (see Higher Education Quality Council’s Improving Teaching & Learning Project: Interim Guides to Good Practice 2002). To my knowledge there has been no systematic investigation of South Africa’s external examination system. There is work-in-progress being conducted by Lee Sutherland of the University of Zululand on the role of moderation in higher education.

What is the basis of external examiner authority? The principles of the field determine that their basis of legitimacy/authority lies in their disciplinary expertise. Piper's (1994) survey found that in the United Kingdom the primary reason for the selection of a particular external examiner was his/her eminence in the disciplinary field. The second most frequently cited reason was that the external examiner was personally known to the internal examiners. Thus from the internal examiners' perspective the external examiner is 'one of us'. But a second basis of their authority, I would suggest, lies in their alleged impartiality as 'external' to the particular context, in other words, they are independent of the university and thus perceived to have no invested interest in the students' performances. This principle of independence from the university concerned is stressed in SAU's policy manual; only in rare exceptions can an external examiner be appointed who is internal to the university and in the case of doctorates (where the stakes are highest) an external must be independent of the university (University policy manual 5.4.1.3).

This prevalent belief in the privileged status of an external examiner's professional judgment remains largely unchallenged until a disagreement arises. In the event of a disagreement between an internal and external examiner who has the final say? What is the 'official' position? In fact, at SAU there is no 'official' position at university, faculty or departmental level. The authority of the external examiner appears to be a matter of case by case interpretation. Similarly Piper's study (1994) found that for British universities there was no national regulation on this matter; the position varied from institution to institution. In his study both internal and external examiners were asked, 'Could an external overrule an internal examiner's mark?' The response from internal examiners was: 37% agreed that externals could overrule, 37% argued they could not and 26% offered no reply. The response from external examiners: 64% of respondents agreed that externals could overrule. Piper (1994) notes, however, that it was clear from interviews that external examiners seldom had to evoke these powers since differences were resolved by "mutual adjustment" (p. 211). Piper's study suggests that the very founding assumptions of the external examination system lie in a powerful belief in the

privileged status of the external perspective. However, when it comes to conceding final authority, the academic community appears to be divided, reluctant to relinquish their right of autonomy to the 'external', albeit 'one of us'.

5.2.4 Conclusion

The aim of this analysis was to develop some understanding of the social conditions of constraint and possibility which ensure the on-going production and re-production of academic validating practices. In Bourdieu's terms this involves sketching out the contours of the field, identifying the forms of capital, demarcating positions and exposing some of the principles of vision and division which inform academics' classificatory acts. I argued that the higher education field is highly differentiated so that the epistemic principles which inform academic validation systems are not homogeneous across the field. Higher education is also a field in the midst of turbulent changes which suggests that validating practices will feel the strain of pull from competing discourses. From Bourdieu's perspective, positions within this differentiated field are based on access to forms of capital, and I have argued that for higher education these are principally knowledge and autonomy. Thus institutions which have accumulated vast amounts of capital (measured, for example, through scientific ratings, publication counts, and research funding) are advantageously positioned in the hierarchy of institutional prestige. The same applies to individual positioning within the field; academics' positioning (or professional identity) within the field is, to a large extent, determined on the basis of their accumulation of specialist-based expertise. Furthermore, since capital is a form of power, it is largely on the basis of this expertise, that professional authority is determined.

Not only does capital define institutional and individual positioning, but it also specifies the "principles of vision and division", that is, "the legitimate principles of legitimation" (Bourdieu 1996, p. 265). These are the principles which inform academic classificatory acts: what is perceived to be legitimate knowledge, the legitimate criteria for assessing this knowledge, the legitimate assessors of this knowledge, and the legitimate validators of these assessments. Fields are inevitably sites of struggle, although the degree of contestation within and from outside the field over these positions

and principles will vary. Despite the turbulent changes occurring in higher education more globally and in South African higher education locally, I have argued that because of its relatively privileged position in the higher education hierarchy, SAU has been able to determinedly resist calls for change. In particular, it has been able, to date, to resist calls for increasing accountability which challenge the legitimacy of traditional modes of validation practice. An example of the latter is the external examination system, a deeply entrenched academic social practice which makes best sense within the logic of the higher education field.

Trowler (2001) argues that universities are dialogical entities, “composed of a multiplicity of discourses with plurivocal meanings” (p. 191). He suggests two sources of this plurivocality: first, there are the conditioning structures within and beyond the university itself. These conditioning structures have been the focus of part one of my analysis. But, secondly, there is the operation of the activity system, that is, how individuals and communities contest, negotiate, and re-construct these conditioning structures. He argues that is it within the communities that “recurrent practices become shaped and embedded, where developing meanings are constructed as individuals work together on the issues of professional life” (p. 194).

It is to these academic communities which I now turn.

5.3 Humanities Case Study

5.3.1 Validating practices

My first case study focuses on the community of practice whose enterprise is the assessment of final year projects in one of SAU’s Humanities departments (henceforth referred to as the Humanities department). This community involves most of the academic staff of the department (nine members of staff in both 1998 and 1999), an administrative assistant and two external examiners who are academics from other South African universities. (See Appendix 1 for data on Humanities staff who participated in the study).

A series of both national and institutional events in the late eighties and early to mid-nineties left this community of academics fairly bruised. During the mid to late eighties historically white English-speaking universities experienced a significant influx of students coming from educationally disadvantaged backgrounds, students who under apartheid legislation had been denied access to these institutions. A number of staff interviewed commented that the Humanities faculty had become an institutional 'catch-pool' for these students since their generally low matriculation points¹⁰ denied them access to other faculties. There was also a perception among some of the staff that the Humanities department itself served as a faculty 'catch-pool' for the poorest quality of student, what the course convener (in 1999) referred to in an interview as "the bottom of the barrel" (Kei 99-3). He argued that the Humanities department was "vital" in the context of a university that was taking in increasing numbers of students who for one reason or another failed to be admitted to or to make satisfactory progress in the major of their first choice (usually commerce, engineering or medicine). "(The Humanities department) became the common last chance for a degree for many students... This meant that the department accumulated a large proportion of students who did not intend to major in it but were stuck in it and/or were not very highly motivated academically..." (e-mail, 1/19/02).

At the institutional level the department had also suffered badly through a programme re-structuring policy which resulted in significant loss in student numbers between 1998 and 1999. Given the loss of revenue from student enrolment and institutional plans for rationalizing and down-sizing, threats of mergers and retrenchments loomed large for this particular community, and staff morale was generally low.

The final year project is located within the context of a third year first semester course (referred to as Hum300) which has been in existence since the late seventies, largely unchanged in its aim to equip majors in the department to investigate social

¹⁰ 'Matriculation points' refers to the score which students received on their final secondary school examination. A matriculation exemption (i.e. a pass) is a necessary requirement for university admission. Faculties then have specific minimum point requirements for admission to particular programmes of study.

phenomena in a wide range of professional fields or to carry on for higher degrees. At the time of the study the course had annually enrolled on average 150-200 students over the previous few years. As a requirement for the course (and for graduation) students conduct research projects which account for 55% of their final mark¹¹. Students are assisted with these projects by a series of lectures on research methodology. The project topics are set by the academic staff and generally reflect the staff member's particular research interest or area of specialization. Students select from the range of topics (or supervisors) and are allocated to groups on the basis of their selection. The groups (of approximately 20-30 students) are convened by the staff member who is fully responsible for supervising the students and for the assessment of the projects. The actual supervision of projects varies. In some cases students work as a class on a common research problem which has been identified by the supervisor. In other cases students choose a research problem within the area of a specialization. Some supervisors strongly encourage group work and others prefer students to work independently. Three weeks into the semester students hand in a project proposal (worth 15% of the final mark) and towards the end of the semester they hand in a final project (worth 40% of the final mark). The remainder of the final mark is made up of a statistics test (worth 10% of final mark) and the final examination (worth 35% of final mark)¹².

In terms of the internal moderation of the project marks, the standard procedure is that supervisors submit the project marks for their groups to a central administrator. A spreadsheet of marks is compiled of all course marks: the proposal mark, the project mark, the statistics test mark and the exam mark. Supervisors then come together in a meeting to discuss borderline cases, that is, cases where a student's final mark straddles a class boundary (e.g. pass/fail, upper second/first). Supervisors are asked to make and justify a particular judgment call. At the end of this process a sample of the projects and examinations (i.e. a selection of good, poor and middle range marks) as well as all the borderline cases of projects and exams are seen by an external examiner. In 1998

¹¹ Technically to fail the project is to fail the course. However the actual application of this rule varied from year to year. In 1997 it seemed to have been applied, but in 1998 it was not.

¹² There was a slight difference in how the final mark was allocated in 1998 and 1999. In 1999 the course convener introduced a series of in-class tests on the course readings which comprised a very small percentage of the final mark.

Hum300 had two external examiners¹³, both from other South African universities. Their role was to give feedback on a draft of the examination paper, and to confirm that the marks awarded by the internal examiners for both the examination and the projects were appropriate. If the external examiners felt that a particular mark was not appropriate, they would adjust it. This adjusted mark was considered by the department to be a recommendation. While it was common practice to defer to the external examiner, particularly in the case of an adjusted project mark, the course conveners (of 1998 and 1999) were clear that it was the department's prerogative to accept or reject the recommendation of the external examiner. I return to the issue of the role and authority of external examiners in the next case study.

During the time of my study (1998-1999) there were several modest attempts by the course conveners to strengthen the course's validation practices¹⁴. These attempts need to be understood in part against a double backdrop: national and institutional accountability agendas which emerged in the mid to late nineties and a painful and protracted departmental inquiry. At the national level the government implemented a National Qualifications Framework, in part, a strategy for requiring educational providers (including higher education) to be more accountable to the government for their academic offerings. In parallel with national policy at the institutional level SAU proposed in 1996 an Academic Planning Framework which required a more explicit articulation of academic programme purposes, objectives, learning-outcomes, methods and criteria for assessment.

An (arguably) more significant event was a high profile inquiry (1996-1997) into accusations of racism by a group of post-graduate students in the department. One of the issues at stake in this inquiry was the minimum requirements for entry into postgraduate level studies. From the perspective of the accused members of staff the faculty had admitted students to postgraduate level who "shouldn't have been there". They argued

¹³ The rationale for two external examiners was to cater for students in Hum300 who were registered for the general degree and those who were registered for a specialization within the degree.

¹⁴ As noted in Chapter 1, validating practices refers to all the mechanisms which a community of practice puts into place to ensure the reliability and validity of the assessment of student performance.

that these students simply did not have “the basic conceptual skills” (Jor 98-1). As teaching staff they felt that they were placed in an impossible situation of having to teach students who, even with extra assistance, were not able to cope at the expected level. The individual academics accused were cleared of the specific charges, but fundamental educational issues of access to postgraduate studies were left unresolved.

Thus against the backdrop of these external and internal pressures, I noted several attempts by the course leadership to strengthen the Hum300 validating practices. The course convener in 1998 was particularly concerned to establish some benchmarks. At the first Hum300 staff meeting of the year, he argued that the marks which students received for their Hum300 projects needed to signal the extent to which students were capable of postgraduate level study. His proposal was that students should not be given a mark above 65% for their project unless they were “honours material”¹⁵ (field notes, 16/2/98). His concern was that over-inflated project scores might result in students being admitted to postgraduate level studies who were not capable of succeeding. Like his predecessor, the course convener in 1999 was also concerned about consistency between assessors. He proposed a set of course objectives which appeared in the 1999 course outline. He also recommended that all Hum300 staff use a standard marking memorandum: he proposed, in fact, that they use his.

Staff responses to these attempts at standardization were varied. Some felt that standardization was important. Mary recalls orienting some postgraduate students who assisted her with first year marking. “I told them 65 is an important mark in this department...I let them know because otherwise you don’t know what a particular institution does with a particular percentage. You have to tell them how the institution feels about these marks” (Mar 98-1). Others were more skeptical. Joseph conceded that officially a ‘65%’ was a minimum requirement for entry to postgraduate level, but in reality he admitted, “I haven’t the faintest idea what a 65 means...someone who has got cultural capital and knows how to write nicely, make pretty little borders, puts the right references and has got all the things from short-loan on the right day, not the wrong day,

¹⁵ The Honours year in this faculty refers to the first year of postgraduate study.

and packaged it nicely, then that's 65. And somebody who has made a total mess of something (but) is clearly working with something...you can kind of push it into 65. And somebody who is learning how to deal with the way of doing things in the system, you know, competently is 65 (Jos 98-1). These two responses reveal something of the social-situatedness of the assessment process which I explore in more depth in the context of actual validating arguments¹⁶.

5.3.2 Validating arguments

In 5.1, I explored the social conditions for academic professional judgment, that is, the ways in which the academic profession in general and the disciplinary community in particular constitute academic 'professional vision'. In Bourdieu's terms, this is an exploration of the relationship between positions constituted by the field and agents' dispositions. Bourdieu's "logic of practice" helps us to make sense of the homology between the academic "game" and academics' "feel for the game", that is, academics, as a result of inhabiting the same field, do share a common "professional vision" (Goodwin 1994). Through the case studies, I now turn to a closer exploration of this professional vision. In this next section, I explore assessors' disposition, that is, how they take up positions constituted for them as academic assessors. I explore these dispositions through assessors validating arguments which emerge in two different contexts: supervisors' own accounts of their interpretive process, and the collegial discourse of a marking workshop.

5.3.2.1 Supervisors' accounts

In the second year of the Humanities study, I asked my interviewees to bring to the interview examples of student performance (in this case, the research proposals) which had raised particularly interesting assessment issues for them. These four accounts arose from these interviews, offering a close-up view of the assessment context and its multiple contingencies.

¹⁶ As noted above (4.3.4) validity argument refers to the actual discourse which staff use to justify their interpretations – whether to students, colleagues or the researcher.

Keith

Keith¹⁷ gives his students considerable latitude in identifying a research question, and he strongly encourages students to work on their own rather than in groups. He is looking for “self-initiators”, students who come with a research interest and can work fairly independently. Keith’s account of his marking process suggests a systematic, largely inductive process of assigning scores to parts, adding up those scores, arriving at a total score with some adjustments in light of the whole. Having given a score, he noted that he seldom makes adjustments for consistency across the whole range of students’ scores. He figures this will “come out in the wash”. He acknowledges, however, that sometimes he will start off with a “certain level of expectation and (after) I’ve done five or ten I think ‘Wow, hold on here I’m being too strict or too lenient or something’...but I mean a lot of the markers do that. It’s like taking a sample and trying to establish whatever the standard that you are going to mark at, and then you go back”.

However, when our discussion turns to actual projects, Keith acknowledges, “It’s a complicated story, because we are about teaching them (the students) something”. He elaborates, “A lot of students, for example, didn’t have time to do their pilot work and so...they got zero for pilot work...I say ‘Look, I’ll give you a week to go and test your question out on two or three people, get some comments on it and come back and tell me how you are going to revise it’. So it’s like that. A snowball...”. With reference to a project that initially got 16% on the draft, he argues “It’s a jolt... they (the students) are all used to getting passing marks because they are now in third year... you get a bright student. here’s one ... she only got 16 percent! So they sober up and they say ‘Hey, this is not a Mickey Mouse course, we had better get going’. Now (on the final proposal) she has got 72...I see the marks basically as a sort of carrot and stick thing”.

Keith had given another student a 48.5 % on her final proposal. This meant that she had to re-do the pilot study. “I want her to say a bit more about what questions she asked, how they responded, whether the data is useable, flesh out that pilot study. She’ll bring it back to me then at the beginning of next term and that mark will undoubtedly (go up), I

¹⁷ All interview data in this account is drawn from Kei 99-2.

don't know 58, 62 or whatever depending... I mean I don't want them to fail and be discouraged. I want the person to look at that 48.5 and say 'Hey that's not me'... and they move themselves a bit".

Mary

Mary's¹⁸ typical model of supervision in previous years had been a group project where all the students are involved in the same research project, collecting and analyzing data which was directly related to Mary's own research. Having collected enough data for her own project, she decided in 1998 (the year of the interview) to give students more of a "free rein" in selecting their own topics. Mary follows the same procedure as Keith, collecting drafts, giving students feedback, and asking them to re-submit. Unlike Keith however she does not use a marking memorandum to derive her mark. Mary's process of allocating marks consists of reading through the proposal and allocating symbols, for example, A, B, or B+. When I ask her how she then arrives at a percentage, she says, "I sat there and I thought, 'Well A is going to be 75', I decided to myself, 'D is going to be 50', then I decided 'Well B will be 65 since (this is an) important mark in the department and C will be 60'." In terms of what number (out of 100) the student gets within the range of marks for a particular symbol, she states, "It's absolutely arbitrary...I'm more aware than anyone of how arbitrary it is".

When asked what she is looking for in order to award a proposal, for example, a 65%, she notes things such as "signs that they're actually thinking things through", evidence that "their minds are ticking", movement "from an idea to an idea", "creative intelligence", "evidence of assimilation, not just pumping it out". She also notes, "I think I'd be very unlikely to give a 65 to someone who didn't write English... I think I would be misleading them." When I ask her what she means by "didn't write English", she elaborates, "something that would be embarrassing to show to a stranger and say, 'This is a student to whom I have given a 65'".

¹⁸ All interview data for this account is drawn from Mar 98-1.

In discussing her assessment of the projects, Mary noted one particular project that she was “really worried about”. The student had chosen as the topic of study a particular moral issue, and it was clear to Mary that the student had “a real problem”, something was really “bugging” the student with reference to this issue. The student became very defensive when Mary raised questions about some of the assumptions that were being made in the research design, but the student insisted in proceeding with the project despite Mary’s reservations. Mary is in a difficult position and admits, “I gave (the person) a higher-than-I-should-have mark. I thought ‘I really want to give this (person) about a C but (they) are going to just say it is because I am opposed to their ideology’. I gave (the person) a B or B+. I think I removed the plus before I gave it back to (them). Because (this person) is one of the ones that I really worried about and I went back and thought ‘Okay forget about my reservations. I am supposed to criticize the stuff’”. She then details how she went through the questionnaire with the student question by question pointing out assumptions which were highly problematic. “So once I sort of felt that I had sorted that out in my mind then I felt I could pull (the) mark down a bit. I mean I didn't feel so bad since I'd managed to rationalize my deep unhappiness with this project.”

Joseph

Joseph’s¹⁹ process follows yet another slightly different model of supervision (from that of Mary’s and Keith’s). Like Keith, his students have quite a bit of flexibility in their choice of research focus. But in contrast to Keith, he strongly encourages students to work in groups. Students hand in a draft proposal which can be a group proposal; he gives feedback; students re-work the draft and hand it in. Joseph marks the final proposals but does not put the mark on it – he prefers to have a discussion with the students about their proposals without having the mark “in the way”. Joseph feels that this system allows some discussion and feedback to take place which is not completely overshadowed by the grade. “In an ideal world, I’d like to do away with marks”, he says.

¹⁹ All interview data for this account is drawn from Jos 99-1 and Jos 99-2.

In terms of criteria for assessing the proposals, Joseph admits to a tension. On one hand, he resists the reduction of complex performances to 'labels' and 'numbers'. On the other hand, he acknowledges that students might benefit from those 'labels'. So over the years he had produced a number of different marking memoranda ("grids") which he uses with some degree of ambivalence. "I'll continue this year using a combination of a grid and not using it. I'll have a grid which I'll give to students but when I am actually marking the (proposal) I'm not going to tie myself to that grid. I can't. I can't work out a tight enough grid to be comfortable tying myself to it".

Joseph discusses a group proposal which he had failed, which was "a big mess". He admits that had he been using some kind of "grid" he would probably have given it 35%. Instead he gives the group a 47% so as not to "discourage them". But as a "product frozen in time", it was not a "near fail" but a "definite fail". He justifies this judgment on the grounds that, firstly, it fails to meet a number of the requirements of a research proposal, and secondly, "There's absolutely no basis under which this will work as a plan for a research project...if they go and try and do this it will just be a complete disaster". The overall message which Joseph wants to convey to the students is that they must "disassociate themselves, free themselves from what's in here (the proposal), because it's not going to work". Out of what is a "big mess", they need to "pull bits of what is either in here, or was meant to have been in here...and start moulding those bits into a plan which will actually make a successful third year research project".

On the basis of Joseph's account, it is clear that he is drawing on other 'texts' than the draft proposal. "I can see something of a kind of process going on...I've actually seen those people work and I've actually listened to them working...it's dramatically better than the draft proposal." Joseph is also 'reading' the students: "They are not lazy". It is rather a case that "This whole thing (research) is alien to them...they are separated from some of the things which others, using different cultural capital, just smoothly slide into...". On the basis of these multiple 'texts', he speculates about the "mess". "What I think has happened in this particular thing is that just at the point where they start to get somewhere, a whole lot of different ideas start crowding in which they are unable to

discipline or handle or organize, so it's exactly because they've reached a point of progress that this thing has got more messy than it might otherwise have been." He adds, "But I'm sure that they can do it".

Zora

Like Joseph, Zora²⁰ also believes strongly in the value of group projects. In terms of her supervision, like the other supervisors, she takes in draft proposals, gives them extensive feedback, as well as a mark. She also has regular weekly meetings with the students. What is clear from her account is that the projects which the students conduct are issues which she herself is clearly invested in (e.g. black identity) and has grappled with personally and professionally. Many of the students in her option have done a previous second year course with her. She has established a relationship with them. There is a history which she is building on.

In terms of marking criteria, Zora does not use an explicit marking memorandum. She described her mark as the outcome of an impression that accumulates through the process of the proposal development. She is looking for certain qualities, for example, "focus – have (they) managed to narrow down to a specific research question", "literature review – does it link with the things they are trying to research", the "quality of the interviews" – but she admits, "Maybe I privilege process...because I do mark the final in relation to what I got from the draft, and in relation to the amount of work they've done". For Zora, the mark means something, "in relation to (the) draft proposal, as well in relation to my expectations".

In relation to particular projects, Zora discusses one group project, where the students got a 55% on the draft proposal. She had made copious suggestions on the draft for how the students could take the proposal forward. But on the final, "I gave them 58 because I felt I couldn't give them 60, but they worked so hard that I want to give them 60". She describes a process of re-ordering one of the paragraphs in order to create more coherence. She asks the students to go through the proposal and follow her example. "So

²⁰ All interview data for this account is drawn from Zor 99-1.

what I say to them is, if they do the basic minimum changes, on the basis of my restructuring, then I would consider giving them 60% because they've just been working so hard...And I feel as though I want to reward that, but I also don't want to give them wrong messages so that if they were to go to someone else with their project, the person would say, 'No, but this is definitely not a 60'. So it's that kind of balance that I'm trying to work with." She gives them a chance to re-work the "final" proposal again for a 60%. "Even if it's just a 2%, (it's) just sort of pushing them beyond a third into a lower second; it's also about motivation...so that they don't lose hope." She concludes, "So it's hard, you know, cause you're also dealing with people".

5.3.2.2 Dimensions of disposition: assessors' interpretive matrices

These accounts give some insight into a complex set of contingencies which shape assessors' readings of student performance. Each interpretive act involves a whole complexity of contingencies which influence the final outcome. I introduced in Chapter 3 the concept of interpretive matrix to describe this complexity of contingencies, where 'matrix' is understood as a "context or environment" or "womb" (Oxford Dictionary), a place where something "originates or develops" (Webster Dictionary). A crucial part of understanding academics' validation practices comes from exploring these interpretive matrices and how the local exigencies of an interpretive act may result in an adjustment or even a divergence from tacitly agreed-upon points of reference or criteria. On the basis of the accounts I tease out different functions which assessment is serving, and suggest that different functions might privilege different kinds of evidence.

The four supervisor accounts suggest that within each of these interpretive acts assessment is serving simultaneously two different functions. At one level the mark is an interpretation of the value of a product. It is a measure of a product against some notion of what the community believes to be 'good' research, some standard of what is 'competent' performance at this level. These marks also come with pre-assigned institutional and societal meanings, and accompanying messages. Thus Mary cannot give a 65 to a project where the "level of English would be an embarrassment". Joseph has to fail the proposal because it does not meet the "requirements"; the project is simply "not

going to work". Despite Zora's desire to honour her students' hard work she cannot give them an extra 2% since someone else might look at it and say "This is not a 60%". These accounts all point to constraints on the interpretive process, constraints which are to some extent imposed by intersubjective notions of what is 'good', for example, good research and good English. The function of the mark is to convey to students (and other recipients) something about this community's notion of what 'good' is, what it values.

But serving as a measure of product is not the only function of marks. The mark is also a measure of a process; it serves as currency in a communicative exchange between assessor and student. In Keith's words, as "carrots and sticks"; a means of "jolting" students into action (Keith); a means of communicating disapproval (Mary); a means of holding up a picture which students can choose to "disassociate themselves from" (Joseph) and say "That's not me" (Keith); a means of affirming hope (Zora). For these purposes, the mark is, to use Keith's metaphor, a "snowball"; it is not a solid entity with fixed and stable meanings but an evolving entity. Just as the snowball changes size as it rolls so too a mark gathers meaning as it 'rolls', the changes reflecting different sets of contextual considerations.

The tension for assessors is that at some point in the mark's 'travel' meanings may become temporarily fixed. A mark becomes a *final* mark. It gets entered into a spreadsheet. It gets discussed at a departmental meeting, reviewed by an external moderator, aggregated with other scores, printed on a transcript. It becomes reified as a signifier of achievement. Decisions are made on the basis of these scores. Students are passed or failed, they are admitted to postgraduate level or not, their transcripts are looked upon favorably or otherwise. In its reified form the mark's meanings from previous contexts may be eclipsed. Where once the mark was a "carrot", it might in another context become a "stick".

All four accounts suggest that this tension between marks as an interpretation of a product and marks as an interpretation of a process is an unavoidable dilemma for supervisors. One of the most insightful articulations of this tension comes from Joseph.

He notes the difficulty assessors face when they receive “a written product (that) doesn’t reflect what I know they (the students) have done” (Jos 98-1). He notes that it is a particular feature of a course like Hum300 where “you’ve worked with those people and you’ve listened to them, you know the difference between what they’ve said and what they’ve written...you know the artificiality of what comes onto the paper” (Jos 99-1). This creates a dilemma for assessors and they consciously or unconsciously adopt strategies to deal with this. Joseph’s strategies are conscious. “I give them (the students) a mark which tries to account for both (what they have done and what they can do) – I don’t want to give them signals which are not sufficiently located in the product... I tell them ‘the mark I give you has to be a mark for what you’ve written’. But I also try to take account of what they haven’t written, or what they write in another thing.” Thus, for Joseph, the mark is not irrespective of the product – but an interpretation that results from a particular view *through* the product to a person who is in process. Joseph articulates the supervisor’s position of assessing a work-in-progress and not only progress through a course but the progression of their undergraduate career. For Joseph, this is the supervisor’s privileged position.

Joseph argues that despite these “subjectivities which we cannot escape”, there is a kind of “balancing which goes on which I find very difficult to understand...what is on paper and the (assessor’s) judgment of the person and the process, they are not far away from each other”. But this has to be “secured” and the Hum300 offers that kind of opportunity because of the active engagement with students. It is that active engagement with students, “the extra hour with that person”, that brings about that security, that consistency. So Joseph feels an obligation to students (as a supervisor) to help them pull out what they know, what he knows they know, that is not yet committed to paper. It is this process that secures consistency between the judgment of a person, a process and a product. For Joseph anyway, this is a validation process.

While other staff also expressed the same tension – between assessment of person, product and process – they adopt different stances. In his assessment Frank acknowledges that “the student is on a journey, it is a process”. “But,” he argues, “I also

recognize that this is a product and it does in some ways need to stand on its own. It's a proposal. And so I don't want to read too much into it which isn't apparent, (that) isn't explicit" (Fra 00-2). Elsie feels quite strongly that it is *only* product that should be assessed. Her perception is that too much emphasis on a student's process runs the risk of under-prioritizing the quality of the product and this is potentially threatening to both her and the student. "I know there are students who struggle. I can see they are making an effort, but at the end of the day, when I look at the reports I struggle within myself. Am I doing this student justice in passing this report? Let's assume the report passes, that the student gets a job and the next thing when they get to the field, then people start blaming me, you see. 'Those black lecturers they employ at (SAU), now she taught us nothing'. So I'm scared for my reputation and also for the student. Professionally if I do something like that, it's unethical" (Els 99-1).

In summary, these validating arguments give us some insights into supervisors' interpretive matrices, something of the 'environment' out of which their interpretations are born. The accounts expose a tension between the multiple functions which the assessment is serving. On one hand, the mark is a measure of product against some tacitly agreed notions which this community (within the wider context of the field) holds about 'good' research. On the other hand, the mark is a measure of process, a communicative exchange between a supervisor and a student. These functions create tensions because they privilege different contextual considerations, different forms of evidence. In its function as a measure of product, assessment interpretations may foreground products (e.g. the research proposal) and background processes (e.g. the drafts, the class discussions, one-on-one time with the student). In its function as a measure of process, assessment may privilege evidence from the student (e.g. their effort, progress, and attitude) and the immediate, local context of the classroom (e.g. the student's ability in relation to the class). Of less concern from this perspective is the meaning of this mark outside the classroom. As Joseph argues, "It's a massive tension, a permanent tension" (Jos 98-1), but, I would argue, one which must be held, for neither function is dispensable, neither context can be ignored; all these forms of evidence constitute the supervisor's interpretation.

5.3.2.3 Collegial discourse

I now turn to an analysis of validating arguments which emerge in a different context – the context of a marking workshop. Within the marking workshop assessors occupy different roles, for example, some are supervisors, some are not (I explain this below). My analysis of this collegial discourse seeks to establish how an assessor's role shapes his/her interpretive matrix. Do supervisors draw on different forms of evidence for their interpretations than assessors who are not supervisors? Does this result in different interpretations of student performance? If so, what is at stake in the negotiation of these differences?

For my research purposes it seemed important to have access to collegial dialogue, that is, dialogue where assessors justify their interpretations to other colleagues. Although some Humanities staff would informally swap projects, a more formal occasion, such as a marking workshop, was not a feature of this particular community of practice. I thus made a request to the course convener for such a workshop and he agreed to organize and facilitate it with my assistance. He felt that such a discussion could be particularly beneficial to the new staff members. All staff received a letter from me inviting them to attend the workshop. The purposes given for the workshop were stated as:

... We thought a discussion of the marking of (Hum300) projects would be useful for two reasons:

- 1) (The convener) felt that this would be useful given that there are new people teaching on the course who might appreciate some discussion around the criteria for marking.
- 2) For the purposes of my research it would be useful for me to hear both the common as well as divergent assessments of those involved in marking....

(excerpt from letter dated 14 May 1998)

Time constraints necessitated the selection of two research projects from the previous year (henceforth referred to as projects A and B) which had already been

marked by two members of staff who were in attendance at the marking workshop. As is common practice in marking workshops, the marks, the students' and supervisors' identities were all removed from the projects. As there were no commonly-agreed upon marking criteria, the workshop participants were simply asked to read the projects, allocate a mark and come to the workshop prepared to discuss and justify their mark to the group. A couple of weeks before the current year's project marks were due²¹, a group of eight markers (out of ten), plus myself, gathered for one hour in the departmental tea room. The discussion was facilitated by the course convener, Dan. The first half of the workshop consisted of each assessor presenting their argument, that is, their mark and any justification. This was followed by a general discussion which consisted largely of assessors presenting further justification for the marks they had assigned.

There was no final agreement on the mark for either project by the end of the workshop. The range of marks for project A was between 41% and 60% (with four out of the seven marks clustering around 50/51%) and the range for project B was between 40% and 67% (with four out of the seven marks clustering between 55-60%). The focus of the debate for project A was whether the project should be passed (as the supervisor argued) or failed. Given the significance of 65 in this department, the focus of the debate for project B was whether the project warranted a mark above 65% (as the supervisor had awarded).

It became clear both during the workshop as well as in succeeding interviews that staff had their own 'theories' about why the workshop was unable to achieve agreement on what marks to award the projects. I noted at the outset the artificiality of the exercise, that is, assessing projects that had been assessed by the supervisors in the previous year may have reduced the 'stakes' in this discussion²², and curbed any push towards consensus. In interviews with staff (some but not all of whom had attended the workshop) other 'theories' for difference were put forward. In the context of the

²¹ These marks refer not to the projects under discussion, since as I noted they had already been marked, but to the projects which were being submitted in 1998.

²² Mary argued in response to a draft analysis, "If we had been assessing instead of talking about assessing (if the students' futures hung on it) that would have been different" (e-mail, 4/21/02) (parenthesis in the original).

workshop discussion, Frank proposed that staff were either drawing on different criteria or weighing the same criteria differently. From the perspective of a new member of staff, he found this unsettling. Keith attributed differences between staff to “irreconcilable positions”. He argued that staff comes with different ideological, political, epistemological positions, and those influence the “criteria” which staff use to assess student writing. For him these differences, while understandable, were “indefensible” (Kei 99-3). Similarly Elsie argued, “We have different research values, we have different opinions, we have different expectations, so it’s very difficult to maintain consistency” (Els 99-2). Although consistency was desirable, she doubted whether it was achievable. In contrast, Dan argued that it was not different criteria for ‘good’ research that were at stake in the workshop. Staff, he argued, share a common understanding of what ‘good’ research is. Differences in assessment emerge as a result of different contextual considerations when assessment criteria are applied to student performance (Dan 98-2). And finally, Joseph agreed that there were indeed differences but from his perspective in a course like Hum300, consistency may not be desirable. “I think (Hum300) gives people (students) the opportunity to develop something which is their own. To work on it consistently with somebody who is marking it. To respond to the guidance of that marker knowing full well that it wouldn’t be happening with another person. That something else would be happening with another person” (Jos 99-1).

I turn to the analysis of validating arguments in the context of the marking workshop. The excerpt of argument which I have selected is taken from the very opening of the workshop where Keith and Mary present their marks and justification for them (see Appendix 2). For the purposes of this analysis my analytical point of entry is the text from which I move iteratively between text and *context*, drawing on other sources of data for the interpretation of the text. This context includes the workshop, the context of participants’ practices (or the community of practice), and the broader departmental and institutional context.

As noted above (4.3.4) topic analysis is a form of textual analysis which attempts to get at what the speakers are talking about. Of particular interest are the ways in which

speakers chain topics together, and what these topic chains reveal about speakers' common sense notions. A cursory reading of this opening marking workshop discussion suggests that the topic is about Keith and Mary's assessment of the projects, or more specifically a disagreement between them over the assessment of project B. Keith has failed project B and Mary has passed it "generously" (MW 32²³) with a 67%. However, an analysis of the various 'links' in the 'topic chain', that is, how the topic shifts, gives insight into some of the inferences which underlie assessors' interpretations. An analysis of the topic chain exposes the multiple functions which the discourse is serving. I focus on two main functions which this text is serving. This discourse is functioning ideationally, in what I argue is a contestation about what constitutes legitimate evidence, and simultaneously the discourse is functioning relationally in the interest of collegiality, what linguists refer to as facework. As noted above (4.3.4) 'face' is a term used by linguists to refer to the public self-image that people want to claim for themselves. 'Facework' refers to the various strategies which people use in interaction to manage and minimize the threat to their own face and that of others (Ting-Toomey, 1994).

In the excerpt of text I focus on four links in the topic chain, where each 'link' is a new topic. As noted in Chapter 4, of particular interest in this analysis is how the topic shifts and the ideational work which underlies these shifts. The first link (MW3-40) consists of Keith's argument and the topic of this text unit is clearly the projects, that is, the products which all the assessors have before them. Having pronounced them both "terrible pieces of work" (3) the topic then focuses on project A (4-22): "the other one"²⁴ (4), "the A one" (8), "this one...is very weak" (10), "(it) just trots out" (12), "it just kind of ends" (14), "I think it has one reference" (18), "it sort of it says" (19), "(it) doesn't refer to a single source" (20), "it doesn't locate the topic" (21), "(it's) so very weak" (21). The only exception is "if these were my students..." (15-17), where the implied topics are Keith's supervision practices and his students. Then from lines 22-40 where Keith discusses project B, once again the topic is the project: "this one...this is a

²³ References to data from the marking workshop are labelled 'MW' and followed by the line number. Where it is clear from the context that I am drawing on data from the marking workshop, I have omitted the 'MW' for ease of reading.

²⁴ I have underlined the actual referent to assist the reader in following the topic chain.

terrible piece of work” (22), “you passed it” (24), “it just goes nowhere...it just ends” (36-37), “it’s got an appended...” (38-39). Again there is an interesting exception in line (30-32). When Keith is asked to justify his low mark (29), he shifts the topic from the report to the student writer of project B. He moves from “this is a terrible piece of work” (22) to “well, I mean, this student just is going through the motions but doesn’t understand what he or she is doing...misreads tables, doesn’t know how to actually do anything with the data...contradicts what’s in the tables...” (30-33). It would appear that even though Keith does not know the student writer of project B, there is an inferred writer in view, one who in his judgment gets “plug” (40).

In providing the other participants with a justification for his marks, Keith appeals to his “old marking schedule” (MW 4). What he means is that he has assessed these projects using his marking memorandum, the one which he uses to assess his own projects. It would appear that for him, this “marking schedule” serves as a kind of code for a whole range of expectations he has when he approaches the projects. He is looking for a certain level of methodological and analytical discussion, though what level is not made explicit; he is expecting an interview questionnaire in the appendix; he is looking for a certain level of coverage of the literature, though again how much coverage is not specified. Thus his marks signal the value of these projects in relation to a particular notion of a ‘good’ research project, which for Keith these projects clearly are not.

Later in the workshop, Keith again defends his assessment by giving some insight into the rationales that inform his mark:

“Can I just say something...it sounds very mechanical and technician but it’s part of skills training for me. I want a student when they graduate from (Hum300) to know what an abstract is...to do a table of contents, to do a decent cover, you know, with all the information. The same with an introduction, the review of the literature, you know, so every little brick of the report is important. And so I don’t want a student to get away with just sort of doing a beautiful review of the literature (but) never learn to do research. The same thing goes for... reflexivity, reactivity, ethics

that must be all those things must be there. I must tick them off...the problem comes into honours as you know yourself, even PhD ask a student 'write us a proposal' and they write us rubbish as though they've forgotten everything" (MW 264-268*²⁵).

In justifying his mark and his score sheet Keith gives us insight into the interpretive process that informs his assessment. For Keith the purpose of the course is skills training, providing students with a repertoire of research skills that they must be able to have and use to graduate so that they are able to transfer those skills to further postgraduate study. Thus for Keith a crucial contextual consideration is what the department values in its graduates.

I now turn to the second link in the topic chain, Mary's argument (MW42-48) where a different picture emerges. Mary's discourse starts with the topic as the project: "I gave this one 67%" (42), "this is the one I did...that one" (44). But then we note a topic shift from project B to the writer of project B: "this girl", "a really clever girl" (46), "I can't get her out of my mind" (47). We see in Mary's justification of her mark a shift in topic from 'project' to 'person'. We note the same move later in the workshop when the supervisor of project A, Larry, justifies his mark which is higher than other markers have awarded. He argues, "You can't over-intellectualize these things, as a supervisor you know where the thing comes from and you know what the student has done..."(MW 107*), and later in the workshop he defends the student even more emphatically, "Let me tell you something about this person..." and goes on to talk about the student's "enthusiasm", that he was "verbally very impressive", "very persuasive" (MW 188*), and that he put a tremendous effort into his interviews. Implied in Larry's defence of the student is that as a supervisor he is assessing a product but he also has a person and a process in view, that is, the manner in which the student and the research progress over time.

²⁵ All references followed by an asterisk refer to parts of the transcript outside the selected excerpt. The numbers in brackets refer to text units in NVivo, not line numbers.

It is interesting to note that in presenting her argument, Mary does not refute the “criticisms” (MW 29) made by Keith. It would be tempting to speculate that this is a case of gendered relations, that is, a female of less seniority who may feel disempowered by a more senior male in an authority position²⁶. I do not believe that this is the case however²⁷. I would argue that Mary actually agrees with Keith’s assessment; she accepts that the project is not strong. She concedes, “There are lots of mistakes in here, this is wrong, that’s wrong”(MW 59). It is rather the case that her interpretations of the student’s performance are justified from a different evidential base. I return to this point below.

In the third link of the topic chain (49-58) we note another significant shift from the student to Mary herself, specifically her supervision practices: “how much prompting did you prompt her on the analysis?” (49), “did I prompt her?” (50), “did you say...” (51), “we all put the data in” (52), “you leave it up to them” (56), “we were running at a different pace” (57-58). We can only speculate about the motives underlying this shift of topic, presumably Oliver and Dan are trying to place the student performance in a wider supervision context. Dan follows Oliver’s lead by asking, “then you leave it up to them, then you don’t get a draft in and tell them to redo it?” (56).

At one level Oliver and Dan are asking for information from Mary. I would argue however, that they are not simply asking for information; here we see the ideational and interpersonal functions of language simultaneously at work. I would argue that Dan suspects that Mary’s supervision is not as thorough as it could be²⁸. But he is also aware of the risk that this inquiry may be perceived as a threat to Mary’s face. In view of the

²⁶ As noted in Chapter 2 Starfield (2001) shows the effect of gender relations on assessment discourse.

²⁷ When I suggested to Dan in an interview (Dan 00-1) that this might be an instance of deference on Mary’s part, he discounted it. He noted that there was a “bedrock of respect” between Keith and Mary (e-mail 18/9/99). In response to an earlier draft of the analysis Mary also strongly denied that she had deferred to Keith (e-mail 4/21/02). Keith also argued that this should not be interpreted in the “clichéd gender/power way”. He noted “I think (Mary) views me as her intellectual inferior (which I think I am)” (e-mail 1/10/02).

²⁸ Support for this argument comes from Dan in an interview (Dan 98-2). However, in response to an earlier version of this analysis he clarified that it is not that he is questioning Mary’s competence as a researcher but that under the pressure of time, Mary’s supervision is not as thorough as it ideally should have been (conversation, 27/9/02).

potential risk of threatening Mary's negative face, that is, the threat of intruding into her classroom practice, Dan constructs his question in such a way as to minimize the threat of interference²⁹. The tone of his inquiry is non-confrontational, almost apologetic. He foregrounds Mary's agency in relation to giving students independence, "you ... leave it up to them" (56). Ultimately Dan offers her a face-saving line of escape which she takes, "Um you know I really can't remember that" (57).

It emerges in an interview with Mary that there are a whole range of factors influencing her interpretation of project B. Unlike most of the option supervisors who allowed students to choose a particular question or problem within a designated interest area, in 1997 (the year project B was conducted) Mary and her students were involved in a group research project around Mary's own research. In my interview with her she describes the research process on which she and the group were embarked and how it set up its own particular set of possibilities and constraints. In terms of its possibilities students got experience of an authentic research project; this was a genuine problem to which real social science researchers were applying their minds. Students knew that the data they were collecting was data that might be used by Mary and others interested in this problem. This translated from Mary's point of view into a genuine interest in students' findings, as noted in her feedback on one student's proposal, "I will be very interested and so will other people researching this little understood contemporary South African (issue)" (excerpt from Mary's feedback on draft proposal).

On the other hand, the authenticity of the research also threw up problems for the students. She explains that her group was "running at a different pace from the other (groups) because we were sort of held up by each other" (MW 57-58). She related in the interview how what had appeared to be a productive research route turned out to be a "cul-de-sac" (Mar 98-1). The students panicked when they perceived themselves to be far

²⁹ Brown and Levinson (1987) present a range of strategies that speakers draw on to minimize the threat to face. Some of the ways in which negative face is redressed are with "apologies for interfering or transgressing, with linguistic and non-linguistic deference, with hedges on the illocutionary force of the act, with impersonalizing mechanisms (such as passives) that distance the speaker and the addressee from the act, and with other softening mechanisms that give the addressee an 'out', a face-saving line of escape, permitting him/her to feel that his response is not coerced" (p. 70).

behind their peers in other groups. Mary was also conscious that the project required types of quantitative analysis that many of the students were not adequately prepared for. Thus her supervision approach was closer to that of an apprenticeship, where the students were to some extent thrown into the deep end to experience first-hand the difficulties of social science research, albeit with her support and support from the group.

So in her assessment of project B, Mary is considering a host of factors which inevitably influence her interpretation of this particular student's performance. She is keenly aware of the difficulties the students have had to overcome, some of which are not of their own making. She is also conscious of the fact that the students' projects are to some extent her project. Thus in assigning a value to the students' projects, Mary is drawing on a range of contextual evidence, including her own role as supervisor.

The final link in the topic chain (MW 60-70) is constituted by multiple and complex shifts. In line 59 Mary shifts the topic back to project B: "in here", "this is wrong", "that's wrong" (57-59). Mary shifts the topic to her marking practices: "I take", "I think", "I rank", "I chuck", "if you want me...I can..." (60-61). Keith contests the topic of Mary's marking practices: "that's not what we're here about" (63) where "that" specifically refers to "jack up my standards". Mary reasserts topic control, "no no that's what I'm saying" (64), and shifts the topic to academics' culpability ("our", "we", "our courses") in student failure ("the students", "them", "their", "25%", "wastage", "material") (64-69).

For the purposes of this analysis I focus on the point at which Mary re-asserts control over the topic "no, no, that's what I'm saying..." (64). I take this to be an instance of formulation. A formulation is a type of interactional control where a speaker attempts to "win acceptance from others for his/her version of what has been said, or what has transpired in the interaction, which may then restrict the others' options in ways which are advantageous to the formulator" (Heritage cited in Fairclough 1992). Mary shifts the focus from herself to the staff as a group, "if students are doing so badly that's our fault", "we don't fail them", "we have to say", "our courses" (64-67). In part Mary seems to be

formulating an explanation for student failure which attributes collective responsibility to this group of academics, constructing a version of reality where academics are complicit in and culpable of a failing educational project, that is, the production of good graduates.

As noted earlier there is a commonly held perception that the department is a faculty 'catch pool' for some of the poorest quality students. It emerges (outside the workshop) that Mary feels that her research option is, in turn, the 'catch-pool' for weaker students in the course who for a variety of reasons self-select into her group. She refers to them as "the rump end of student talent" (e-mail, 28/5/98). According to her calculations 32% of the 1998 student cohort for the Hum300 course had to repeat their second year. The failure rate in her group, 38%, is above the average. She remarks, "I have the impression that most of my (students) scrape along at 52, and (I) look enviously at those (colleagues) who've got the highfliers" (e-mail, 28/5/98).

Against this background, Mary formulates a version of reality which couples students "doing so badly" with "our fault" (MK 65). She attributes collective responsibility to this group of academics (and perhaps the institution) for the production of poor graduates. I speculate here that Mary is not inferring that student failure is a result of her colleagues' poor quality of teaching, but rather the "fault" lies in admitting weak students in the first place, and passing them from year to year. Thus the dilemma which Mary poses is the inescapable realities of the social-political context as an influence on assessors' interpretations of student performance. What academic in good conscience can fail a student in third year who should not have been passed into third year, or who should not have been admitted to the university in the first place? This illustrates the dilemma posed by competing discourses, which I noted above (5.2.1), of equity and access, on the one hand, and efficiency, excellence and competition, on the other.

In addition to creating a particular version of reality, Mary's discursive shift from herself to the group may also be a strategy of minimizing threat to her face, in other words, diverting attention away from herself to the culpability of the group. Reflecting

back on this “complicated interplay” between Mary and Keith, Dan notes that meetings such as this can be “dangerous moments” for academics (Dan 00-1). Perceptions of professional incompetence can result in “real sanctions”, what Dan refers to as the “power relations of collegiality”. “People can close doors to you in areas you want to teach in, close committees to you where you want to be, close sabbaticals to you...all sorts of things...these are things that academics struggle very hard over” (Dan 00-1).

To what extent is Mary’s version of reality shared by others in the group? Oliver distances himself from it or at least aspects of it. It is “certainly not all our fault” (MW 70) he declares. In contrast Dan expresses solidarity, “No, I think that is important” (MW 71). Later in the workshop, in relation to justifying why report A should get a “50”, he elaborates “I think this is not good enough for third year research. But I would pass it. I’d give it 50...it’s coherent enough to not be (too) terrible. But secondly I think that (SAU) needs to exclude this kind of person from getting into (here), and we’re not. We’re under pressure in this faculty to accept these sorts of people. So we accept them in first year, they just get through and then (they) reach their third year and I think it’s a problem to then fail them” (MW 158*).

Sensing a rising tension in the workshop, Dan attempts to bring about resolution by affirming supervisors’ “freedom of approach”:

“I’m not against a different approach. I think Mary gave them the experience of a survey, and then said, ‘Look, go for it and see what you come up with’. And if you take that approach, I think you have to mark within that approach. That’s why I wouldn’t judge this (project B) by Keith’s (criteria) because they (the students) didn’t get what Keith told them...my own personal opinion is there has to be freedom of approach... I think you to some extent have a personal way of marking, depending on what you’ve told them” (MW 248*).

I would argue that Dan’s argument for “freedom of approach” is discourse functioning both relationally and ideationally. In terms of its relational function, Dan, as chair of the meeting, is attempting to minimize any further threat to participants’ face,

particularly Mary's. But Dan is also trying to articulate something about the nature of academic professional judgment. Later in an interview Dan reflects back on Mary and Larry's assessments in the marking workshop and re-iterates the importance of giving the supervisor "freedom of approach" (Dan 99-1); it is the supervisor who defines the problem, they know what they have told the students and "(they) have to mark within that framework" (Dan 99-1). For this reason he argues, "the culture of (Hum300) is to give high autonomy to the lecturer (supervisor) because only they know what was defined" (Dan 99-1). But, in yet another interview, he qualifies this "freedom of approach"; "there is (however)", he argues, "such a thing as better research...it is not just an arbitrary thing...things are not totally relative" (Dan 00-1)³⁰.

Some of the participants in the workshop are disquieted by Dan's "personal way of marking" approach. In response to Dan, Mary responds, "But we ought to agree to norms, (to) standardize...otherwise students are at the mercy of our idiosyncrasies" (MW 250*). Frank, who describes himself as the "new boy on the block" (MW 199*), argues for commonly agreed benchmarks, for example, the minimum requirements for passing and for achieving 65%. He concedes, "Maybe it's an impossible task but...for me it would be better if we had some sort of common understanding" (MW 252*). Later in the workshop, he proposes again, "I really feel it would be beneficial for students if there was some sort of common ground that their marks were occupying, rather than huge differences" (MW 312*).

In an attempt to move toward consensus, Dan suggests that they go around the room and each participant put forward the conditions under which they would fail a student (MW 283*). He meets resistance from the group. Mary responds, "In the abstract I find it very hard to say whether (a student) is going to fail" (MW 287*). Dan's suggestion is not taken any further. The topic changes to how to deal with projects which are blatant plagiarism and the workshop adjourns.

³⁰ In both the Dan 99-1 and Dan 00-1 interviews Dan is giving me feedback on earlier versions of my analysis. Dan wrestled in particular to articulate his position as an assessor, a position which was neither objectivist or relativist.

5.3.3 Conclusions

I have explored in this analysis assessors' socially-situated validating arguments. My analysis set out to expose some of the dimensions of disposition, what I refer to as the assessor's interpretive matrix. Socio-cultural and interpretivist approaches to assessment have stressed that 'context matters'. Thus my study set out to explore these contextual 'matters'. Supervisors' accounts of their own interpretive processes reveal a complex set of contingencies which they are balancing. These include the tension between the mark as a measure of the product and the mark as a measure of the process. The analysis of collegial discourse suggests that different assessor roles will inevitably give differential priority to these assessment functions. These functions privilege different kinds of evidence, and may generate different interpretations. Keith draws on evidence primarily from the project and his point of reference is the context of the course and the kind of graduates the department values. His interpretation appeals to a set of implied standards against which the performance is to be measured. Mary's interpretation is premised on a different evidential base: what she knows about the student, the class, the research process and her particular role as a supervisor, as well as the broader social-political context and the inevitable compromises this poses. While not denying the standards to which Keith appeals, Mary privileges the 'particular' in its more immediate setting.

A consequence of these different matrices is that assessors will not necessarily agree: Mary passed the project and Keith failed it. Thus I posed the question, what is at stake in this negotiation of these differences? I argued that the discourse is simultaneously functioning relationally, that is, in the interest of collegiality, and functioning ideationally, that is, there is a contestation between the different purposes which assessment serves. One of the stakes in this discourse is collegiality. While these academics argue strongly for different interpretations of the performance, they do so conscious of the potential threat of these arguments to their own and others' face.

The other issue at stake in this discourse is this community's perception of themselves as a community that can come to consensus, a community that does share a

common professional vision of what constitutes 'good' research. Underlying this discourse may lurk some form of Cartesian anxiety (as discussed in Chapter 3). At the heart of this anxiety is the belief that as assessors we are faced with a choice between two alternative forms of rationality. One (what Bernstein refers to as 'objectivism') is the quest for 'standards', for 'fixed points', for 'common ground' against which performance can be measured. The other alternative (what Bernstein refers to as 'relativism') is a resignation to the reality that there is no 'overarching framework' to which we can appeal, there are no 'fixed points'; there are only 'your' standards and there are 'my' standards. For this community another significant stake in the resolution of difference is the implications of dissensus. What does dissensus imply for the kind of rationality that informs their interpretations of student performance?

While the marking workshop discourse constitutes this anxiety, that is, the perception of these forms of rationality as an either/or proposition, I argue that the members of this community do not hold either of these alternatives as viable options. Indeed their very own practices are a testimony to a different kind of rationality, one which is "beyond objectivism and relativism" (Bernstein 1983), one which eschews the antinomy of objectivity and subjectivity (Bourdieu 1990). Their accounts are articulations of an alternative rationality, one which moves iteratively between the general and the particular: between on the one hand, a community's abstract and latent notions of what is good, acceptable and appropriate, and on the other hand, a specific instance of a person's product as an outcome of a particular process. Academic professional judgment is constituted out of this tension. I turn now to my second case study in order to further elaborate on the rationality which characterizes academics' assessment of student performance.

5.4 Engineering Case Study

5.4.1 Validating Practices

My second case study focuses on the community of practice whose enterprise is the assessment of the final year projects³¹ in one of SAU's Faculty of Engineering departments³². These projects are a requirement for all of SAU's engineering programmes and a standard feature of many engineering programmes around the world (Jawitz et al 2002). This community of practice consists of all the academic staff that are members of the Engineering department (a group of about twenty in 2000), the external examiners (a group of seven in 2000), and one administrator from the department (see Appendix 3 for data on Engineering staff who participated in the study³³).

Like the Humanities department, the Engineering department had experienced a significant growth in both numbers and diversity of students in the 1980s. More recently there had also been some changes in the staff composition as the department had been able to appoint a number of black academic staff³⁴. In contrast to Humanities, however, the Engineering department had been largely unaffected by both the institutional and national policies for programme re-structuring. Their primary external accountability is to their professional body (Engineering Council of South Africa) which had in the year of my study conducted an audit of SAU's Engineering faculty. Although one of the objectives of the audit was to focus specifically on the alignment between learning outcomes, curriculum and assessment, for a variety of reasons, it had very little impact on Eng400's validation practices, the course in which the final year project is located.

Consistent with the logic of the field and SAU's position within it, it became clear from my initial round of interviews that (with the exception of the administrator) the members of this community's primary professional loyalties lie with the communities of

³¹ Since in the Engineering department staff often refer to the final year projects as 'theses', I use these terms interchangeably.

³² Henceforth when I am referring to the department I use the term 'Engineering department'.

³³ As in the Humanities case study, all the names have been changed to protect participants' anonymity.

³⁴ This is consistent with SAU's transformation agenda, as well as meeting the requirements of the Employment Equity Act of 1998 which required institutions to draw up employment equity plans and set equity targets against which future employment profiles would be measured.

practice where they hold membership on the basis of their research specialization, namely, their research communities. These research communities take many different forms: some are SAU-based research units which members of the department either direct or are affiliated to. These research units represent the archetypal accumulation of academic cultural capital. The successful units attract large research projects with multinational companies, generous funding, highly qualified and experienced staff, and top-ranking postgraduate and undergraduate students. For others, not affiliated with a research unit, the research community is a less visible but no less significant specialization network which includes colleagues (nationally and internationally), postgraduate students, conferences, journals, and sources of funding.

EF³⁵, as a field of study, was characterized in many of the interviews as vast and highly diverse. The ex-HOD noted that the differences between two different areas of specialization within the department were as great as the differences between accounting and physics (e-mail, 21/5/02). Another member of staff noted that demarcating the boundaries between EF and other fields within Engineering was “contested territory” (Ste 00-1). He contrasted the field of EF to, for example, Physics, where, he argued, there would be a fair degree of consensus about what should be taught in the undergraduate curriculum. In contrast, “EF isn't like that at all. It is much more diverse”. To illustrate this diversity, he noted that two of the top rated EF departments in the world had virtually no overlap in their undergraduate curriculum (Ste 00-1).

The diversity in the field was reflected in the wide range of specializations which the staff in the Engineering department represented. In fact, there was a common perception among the staff that there were as many specializations within the department as there were members of staff (Rob 00-1). The ex-HOD noted however that, as staff left or retired, there was a deliberate strategy to recruit new staff to an area of specialization already present in the department. The department was home to four broad areas of specializations although the alliances within these areas were relatively loose. One

³⁵ When I refer to the specific field of engineering which is the focus of this department, I will use the designation ‘EF’, i.e. engineering field.

interviewee characterized his area as a group of people who share “the same passage way and have similar interests” but at the end of the day “people tend to go their own way” (Tho 00-1). A member of another area of specialization characterized it as a group of people who were working with the same technology but applying it in very different ways. Thus there was very little actual overlap in their respective research projects. He commented, “Everybody forms a group independently!” (Pet 00-1). Illustrating this independence, the course convener noted that the assessment of the final year projects was one of the few occasions when members of the department came together as a group for significant business. All other activities (e.g. research, teaching) required very little interaction amongst colleagues (Dav 00-3).

The process of assessing the final year projects proceeds as follows: in May the course convener calls for topics from the academic members of staff. There is no screening of these topics³⁶; they are simply posted on a web-page and students select their top preferences. The convener then allocates students to supervisors depending on the topics they have selected. Students then have approximately three months to complete their research. The actual supervision process varies widely. Some supervisors meet regularly with their students individually or in groups, others make themselves available at students’ request. Some supervisors require drafts, others do not. Students submit their thesis by the middle of October. After they submit, students are also required to give an oral presentation, and make a poster presentation. However, only the thesis counts towards the final mark.

The supervisor is the first to assess the thesis. His mark and any comments he may have about the thesis³⁷ are recorded on the official marking memorandum, not on the thesis³⁸. Once the supervisor has completed his marking, the thesis then goes to a second internal marker allocated by the course convener. There is an attempt to allocate a second marker who has related expertise on the thesis topic. The second marker assesses

³⁶ I note this because pre-screening of topics is a practice in other departments in the Engineering faculty.

³⁷ In the year of my study all the academic members of staff were male.

³⁸ This is to ensure that the second and third markers are assessing a ‘clean’ copy of the thesis and are not influenced by the supervisor’s commentary.

independently of the first marker's mark. In addition to the thesis itself, the second marker is expected to attend the student's oral presentation, as well as the poster presentation³⁹. If there is a significant discrepancy between the first and second marker, the thesis is passed on to a third internal independent marker. In 2000 the HOD requested that the first, second and third markers meet to arrive at a consensus mark prior to the departmental examination meeting. This was to avoid the situation where internal markers argued with one another in the final meeting with the externals. These gatherings are referred to as 'consensus meetings'.

Finally, the theses are assessed by a team of external examiners. (I describe this process in more detail in 5.4.3.1). These examiners are chosen for their expertise across the areas of specialization. In 2000, five of the external examiners came from three different South African universities and two from industry. Final discussion is held and decisions are made at the departmental examination meeting where the entire community of assessors, both the internal and external, is present. This whole process, from the supervisors' assessment to the examination meeting, takes place in a week. Thereafter some final adjustments may be made at a later departmental examination meeting where all the students' final year marks for all their courses are considered.

The procedure for assessing the final year undergraduate thesis has evolved over a number of years under the leadership of the ex-HOD. He explained that a number of changes had been instituted to the system "to make the mark more fair and more representative of the value of the thesis" (Rob 00-2). For example, in the past supervisors would assess their students' theses and then simply ask a colleague to "check it" (Rob 00-2). There was no external examination of the process. While there was seemingly more agreement in those days between the first and second marker (though it was impossible to verify), the current HOD felt that there was possibly "too much agreement" (Ken 00-2). Commenting on the previous system, he noted that it was too "palsy", with supervisors selecting year after year the same second marker. This resulted in markers "understanding each other's marking system and being able to guess very accurately what

³⁹ It was my impression that the majority of staff in the department did attend these presentations.

the (other) examiner is going to give the thesis” (Ken 00-2). From his perspective this constituted “bias”. In contrast, in the current system where second markers are centrally appointed and assess independently, he noted, “If you give a thesis to someone who doesn't normally second mark your thesis topics, they don't know how you think ... I think that's quite a useful exercise. Because if nothing else it stirs the pot a bit and makes people think about what marks they're giving. Once you've got the two marks in front of you, (you) now have to worry about what the fair mark is for the thesis. If they're close, it's an average. If they straddle (a class) or clash there's bit of a problem...” (Ken 00-2).

This innovation of a second independent marker did result in some very wide discrepancies between supervisor and second markers. The ex-HOD believed that these discrepancies resulted from second markers reading the thesis “cold” (Rob 00-2). Thus the system was improved by adding more contextualization for the second markers. An oral presentation was introduced, not to test students' oral skills but to give second markers more information about the thesis. In order to get students to take the oral presentation seriously, it was given a mark but these marks were not included in the final thesis score. It was additional information that could be used in borderline cases, that is, when decisions hovered between a pass and a fail or between classes. Thus while the students had several opportunities to display their work, the final mark was “100% for the thesis” (Rob 00-2). The most recent innovation, again to provide more contextualization for the second marker, was that students were encouraged to invite the second marker to a demonstration⁴⁰. The ex-HOD, who was responsible for introducing this innovation, felt that this was “tremendously valuable. It answers a lot of questions, makes a big difference and helps you mark the thesis. It gives you some of that insight that the supervisor has got” (Rob 00-2). Another member of staff commenting on the value of this demonstration noted, “You see the project working at its best, which is good, I mean its fair to the student that he⁴¹ should be allowed to show the thing at its best” (Ant 00-2).

⁴⁰ I was not able to establish how many students actually took advantage of this opportunity. I was under the impression that not many did.

⁴¹ While not all students were male, I noted that it was common for staff to refer to students using masculine referents.

5.4.2 Dimensions of 'disposition'

The analysis of the Humanities case study seemed to suggest that depending on their roles (i.e. whether they were supervisors or non-supervisors), assessors will prioritize different assessment functions and different contextual considerations will be privileged. These different interpretive matrices may lead to different interpretations, as noted between Mary and Keith. In this second case study, I now turn to a more in-depth exploration of these differences of interpretation.

The Engineering community of practice presents a particularly good opportunity for the exploration of difference for two reasons. Firstly, their validating practices open up these differences for scrutiny given that there are clearly assigned assessor roles (e.g. supervisor, second marker) and given that all staff come together to discuss these differences in the departmental examination meeting. Secondly, the year before I interviewed, this community had experienced a particularly problematic examination meeting which, I contend, served, in part, to de-naturalize their practices. (I elaborate on these problems in 5.4.2.1). The problems which emerged from the 1999 examination meeting became a useful springboard for staff to explore in interviews their own explanations for why differences emerge between markers. On the basis of interviewees' own explanations, I draw out some of the features of assessors' interpretive matrices⁴². I stress again that interpretive matrices are highly complex. My study makes no claim to comprehensiveness. Certain features have been highlighted and others have been left out. The features which emerge from my study are constituted, in part, by the particular contexts which I explored, as well as by the particular questions which I pursued; in other words, a different set of interview questions would have possibly raised a different set of features. I turn to a discussion of three features of assessors' interpretive matrices: specialization, experience, involvement.

⁴² These features emerged from my NVivo analysis (described in 4.4.2.3.3) where I clustered the interview data into 'reasons' offered for discrepancy between markers. The 'reasons' which appeared most consistently across interviews were the ones which I pursued in further interviews. These then became themes which developed into features of the interpretive matrix.

5.4.2.1 Specialization

The ex-HOD noted in an interview that although the system of assessing the final year projects had worked “fairly well” over the years, “the past two years (1999 and 2000) there have been some problems” (Rob 00-3). An exploration of these “problems” exposed some divides in the community about what constitutes a “real engineering project”⁴³ as well as what constitutes legitimate evidence for the assessment of projects. These divides manifested themselves in heated debates at the 1999 and 2000 departmental examination meetings. In 1999 the contestation centred on some theses supervised by two members of staff, Richard and Victor. In the case of the former, he had given two of his projects first class passes whereas the recommended marks of the second marker and the external examiner were upper seconds. In the end he reluctantly accepted the recommended marks. (I return to his justification below). Victor’s marks however were consistently higher than those recommended by the second markers and the external examiners, and he was unwilling to accept these recommended marks. The HOD at-the-time decided to send Victor’s controversial theses off to a colleague at another university for yet another independent assessment. This independent assessor “vindicated” the lower mark of the second and external examiner (e-mail, 10/5/02). One of the members of staff gave his account of what happened:

So maybe the supervisor who had clearly in his mind what exactly it was that he was setting, gave it (the thesis) a good mark. It would then go blind to a second staff member, who would look at it and say ‘This is a strange thesis. It is just an essay. What has the student done? He hasn't done anything, (though) he has written fairly well, so we'll give it a 50. We'll pass it, we can't fail it’. But meanwhile the first examiner had given it (the project) a 70 because it's done exactly what he wants. Then it goes to a third (external) examiner who is in a quandary, who says ‘Well, if this were in my university, I would fail it because it is not a thesis at all. But he (the student) seems to have done what the supervisor wants, so we had better give it 60’. And that happened again and again. And external examiners who were in that

⁴³ The 2000 Faculty of Engineering handbook describes the thesis as “an opportunity for the student...to tackle a real Engineering project” (p. 177).

position, who had to make that judgment, complained and said 'We think there should be more clarity on what is expected' (Ste 00-1).

Although the issue of what defines a "real engineering project" explicitly emerged as a problem for the first time in the 1999 examiners' meeting, Steven noted that this was not a new issue.

As I probed, some interviewees felt that the contestation in the 1999 examination meeting was an issue of content and form. Richard was part of the ES⁴⁴ area of specialization and some staff members referred to projects from this area somewhat disparagingly as "essays" or "studies" in contrast to a "design product". In trying to clarify this distinction the ex-HOD elaborated:

Generally an engineering project involves a design process – identifying a project, doing the design and completing the project. And it might be a piece of software that does a certain thing, it might be a piece of hardware. It could be something very practical...but generally it involves design. Whereas a number of projects were being offered (by ESS) which were a study: 'Investigate the efficiency of (a particular product). ...Investigate this as a viable product'. So it doesn't involve any design. It is a survey, it's a study (Rob 00-1).

Reflecting on the 1999 contestation, he noted:

We started to find big discrepancies between the supervisor and the second examiner, because the supervisor thought (these projects) were worthwhile, it must be worthwhile because companies like (X⁴⁵) want it to happen. Whereas the second examiner said 'But this isn't a design project, it is not what engineers (do), it doesn't test the design process. So it is lacking'. And interestingly enough the second examiners took the side of the externals, and in fact they advised against (these kind of) projects, because it doesn't test the design aspect, which is a big chunk of an engineering degree...The staff members concerned tried to defend this and argued

⁴⁴ While 'EF' refers to the field of study, I use the designation 'ES' to refer to an area of specialization within the field, i.e. engineering specialization.

⁴⁵ To protect anonymity of speakers, I have omitted any references to information or specialist discourse which would give away their disciplinary identity.

that if it is the sort of thing that people do when you go to industry, surely it ought to be a legitimate project” (Rob 00-1).

One of the staff members whose students’ theses were received unfavourably had a different view on what was legitimate. He described one of the particularly contentious projects from one of his “very good students” (Ric 00-1)⁴⁶. The student’s project had consisted of tackling a problem which had been troubling a particular industry for some time. He described it as “exploratory research” where the student needed to investigate a problem which would require a team of people and a number of years to tackle fully. But as a preliminary exploratory research, the project required the student to collect and synthesize theoretical, empirical and practical information which is intellectually complex. It was also logistically difficult to retrieve this information since “it’s not written down”, and “when one speaks to people...you don’t always get a consensus” (Ric 00-1)⁴⁷.

But the project was not assessed favourably by either the second marker or the external examiner. Commenting on their perspectives, the ex-HOD noted, “The external examiners (supported by a number of SAU staff members) argued that ‘just because industry employs engineers to do this work does not justify it as a suitable engineering theses topic’.” He argued that from their perspective these theses lacked the “rigour” of the “‘design and build and see how well it compares to your predictions’ thesis”. The external examiners concluded that while this type of thesis might be “acceptable”, it would be difficult for the students to achieve good marks (e-mail, 21/5/02). From Richard’s perspective this had a very unfortunate consequence: “One of our top students had done an important piece of work and got a second for it”. As far as Richard was concerned, “he had been done an injustice” (Ric 00-1).

⁴⁶ To justify his confidence in this student’s ability, Richard noted in a later interview that the student went on to Oxford University on a Rhodes scholarship.

⁴⁷ As further justification for the kinds of projects which ES set, Richard noted (in response to an earlier version of this analysis) that there are very real practical constraints (e.g. the cost and size of equipment) which made it impossible for those in the ES area of specialization to produce the kind of ‘design, build and test’ type of projects which some staff were arguing for.

The contestation was clearly not one of form and content only, but as Stephen noted, also methodological approach. “It isn’t only in the actual subject matter; it is how one approaches the subject matter” (Ste 00-1). He describes the kind of projects set in his area where students “build a gadget”. They spend all their time “getting it to work” and then “quickly write a report”. In contrast the ES projects “look more like a sociological exercise to me than an Engineering one ...it seems to draw very little on anything that they have learned in their degree”. Anthony broadens the contestation even further, beyond form, content and methodology, to the very nature of engineering as a field of study. He argues, “A major cause of disparity comes back to that personal understanding of what constitutes engineering, what constitutes a thesis, and what constitutes good work”. Citing, as an example, one of the ES topics which focused on the “economic aspects” of some engineering problem, Anthony commented, “A few years back it made me raise my eyebrows a bit, ‘Where is the engineering?’. But now I come to realize that it is engineering. Engineering and financial aspects are fairly inextricably entwined for them (ES staff).” He added, “But not everybody shares my broadness of view” (Ant 00-1).

Allan offered further insight into the contestation. He agreed that there may be different value systems which are shaping, on one hand, what Richard (and the industry with whom he is closely linked) would understand to be a ‘good’ project, and on the other hand, what other members of the community of practice would consider to be a ‘good’ project. The problem is that neither proponent is being explicit about those values. Allan argues that Richard is locked into a “particular perspective” which is “exceptionally important”. This perspective shapes how he designs his assessment and his projects and “industry is happy with that”. But there has to be a “valuing” of the objectives of that project experience; “(these objectives) have got to be made explicit” (Ala 00-1).

Although the externals vindicated the lower scores, the community did not achieve resolution on some fundamental issues: what is “real engineering” and what is a legitimate project? The ex-HOD noted ambiguously, “But we have now, I think we have

resolved it...I'm not sure we have resolved it" (Rob 00-1). One of the staff members whose students' theses were received unfavourably was less optimistic. Reflecting back on the contestation over what a thesis is, he noted, "It is a definite problem at the moment which is still outstanding and hasn't been resolved" (Ric 00-1). Another member of staff commented in response to whether the problem had been addressed, "There has been some tea time discussion, that's all I'm aware of. I think it might be quite difficult to arrive at consensus" (Ste 00-1).

Thus one of the features of assessors' interpretive matrices which emerges from interviewees' explanations is their very particular disciplinary specialization. Different specializations tackle different kinds of research problems from different methodological and epistemological approaches. These approaches will privilege different kinds of evidence, and it also emerges that communities of practice have ways of valuing particular kinds of approaches and marginalizing others.

5.4.2.2 Experience

As the interviewees reflected in particular on the 1999 examination meeting and more generally on the differences between markers, another feature of assessors' interpretive matrices which emerged was experience, that is, the extent to which individual markers have been socialized into the community of practice's 'ways of seeing'. Thus one of the theories for why (in some cases) such wide discrepancies emerged between markers was that new staff had not yet developed the common professional vision of the community⁴⁸.

The HOD felt quite strongly that the extent of difference between first and second markers was directly related to the number of new staff, and the department had had a number of recent appointments. His view was that it is "relatively easy" with experience to decide whether a project deserves a first, second or third class award. But he conceded, "not the first time". The problem of new staff was, he argued "symptomatic of university employment...when a lecturer gets appointed here there is very little guidance

⁴⁸ It is not surprising that this 'theory' emerged given that one of the members of staff, Victor, who had been at the centre of the 1999 controversy was new to the department.

given to the person as to how to go about teaching, how to go about assessing, how to go about marking theses. If one comes in from an industrial environment, or another university environment, or another department within the university, the chances are that the things that one looks at and looks for in a thesis will be different..." (Ken 00-1). Similarly another senior member of staff noted, "I think it is very hard, if you haven't assessed something like that before, how do you go about it? I don't know how you do it, they (new staff) don't get taught, they don't get instructions and they've got the sheet (marking memorandum) (but) they can interpret the labels which ever way they want to" (Pet 00-1).

To the department's credit, new members of staff who were inexperienced as supervisors were given a light marking load, and in some cases shared a supervisory load with a senior academic. Nonetheless many of them admitted to struggling. Kevin, who was supervising for the first time, admitted to being unclear about how to go about the marking. He raised the dilemma of the student who does "good research" but produces "a fairly shoddy report". As a supervisor, "you know...this is a very competent person. You would employ them. You would welcome them into your research group for post-graduate work." But Kevin queries, "Are we simply trying to assess the report that they have produced or are we trying to grade them as individuals, as a measure of competency in carrying out a project?". In my first interview with him, he was hoping to find out what the "official policy" was on how to formulate a mark (Kev 00-1). When I interviewed him again several months later (after he had begun to assess), he noted, "I asked our HOD and several others, on what basis the thesis is judged...there actually are no guidelines" (Kev 00-2).

David gave a personal illustration of how his socialization as a marker happened. As a new member of staff he recalls feeling very unsure of his marking initially. In his first year he was "higher" than everyone, his second year he was "lower". A senior member of staff assured him that this "pattern" happens to everyone. "This is the way in which people learn to adjust their marking," they told him. "You get in, you see 'oh this is how it's done', and you change..." David points on the spreadsheet to the scores of a new staff

member who had given a student a 25%, "I mean, he won't do that again! He won't go there again and give a student 25! If he really wants to fail him, he'll probably give him around 40 because he will get to see that people don't do that". When I commented that this trial-and-error system might be a somewhat publicly embarrassing for a new member of staff, he responded, "Nobody has to tell you anything. I mean...you learn. You are in a place" (Dav 00-2).

Thus the interviewees' accounts suggest that assessors' interpretations are born out of experience; that through membership in a field, over time, one becomes socialized into a community of practices' 'ways of seeing'. As one of the members of staff noted, "one becomes imbued with a form of thinking" (Ric 00-1). The dilemma, of course, is that novice assessors have not yet developed this vision, and thus seek out an "official policy" (Kev 00-1) and guidelines which as the HOD noted "are not written down" (Ken 00-2).

5.4.2.3 Involvement

I turn now to the final feature to emerge from interviewees' explanations. Consistent with what emerges in the Humanities case, the analysis of the accounts suggests that the supervisor's involvement in the performance constitutes a significant feature of their interpretive matrix. That is, the supervisor is invested in the topic. The supervisor has access to a wide range of different performances which other markers do not have. The supervisor is not simply responding to a 'text' but to a person. What emerges from the Engineering interviewees, however, is not simply an acknowledgment of these differences of perspective (between, for example, the supervisor and the second marker), but a problematization of these more external perspectives which are often referred to as 'cold' or 'blind' perspectives.

The supervisor's involvement in the performance begins with the topic. In some cases the topic is directly related to the supervisor's research. Those supervisors who direct research units often set topics which contribute to the overall interests of the unit. Some topics emerge out of a joint research interest between the supervisor and a partner in industry (as noted with Richard's projects). Thus from its very inception the

supervisor has a particular intellectual interest in the project which the other markers may not share. As the topic designer the supervisor has some idea of the objectives, scope, possibilities, limitations (often based on previous attempts at similar projects) which contribute to a particular set of expectations. As supervisor, someone who knows the student, he may also have particular expectations. One interviewee noted, "I know for the student (with) whom I was familiar... I know when I came to read his thesis, I already had a picture in my mind (of) what I thought he was capable of before I read that thesis" (Kev 00-1).

One interviewee responded to this issue of the supervisor's involvement with this insight:

You read the acknowledgments. On virtually the first page of the document the writer acknowledges the contribution of the first marker. Because that's the supervisor, immediately the supervisor is tied up in the work, so the first marker can't stand right back and say: 'I am objectively assessing this report into which I put a whole lot of effort for the past three months'. You have got to be coming at it from a different direction. So the second marker is much more detached from it, (pointing to thesis for which he is second marker) this is a report from anybody. I don't know the person.... I'm not deeply involved in this project and thinking 'I know how it should have gone', or, 'Ja, he's done very well'. (pointing to a project for which he is the supervisor) This student here, I am involved. I identified this as a problem in the first place. I guided him in the study of it and here on the acknowledgment page: 'Gratitude to my supervisor, my first marker for his input'. It would be foolish to think that the first and second marker can both be independent. The first marker is deeply involved in the project, every time (Tho 00-2).

Does this investment impact on the supervisor's assessment of the project? "It should affect the marking because I, as the first marker, will give a final mark which depends on what I saw of the student, how he performed. I know how many of the ideas were mine and how many, if any, were his, for example" (Geo 00-2). Another marker noted, "The supervisor will have spent 2 to 3 months agonizing with the student over

various problems and during that time you get to understand how well the student deals with technical issues, how well they formulate them, how well they look at solving them, how quick they are at actually implementing the solutions and so on. And that definitely has a bearing (on the assessment)” (Ken 00-2).

Further evidence for supervisors’ involvement in students’ projects was noted by several interviewees; they are sometimes responsible for problems which students encounter. Anthony gave a poignant example of this. He set a topic for a particularly “fantastic student” but as the student got involved in the project, Anthony realized that there was not enough “meat” in the project for the student to get a “good first”. He admitted, “This was substantially my fault...a case where a shortcoming on the supervisor's part effects the student’s mark.” Anthony felt that it would be a “travesty” if this student did not get a first class honours⁴⁹ and so he asked to see the student’s thesis outline “just to make sure he had put in everything that he could put in”. The supervisor admitted that this might appear “manipulative”, as well as “unfair” to other students whose work he did not get involved in to the same degree. However his bigger concern was that this student’s overall capability would not be fully appreciated by the second marker. “(The student) would have killed himself in order to get to point X. And the trouble was that there wasn't that opportunity (in this project)”. He acknowledged this as a case where “a shortcoming of the project or a shortcoming of the supervisor causes a student to, let's say, get stuck, or not display his full potential. And we need some mechanism for, for bringing that into the marking process” (Ant 00-2). Another member of staff, Victor, felt even more strongly about this issue. “The blame that is on the failing of the student really should be put on the supervisor and not the student.” He pointed to the terms of reference in the front of the thesis which the supervisor, in theory, negotiates with each student. If the student follows the supervisor’s specific instructions, “Who should be failed?”, he asked (Vic 00-1).

As I probed these differences in assessor roles and the influence on the markers’ assessment, interviewees expressed varying degrees of reservations about the soundness

⁴⁹ The Engineering faculty has what it refers to as the ‘honours ruling’. In order to graduate with honours or distinction a student has to get a minimum number of points on their thesis.

of the interpretations which came from those more distant from the final project, for example, second and third markers and external examiners. As noted earlier the rationale underlying the existing validation system was to seek for greater consistency across multiple independent interpretations of performance. The theory is that the second marker comes to the thesis with expertise in the thesis area, with some contextualizing information, but with less investment in the student and the project, and thus, in theory, is more 'objective'.

However interviewees commented on a number of aspects of the official procedure which did not always work out in practice. While in theory the second marker is supposed to be knowledgeable in the area of specialization, in practice, given the number of theses, the limited number of specialists in a particular area, and the need to balance workloads, second markers may not be as knowledgeable in the area as the supervisor. Even if they have expertise they may not know precisely what the supervisor had in mind, as the objectives of the project may only be implicitly signalled in the project.

Nearly all the staff commented on the difficulty of the second marker role. One noted, "You don't know the student" (Ric 00-1). Another commented, "My biggest problem is to find out exactly what the student did" (Geo 00-1). They also pointed to a number of potential 'blind spots' which the second marker may experience. For example, projects which depended on software were particularly difficult for second markers to judge. A student might produce something which is "really quite fantastic from a hobby point of view but there's a question as to how much of that design and manufacture has actually been done by the company....I think that the supervisor would know and see that but the second marker may not, it depends on their expertise" (Ken 00-2). George confirms this, "The danger is because I am the second marker and didn't know too much about the computer aspects of it, I was perhaps too easily dazzled by the computer expertise which these students display..." (Geo 00-2). Another 'blind spot' which the second markers may experience is that a student's language skills might conceal conceptual problems. In other words, the student's actual understanding of the material is very weak, but the thesis is a "glowing communication in the language that the

student has mastered superbly”⁵⁰. This is a “definite problem” which a second examiner may or may not see through (Ken 00-2).

Another member of staff was more critical. In response to a query about a particular second marker’s assessment of one of his students, he noted, “I think he (the second marker) failed to interpret, apparently he is not so sure about what (a certain procedure) is...I think that one of the (reasons) that students are failing here is not because they are supposed to fail or students are passing, not because they are supposed to pass, but simply because the second examiners, they don't know, you see. Theses...they are just thrown at us. ‘Mark this. You are the second marker’, when in fact you don't know” (Vic 00-1).

Assessors also problematized the extent to which second markers’ viewpoints were independent and the extent to which this independence was a good thing. David noted for himself that part of the psychology of being a second marker was to “avoid conflict”. He explained that if he knew that a particular first marker had a reputation of being generous, he would assign his own marks accordingly. “In general I wouldn't want to pick up a fight, I have enough work to do” (Dav 00-3). In contrast Stephen describes a reverse adjustment, “Some of my colleagues are known to be generous in their marks, and certainly if I get a thick one to mark from one of them, I think of pulling it down and vice versa.” David concludes, “Basically what I am trying to get at (is) it's not as independent as, it's not like I go and look at this thesis and I mark it independently of who the student is or who the first marker is and what topic it is.” Stephen concludes, “I think there is a whole complex thing that goes on” (Ste 00-1).

Under the new system of independent markers, dialogue between the first and second marker prior to allocating a mark was “discouraged” (Rob 00-2). The ex-HOD commented, “It would be totally unprofessional to go to the first marker and ask him what he got” (Rob 00-2). Some were sceptical, however. “People talk of transparency ...

⁵⁰ Kenneth was referring to students who are fluent first language speakers of English, who have benefited from secondary schooling which has equipped them well for oral/written presentations, and in addition they may be technologically very competent with Power Point presentation and other software packages.

first and the second marking independently, but you will never know what happens” (Vic 00-1). Others were not sure that it was appropriate, “One talks about trust ... it means the staff are going to assess things as honestly as they possibly can. But they should be able to go to their colleagues and say, ‘Tell me a bit about this’, or otherwise say ‘I don't know, I think this looks like maybe a 63 or 64. What do you think?’ or ‘What have you given it?’ And in fact amazingly often it is very very close. And if the other person said ‘Hell I thought it was worth a first’, then I think you should say ‘Well, perhaps I've missed something’ and discuss it and go away and think about it again. And if you honestly don't think it is worth a first, then you give it to a third person and those can be difficult situations but that is the way to try and resolve that” (Ric 00-1). Anthony problematizes the notion that an independent judgment is likely to be the most fair judgment. “I think it might be a mistake sometimes because in a sense what we are saying is that we are not going to make a fair independent assessment if we communicate stuff outside of the thesis, you know, whereas the prejudice that you get by giving extra communication in context is virtually never against the student. I would say it is almost always in the student's favour so I have a bit of a problem with that” (Ant 00-1).

Similar criticisms were raised against the external examiners. The HOD noted that South Africa is a small country: “It's almost impossible to find enough independent people with whom one is not doing research in one's own area” (Ken 00-1). Problematizing the perspective of the external examiners, Anthony comments, “Many of my theses have a very technical content and I find that's not acknowledged by external examiners and it's frustrating for me. For example last year I had a student who did a piece of brilliant technical work without much theoretical under-pinning and did so much superb technical work that I thought he had to get a first. But firstly the external examiner never sees the hardware, he just sees the thesis, but secondly he said ‘There is very little actual theoretical knowledge displayed in this thesis... to get a first you have to be a good theoretician as well as everything else’ and I couldn't argue with (him)” (Ant 00-2).

In summary, an analysis of these explanations for difference of interpretations suggests three features which contribute to the constitution of assessors' interpretive

matrices: an epistemological orientation that comes from the specialization, particular 'ways of seeing' that are developed through experience, and a privileging of evidence based on different levels of involvement. As with the Humanities analysis, I now explore through the departmental examination meeting discourse how different assessors' roles constitute different interpretive matrices. The interviewees' accounts suggest that supervisors do draw on different kinds of evidence than second markers or external examiners and that these different evidential bases will result in different interpretations. Does an analysis of examination meeting discourse support this? Again as in the analysis of the Humanities discourse, I am interested in what is at stake in these negotiations around difference, and finally, how these differences are resolved.

5.4.3 Where difference plays out: the departmental examination meeting

5.4.3.1 The constitution of external privilege

The assessment of the final year theses culminates in a day-long series of meetings. The first meeting is a gathering of the external examiners. The procedure in 2000 was as follows: the external examiners gathered in the staff room where they had available to them all the theses, the internal examiners' reports, and a marks spreadsheet. The spreadsheet consisted of the student name, thesis topic, the first marker's mark and initials, the second marker's mark and initials, the calculated difference between the first and second mark, (and where given) the third marker's mark and initials, and the consensus mark⁵¹. The spreadsheet of marks was arranged by difference between first and second marker in descending order, in other words, the widest differences and thus most potentially problematic theses were at the top of the list. The externals started at the top of the list and focused on the theses that had the widest discrepancies between the first and second marker, as well as those which were borderlines. They identified approximately 40 theses for moderation out of a total of 84. These theses were divided among the external examiners according to their specializations (on average 5-7 theses each), and a few theses were looked at by two externals⁵².

⁵¹ The consensus mark is the outcome of the consensus meeting, that is, the meeting of the first, second, and third internal markers which happens prior to the examination meeting.

⁵² It would seem that external examiners 'doubled up' on theses which they felt might be particularly sensitive cases, as evidenced by the first dispute (discussed in section 5.3.3.2).

The general feeling among the staff interviewed is that the role of the external examiners is two-fold: firstly, to arbitrate where there are discrepancies between internal markers, and secondly, to assist the staff in selecting the top thesis. One of the members of staff described the arbitration process: "Occasionally you get a vast discrepancy and the way we work is, we don't fix those discrepancies. Those are (given) to the externals...they skim through the thesis and then from comments that are written down, they try and ascertain, well, they come down on the side of one or the other. And generally they pick one or the other of the sort of high or low, they don't average the mark or anything" (Car 00-1).

Although there seemed to be general agreement that the primary role of the externals is to arbitrate between internal markers, what was less clear is what procedurally should happen when an internal marker contested an external's recommended mark. Where did the final authority lie? The HOD argued that there was no single person with whom authority resided; it was the departmental meeting's judgment of the strongest argument. Another member of staff also denied that the externals had final authority. He argued emphatically that if he needed to defend a student who had done "a good piece of work", he would. If he is challenged by the external, "I say, 'I can't accept this'...that's the reason why we are there, to discuss those issues, not just to keep quiet" (Pau 00-1). The ex-HOD noted that there was, in fact, no official policy on this matter; "It wasn't really a problem for us until the last two years...we seemed in the past to get to consensus" (Rob 00-3).

In the afternoon all the external and internal examiners came together in the examination meeting to consider the marks recommended by the externals. The gathering was chaired by the HOD who was chairing for the first time. Everyone had a copy of the spreadsheet of marks (excluding the externals' marks). After introductions the chair opened the meeting by reminding the gathering that the "intention" of the

meeting was to “go through the results of the thesis projects” (DEM 10⁵³). He reminded the participants of the procedure which had been followed. The chair then started at the top of the spreadsheet (the same one ordered by discrepancy) and read out the student’s name and the series of marks allocated to the thesis: the first mark, second mark, the third mark and consensus mark (where given), and finally the external’s mark (see Appendix 4 and 5, line 1-2 for examples). If anyone had an objection, it had be made at this point. “Otherwise”, as the chair noted at the beginning of the meeting, “we will go with the mark recommended by the external examiners” (DEM 14).

Despite the HOD’s perception that authority lies in the strongest case, an analysis of the meeting discourse as well as the exchange structure⁵⁴, suggests that there may be particular structural constraints in the genre⁵⁵ itself which affect the extent to which supervisors feel free to present their case. The discourse of the meeting was relatively formal. This formality was evidenced by particular meeting protocols which were adhered to, for example, the chair was often referred to as “Mr. Chairman”, participants often raised a hand to indicate that they wanted to speak, at various points the chair asked the meeting “to rule” (B 35) in favour of a particular decision. This is a discourse which gives pride of place to administrative efficiency and unambiguous decision-making. At various points in the meeting, members of staff, particularly one new member of staff, attempted to use the meeting to explore some of the ambiguities of the procedure and assessment issues more generally. The dominant administrative discourse of the meeting however, constructed these explorations as divergences, and the chair urged the meeting back to the business at hand.

The typical exchange structure for these disputes is that the chair reads off the list marks, starting with the supervisor’s mark and ending with the external’s recommended

⁵³ References to the departmental examination meeting text which are not included in the appendix are referenced as ‘DEM’ and the number refers to the Nvivo text unit.

⁵⁴ As noted in Chapter 4 (4.3.3) ‘exchange structure’ refers to the pattern of turn-taking between participants of a conversation. Analysis of exchange structure can illuminate interactional control, i.e. are the relations between speakers symmetrical or asymmetrical in terms of degree of the control? (Fairclough 1992).

⁵⁵ As noted in chapter 4 (4.3.3) ‘Genre’ refers to “a relatively stable set of conventions that is associated with, and partly enacts, a socially ratified type of activity” (Fairclough, 1992).

mark. As noted above, unless someone spoke up at this point, the external's recommended mark became the 'final mark'. If a supervisor (or second marker) chose to dispute a mark, the onus was on the supervisor to justify why his mark was different to the external rather than the onus being on the external. The external then offered his/her justification. Thus in terms of the exchange structure the external often had the last word (as illustrated in dispute A, line 32-35).

It emerged in interviews prior to the departmental examination meeting that supervisors were conscious of a difficult dynamic that the meeting sets up. In contrast to Paul's adamant claim that he would "not keep quiet" (Pau 00-1), other supervisors were more circumspect. Anthony noted, "The trouble is that having invited these examiners...one can't then contradict them...they are the final arbiter of what goes on" (Ant 00-2). One of the staff members who had been at the centre of the 1999 contestation (discussed in 5.4.2.1), noted in response to my question about whether he felt he could defend his marks, "It was very hard in a staff meeting of about twenty people, discussing a hundred student project marks" (Ric 00-1). One of the more senior members of staff noted that "it takes quite a lot of chutzpah to stand up and say you don't agree with this mark allocated by the external examiners and they are all sitting around the table" (Pet 00-1).

Another constraint on supervisors presenting their case is what the ex-HOD referred to as the 'wounded supervisor syndrome', that "you just argue for a student because they are your students" (Dav 00-3). Anthony illustrated from his own experience the supervisor's dilemma:

Last year I had two students who got upper seconds and both should have got firsts, I believe. And both on the average of everybody's marks were sitting at upper second level and...you can't argue every one of your students up, even if they all deserve up. In that case I had to say, 'Gee I can argue fiercely for one of these guys and if I argue fiercely for both, I'll run the risk that neither of them will get promoted because I will just come across as a person who wants all the students pushed up a grade'. And so you do a kind of triage you know, you kind of decide maybe I can get him up, maybe I can't

get him up, which is a bit horrible. It's not an ideal situation (Ant 00-1).

Thus I would argue that while officially there was no clear position on where the final authority lies, and while some interviewees argued quite strongly that final authority lay in the strength of the better argument, an analysis of the actual genre suggests that the meeting itself (in its discourse and exchange structure) may have in fact privileged the interpretation of the external examiner, effectively granting the external the last word. (I will illustrate this in the next section when I turn to the analysis of the actual disputes). Interviewees' accounts of their own experiences also serve to endorse these constraints. "I will speak out...but you need to be careful," notes David (Dav 00-3).

5.4.3.2 What's at stake?

It needs to be stated at the outset that the overall dynamic of this particular examination meeting was one of tension. One member of staff declared that the meeting was one of the worse that he had ever attended in his long career at SAU (field notes, 16/11/01). I turn to an analysis of the validating arguments with a particular interest in some of the underlying causes of this tension: what is at stake in the negotiations of these differences?

Before turning to the analysis of the actual arguments, the quantitative data reveals that some degree of discrepancy between first and second markers is a standard feature of more recent departmental examination meetings. Table 1 summarizes the discrepancy patterns for the final year projects over a three year period by class difference⁵⁶. This data indicates that there appears to be less discrepancy between first and second markers in 2001 where a total of 97% of the projects have a difference of one class or less compared to only 82% and 88% in 1999 and 2000 respectively. However a discrepancy of one class (which even in 2001 is 41%) is not insignificant especially when it is the difference between a fail and a third class pass, or an upper second and a first class pass.

⁵⁶ A 'class' refers to the category of mark which the student is awarded, for example, first class is an award of 75% and above, a second class award is between 60% and 74% , a third class award is between 50% and 59%. Thus a 'one class' difference is, for example, where one assessor has awarded the project a second class and another assessor has awarded it a third class.

Thus this analysis takes discrepancies between assessors as a given. The point of interest is what underlies these differences, what is at stake in their negotiation, and finally how these differences are resolved.

Table 1: Final year project discrepancies between first and second markers: (1999-2001)

Year	Same class	One class	Two classes	Three classes	Four classes
1999 T=102	38% (39)	44% (45)	17% (17)	1% (1)	0
2000 T=84	49% (41)	39% (33)	7% (6)	2% (2)	1% (1)
2001 T=103	56% (58)	41% (42)	3% (3)	0	0

My analysis of the validity arguments focuses on the disputes since it is through these collegial exchanges that justifications for assessments are exposed. 'Dispute' refers to the situation where one of the assessors, usually a supervisor, would contest the external examiner's final recommended mark. There were 16 such disputes in 2000, and not surprisingly, the majority of them were disputes over the award of a first class pass or the award of a failing mark. Out of the 16 disputed cases in the meeting, six of them concerned disputes over the award of first class passes (i.e. projects where the supervisor awarded a first class pass and the external marker pulled the mark down). Five of the disputes concerned the award of a failing mark (i.e. projects where the supervisor passed it and the external examiner failed it). In all but one dispute, the external examiner's recommended mark became the final mark.

Crucial to the contextualization of the disputes which I focus on in more detail is the opening dispute. This centres around a thesis which the supervisor, Paul, awarded an 81%, the second marker, Victor, awarded a 40%, and the external examiners (two of them in this case) awarded a 70%. No third marker had been appointed since the course convener did not feel that there was anyone in the community with sufficient expertise in this area of specialization. In his justification for his mark, Paul gives a detailed explanation of the task and the various steps which the student had taken in order to complete the task. Paul stressed that the work was "his (the student's) own development,

nobody else's", "he worked independently...and very little did he draw from the supervisor...in order to do what he did. So based on that...this is in fact a first class pass" (DEM 20).

There ensued a heated debate between the supervisor and the external examiners, who conceded that the thesis had many strengths, but "we felt that there were not many original ideas" (DEM 28). "I would classify it as a good report, I just thought that it fell short of being outstanding because the results, the discussions, that could have been done a little bit better. But it is a good report" (DEM 61). There was also a tense exchange between the second marker, Victor, and one of the externals, Ella. In addition to providing justification for why the thesis should fail (his main criticism seemed to be that the student had not accurately labelled the axis of his graphs), Victor appeared to be questioning whether the externals could do justice to their task. He says to Ella, "I don't know whether you went through it (the thesis) carefully. I'm quite aware your time is limited" (DEM 36). And again, "Your time is limited but if ...you go into the most technical details of (the thesis) and what the student was supposed to do...basically it's just a useless report" (DEM 40). She responds, "I cannot agree" (DEM 42). At a point of seeming deadlock, the chair turned to the ex-HOD for a 'way out'. He responded, "Look, we've got two external examiners who are world authorities in this field, and I, I suggest we respect their opinion, which is 70%" (DEM 65).

I would argue that this opening dispute which was characterized by confrontation between the internal examiners, as well as confrontation between the external and internal examiners, contributed to an atmosphere of tension which pervaded much of the meeting. In addition, reflecting back on this particular dispute the ex-HOD also argued that the fact that Victor's award was so severely out of line with the supervisor's and the external examiners' resulted in him being discredited. He was perceived as the "odd man out" (Rob 00-3).

The excerpts which I chose for analysis are the second and third disputes of the meeting (henceforth referred to as disputes A and B respectively). I chose dispute A

because it best illustrates the typical exchange structure of the meeting. I chose dispute B because it best illustrates a disruption to the norms which constitute the meeting. In the analysis I focus on the kinds of evidence which assessors draw on to support their justifications. I also focus on the facework, that is, the strategies which participants in the meeting use to protect their own face or to maintain or threaten others' face. As I noted in Chapter 4, one of the ways for getting at facework is through the analysis of modality, that is, the degree of certainty or conviction with which utterances are expressed. I noted how modality can signal both the degree of commitment to an utterance as well as the degree of solidarity the speaker has for other interactants. Thus participants' discourse may reveal the balancing of two desires: the desire to express a strong conviction (high modality) with the desire to maintain collegial relations (low modality).

Starting with dispute A (see Appendix 4), I focus on five units of text which correspond to each of the speakers' input in the dispute: (A5-8), (A 8-16), (A22-25), (A26-32). The chair begins with "the supervisor gave it (the project) 88, second marker gave it 72, third marker gave it 69, and the consensus mark was 75, and external examiners recommend a 70" (A1-2).

The first text unit (A5-7) cannot be re-constructed in its entirety due to poor audiotape quality of this section of the recording. (This prevents me from commenting on the modality). The supervisor awarded the student a high first class pass (88%). The evidence he gives in support of this mark appears to be on the basis of the interdisciplinary nature of the task and that the student "seemed to get around and sort out information from a number of difference sources" (A6).

In the second text unit (A8-16), the second marker, Robert, makes a strong case for the award of a first class pass, though his mark is not as high as the supervisor's. He had initially given the thesis a 72%, but had agreed in the consensus meeting to raise his mark to a 75%. The evidence which he offers in support of the higher award is that "it (the report) is entirely self-initiated, had almost no support from the supervisor whatsoever" (A11-12), "the student had worked entirely on his own" (A14-15). This evidence appears

to have been gathered from a discussion with this student where he heard “the student’s story of how he went through this project” (A10-11). Not only does Robert provide evidence for his justification but he bolsters his argument by offering his views on what constitutes appropriate evidence. He argues, “It’s actually very useful to speak to the student and to see how the student approached the problem” (A8-9). He concedes that had he read the report “cold” (A13), he might have also given it a 69 or 70 as the third marker had. But having talked to the student and seen what the student had actually done, “I was convinced to raise the mark from 72 to 75...and I think that it’s important to take that into account” (A16).

An analysis of the modality of this unit suggests that Robert felt relatively confident about his judgment. This conviction is indexed in the following lexical choices⁵⁷: “I was convinced it should be...” (A9), “entirely self-initiated” (A11), “almost no support” (A11), “I was really very impressed” (A14), and he repeats again that the student worked “entirely on his own” (A14-15), “I was convinced...” (A15).

The third text unit (A17-21) provides further insight into some of the values which shape assessors’ interpretations. Andre was not an assessor of this project and it is not likely that he had even seen the project. His input was to “bring out an important aspect” (A17) about engineers being “able to communicate to the outside world” (A18, 20). The inference is that independent work does not constitute sufficient evidence to justify a first if the student cannot “communicate to the outside world” (A18). He qualifies, however, by stating, “but I think both need to be taken into consideration” (A21). Presumably “both” refers to the student’s ability to work independently, as well as his ability to communicate. I would argue that the primary function of this qualifier is to minimize the chance of threatening the face of colleagues, the supervisor and the second marker, who have expressed views different to his.

⁵⁷ I have underlined words that signal modality.

The fourth text unit (A22-25) is that of the third marker, who has given the thesis a 69⁵⁸. The evidence he offers in support of his award is based on the “write-up itself, the thesis” which he argues “is not as good as it should be” (A22). He points to the “references” and notes that the “bibliography is not properly documented” (A23). But again he concedes in closing, “taking everything into account, yes, he did quite a lot of work” (A24). Again I would argue that this concession, like that of the previous speaker, operates as a ‘hedge’⁵⁹ to mitigate what might otherwise be perceived as threatening to others who do not share his views.

Finally, the fifth text unit (A26-32) is that of the external examiner who has recommended a mark of 70. His evidence in support of “not recommend(ing) a first” (A26) appears to be primarily based on the fact that “some of the parts of the project didn’t work” (A27). From his perspective a “distinction candidate” (A29) should have been able to anticipate certain aspects of the functioning of his project, and this student did not. He also challenges the supervisor’s appraisal of the “multi-disciplinary” detail (A31). An analysis of the modality of this text unit would suggest that the external examiner is committed to his view. There is little hesitation or hedging in his argument; “There are a few reasons not to recommend a first” (A26). X is a fairly simple process and the student “couldn’t get that to work” (A28). He concludes “70% is a reasonable mark” (A32). There is finality to his argument; there is no invitation for the others to disagree, nor do they.

Thus we see from the analysis of these text units that different interpretations of the student’s performance are being justified on the basis of different kinds of evidence, offered from the perspective of the different roles which the assessors hold in relation to the performance. Evidence presented by the supervisor and the second marker supports a first class pass. This is evidence from the nature of the task, the student’s independent tackling of the task, the student’s “story” of what he did, as well as a demonstration of the

⁵⁸ Since I did not sit in on the consensus meetings, I do not know whether the third marker actually agreed to the ‘75%’ consensus mark.

⁵⁹ A ‘hedge’ is a linguistic device by which a speaker avoids being compromised by a statement that turns out to be wrong (Oxford Concise Dictionary of Linguistics). These devices are often used to reduce the modality of an utterance.

actual equipment (Rob 00-3). Evidence presented by the third marker and the external supports a lower and upper second respectively. This is evidence on the basis of the report itself, the manner in which the references were documented, and the fact that aspects of the project did not “work”. We also note ways in which the various participants manage the potential threat to others’ face. The second marker clearly feels fairly strongly about his mark. The third marker does not agree, but presents his argument with a concession which serves as a hedge. The external presents his justification, politely, but with certainty. The first and the second markers do not contest his mark and the final mark stands at 70%. This discourse is privileging the interpretations of the more distant markers (the third marker and external examiner), at the same time negotiating these different interpretations in a collegial manner, that is, in a manner which minimizes the risk of threatening individual’s face needs.

In dispute B (see Appendix 5) we see once again how kinds of evidence are foregrounded from the different assessor positions. The external examiner has given the report a 70. His argument that the report does “not merit a first” (B26) seems to be based on what he considers to be an inappropriate balance of theory and practice in the report. The supervisor has given the report an 80 and supports his mark with evidence from aspects of the research process, for example, the amount of time and effort which the student expended, the complexity of the task, the fact that the student met the supervisor’s expectations. The supervisor notes, “I thought he (the student) did exactly what was expected of him” (B36). The second marker chooses not to justify his mark – originally a 65 but he appears to have agreed to a 76 in the consensus meeting. The third marker awarded a 72. He justifies his lower mark by agreeing with the external that the report is “unbalanced” (B48), “there is no in-depth discussion of the results and I think that is what you would expect from a first pass” (B50). Thus in dispute B, as in A, evidence presented by the supervisor, based on the task, the student’s effort, the achievement of expectation, results in an award of a high first class pass. Evidence which the third marker and the external examiner draw from the thesis results in an award of a second/upper second.

As in the Humanities analysis, the analysis of these disputes suggests that from the perspective of different roles, assessors inevitably privilege the different functions of assessment. The external examiners can only assess the thesis as a product. Their role is to measure the thesis against what they believe to be 'good' research at this level of the curriculum. Thus from the 'external' point of view the student concerned in dispute A is not a "distinction candidate" (A29) and the project concerned in dispute B "is not meriting a first" (B26). In contrast the evidence offered by the supervisors suggests that the mark is not only an interpretation of a product but also of a process. Interpretation of the process privileges contextual considerations such as the student's "self-initiative" and "time and effort", and the supervisor's "expectations".

How are these differences of interpretation resolved? In dispute A the external's perspective is privileged and I argue that the negotiation of these different interpretations is done in such a way to minimize the threat to individual and collective face, in other words, collegiality is maintained even at the cost of having one's professional judgment discounted. I now turn to the analysis of facework in dispute B where I demonstrate that in the interest of challenging the external perspective, one member of staff resorts to a number of discursive strategies which threaten both individual and collective face⁶⁰.

I noted above that the typical exchange pattern for the meeting is the initiation of a dispute by the supervisor who justifies his mark, followed by a response from the external examiner which often (though not always) constitutes the last word (as illustrated in dispute A). In dispute B Victor initiates the dispute. He informs the gathering of the outcome of the consensus meeting (between himself and David; it is not clear whether the third marker was present): "We agreed that a fair compromise would be around a 76. Honestly speaking I don't see why this student should be denied a first class, if the externals can justify that" (B6-8). Thus by asking the externals to justify their

⁶⁰ My analysis does not really get at the issue of why Victor feels the need to disrupt the norms of this meeting. I acknowledge that he was responding to a complex set of factors which went beyond the boundaries of the specific issues which I was exploring. By the time I had completed early drafts of this analysis, Victor had resigned from SAU and did not respond to any of my e-mails. I was thus never able to probe these issues to the extent that I would have liked.

mark first, Victor disrupts the typical exchange pattern, “if they can tell me where they think he missed it...please” (B8-10).

In addition to this disruption, implied in Victor’s request is a challenge to the fairness of the external’s mark; the student has been “denied” (B8) a first class pass. However, in contrast to later in the dispute, the challenge is not personalized, in other words, he is not speaking directly to the external examiner but presumably through the chair as is the norm in this type of meeting. His tone is non-confrontational. The significant pauses (B 8,9) suggest that this is a genuine request for justification.

The exchange which follows between Victor and one of the externals (who is not the one responsible for looking at this particular thesis) (B11-17) is a tussle about legitimate sources of evidence, on one hand evidence based on the “report”, on the other hand evidence from the presentation. It is not clear why Ella raises the issue of “presentation” (presumably a reference to the oral presentation which all the students have to give), but her point is to stress that the external examiners’ assessment (“ours”) “is based on the report” (B14). Victor does not accept her point (“no, no” B13) and speaks right over her, arguing that the presentation is an important source of evidence. The chair, however, interrupts him and endorses the external examiner, “We are basing this on the report” (B16) and he is chorused by another member of staff, “on the report” (B17).

The external examiner responsible for assessing this project presents his argument. It would appear that his main criticism of the thesis is that the report relies heavily on the theory and the literature, and not enough on “what was actually done” (B21). However, perhaps as an attempt to defuse the mounting tension, he foregrounds the positive features of the report and backgrounds any specific criticisms which would justify his mark of 70 instead of a first. In the first two lines he repeats “well” three times (although qualified) and again when he concludes: “Well, looking at the report it’s a well written report which explains the theory fairly fairly well, the theory of um (X) that’s actually explained quite well ...” (B19-20). He concludes, “So it was a lot of literature survey done, the work that was done was actually done quite well, but not meriting a

first” (B25-26). I would argue that this foregrounding of the positive aspects of the report and the “actually” which precedes “explained quite well” (B20) serve to soften what the external examiner perceives to be a face-threatening act.

Victor is not prepared to accept this judgment and comes back to present his evidence, as noted above, on the basis of the amount of time the student spend on the project, the difficulty of the task, the fact that the student has met the supervisor’s expectations. But he presents his evidence in an increasingly confrontational manner, “I thought he did exactly what was expected of him, so in this case, I mean, what is wrong?” (B 36-37). The pace of his speech picks up; the incidences of ‘whatever’ increase (four times in B38-39)⁶¹. His challenge shifts from impersonal to the personal: “the student has done a lot of hard work but then you find a way of saying ‘No, I don’t think he has done the whatever’” (B40-41). He is no longer speaking through the chair but directly to the external. In contrast to dispute A where the assessors hedge to avoid face threat, there is little attempt on Victor’s part to minimize face threat. His modality is very high. “No I don’t think, I don’t agree with you, not at all” (B41). There is emphasis on “not” which adds to the strength of his conviction. Nor is Victor simply disagreeing with the external examiner’s (and the third marker’s) mark⁶². He is challenging the fairness of the external examiner’s judgment: “If you are to compare (this thesis) with some of the (other) theses, don’t you think there is too much discrepancy here?” (B37-39). He repeats again later, “I don’t think that is fair” (B61-62).

The exchange structure gives the chair control of how and when a dispute will be brought to a closure. In contrast to Victor’s personal, direct and high modality discourse (“you find a way”, “I don’t agree with you” B40-41) the chair reverts back to the impersonal and indirect discourse of the formal meeting. “Could we ask the meeting to rule on it, are we going to accept the mark of 70 from the external examiners?” (B43-45).

⁶¹ Victor seems to have a somewhat idiosyncratic use of the word ‘whatever’. When I examined his use of the word in other contexts, for example when I interviewed him, he tends to use it when he is at a loss for the right word. This would suggest that the frequency of the word in these lines signals some degree of agitation.

⁶² It is interesting to note that the second marker (who had originally given the thesis a 65 and then “agreed” to a 75 in the consensus meeting, according to Victor) is silent in this debate, except to invite the third marker to offer his justification.

The chair becomes “we”. A response is requested from the “meeting” “to rule”. The third marker presents his evidence in support of the external examiner’s mark, “the report was unbalanced as mentioned” (B47-48). Victor then comes back once again to justify his mark (B 51). Despite Victor’s protest, “I don’t think that this is fair” (B61-62), the chair ends the dispute by effectively silencing Victor: “Thank you, Dr. ___” (B 63). He then continues immediately with the very indirectly formulated question, uttered firmly with fluency and finality: “I wonder if we could ask the meeting that we go with the mark of 70% from the externals?” (B63-64).

5.4.3.3 How differences are resolved

As already noted, the departmental examination meeting in 2000 was generally perceived to have been a highly problematic gathering⁶³. Reflecting back on this contentious gathering, Robert noted that typically at these meetings there was a “general air of politeness” towards the externals. In contrast, Victor had been “very direct...he said what was in his head”. Victor had in the examination meeting challenged the external examiners and the procedure. Robert notes, “Quite a few people were embarrassed by this totally frontal attack ... (they) thought it was insulting to the externals”. Given that Victor’s theses had been at the centre of controversy in 1999 (as noted in section 5.3.2.1) and again in 2000, he was perceived to be the “odd man out”. Because of this “credibility problem” quite a lot of staff members “leaned towards the externals”. Thus the way that differences were negotiated in the examination meeting was that “we agreed we would go with the external examiners ... the externals obviously would have the final say” (Rob 00-3).

Following the departmental examination meeting, there was a constructive attempt on the part of the department (spearheaded by the ex-HOD) to review the system so that this experience was not repeated in following years. In an interview the faculty’s educational development officer noted that although (in his view) there were many issues

⁶³ The departmental examination meeting was followed by a e-mail discussion amongst members of the department which was initiated by Victor. In this volley of correspondence Victor implied that he felt that there had been racist attitudes in evidence in the examination meeting. Later on in the discussion he retracted his statement arguing that he was articulating a “suspicion” rather than an “accusation” and that it was directed at the external examiners and not the department (e-mail, 13/11/00).

that needed to be resolved, the ex-HOD perceived “the bottom line” to be a lack of agreement about the role of the external examiners, specifically “a lack of consensus around who makes final decisions about what” (Ala 00-2). The ex-HOD articulated the problem in an e-mail to all staff: “The difficulty we have is when the internal examiners disagree with the external examiners, how do we arbitrate this?” (e-mail, 13/5/00). His initial proposal (to be discussed at a staff meeting a week after the examination meeting) was that the external examiner should be the final arbiter: “If the internal and external examiners are unable to agree, the meeting will uphold the external examiners’ mark” (e-mail, 13/5/00). However, after consultation with both SAU academic staff and the external examiners, he withdrew the proposal and proposed instead that a committee be established to review the existing final project assessment system (e-mail, 16/11/00).

It appeared that a number of staff members had changed their minds about whether the external examiner should have the final say, including the ex-HOD. Robert conceded that Victor “might well have been out of line”; “He’s new, not used to the system, not used to the students, not used to the standards” but he highlighted that “We’ve got a problem...we don’t know how to come to consensus between the external examiners and the supervisor who believes his marks are correct” (Rob 00-3). Victor had “one very good point” which he raised, both during and after the meeting. Victor was arguing, according to Robert, that a staff member has been:

supervising (a) student for 12 or 13 weeks and he's really got to know that student's work. The external examiners come in and they've got something like 10 or 12 theses each in the space of one morning. How can an external examiner, albeit an expert in the field be so confident that they're correct when they've had half an hour to assess a thesis, and what's more they've never been present at an oral and they've never seen the equipment? And they never interviewed the student, and the only thing they've got is a document. With an under-graduate thesis project with the time scales, that's not adequate⁶⁴ (Rob 00-3).

⁶⁴ In my interview with one of the external examiners, he argued that it was not a case of the externals needing more time or expertise. The problem was rather a case of an “inexperienced person” who was not sufficiently familiar with South African standards (Nic 01-1).

Robert noted, "I had a lot of sympathy with what he was saying. And that's the basis of reviewing the system" (Rob 00-3).

Kevin confessed that he too had felt quite strongly during the examination meeting that the department should accept the mark awarded by the external examiners, but "I later changed my mind". His view now was that the externals' mark should rather be treated as a recommendation to the department, not as final. It was the department's role "taking into account both the internal and externally awarded grades" to make a decision. He justified his change of mind on the basis that "the external sees only the written thesis and does not know factors like, e.g. a) how much input was provided by the supervisor towards the success of the project, b) how much initiative was shown by the student; did he/she drive the project, c) the ability of the student to solve problems encountered during the course of the project." He argued, "To disregard the internal's mark completely assumes that the internals are somehow less qualified than the externals to judge the project. I doubt that is the case." He goes on to add, "Our department does not have a clear policy on how a mark should be awarded. For example, should a thesis be marked on the written document alone, or should the mark include a contribution from the oral, or should the hardware constructed be judged on its quality of construction or on how well it works? No guidelines have been written down" (e-mail, 27/11/00).

The committee completed its review in mid-2001. It had been tasked to address "the one big problem...of the external examiners" (Dav 01-1). One of the final recommendations was that "a thesis project board be set up consisting of mainly senior staff members and other interested members. The job of this board would be to deal with cases where there was no consensus agreement even after the external marking meeting." This proposal seeks to give final arbitration decision to the thesis project board (e-mail, 20/6/01). Explaining the rationale behind this proposal, David argued, "We really do need these people from outside...but do you give these people the power to actually tell you how to do your job?" (Dav 01-1) Last year "they had the power to act. Right now what we are doing (is) taking (that) away. We are saying that they can recommend...we

accept it (their mark) as recommendation” not the final word (Dav 01-1). The proposal was accepted unanimously.

Thus, procedurally there seemed to be resolution. The role of the external examiners had been clarified. I argue however that other substantial educational issues were still unresolved. In a final interview, David admitted that some common understanding among the staff about what a thesis is had not been achieved (Dav 01-1). Nor as Kevin argued was there, as yet, any clear policy on “how a mark should be awarded”, in other words, what kinds of evidence can legitimately be drawn on to support the assessor’s interpretation. These are fundamental validity issues, fundamental to the process of evaluating interpretations, and are yet to be addressed explicitly and extensively by this community of practice.

I now turn to a brief summary of the key insights gained through the Engineering case study in response to my guiding questions.

5.4.4 Summary

The Engineering community of practice provided me with an ideal opportunity to search for explanations for both the common ground which academic assessors share, as well as the differences in interpretation. From the analysis emerge three features which contribute to the constitution of assessors’ interpretive matrices: an epistemological orientation that comes from the specialization, particular ‘ways of seeing’ that are developed through experience, and a privileging of evidence based on different levels of involvement. I then explored how these interpretive matrices contributed to different interpretations; how from the perspective of their various roles, assessors interpret student performance based on different kinds of evidence.

What is at stake in the negotiation of these differences? Consistent with the Humanities analysis, I argue that there are two issues at stake in negotiating the differences which emerge in the moderation meeting. One issue is collegiality: the individual and collective face needs of both the internal and external examiners have

been threatened. Up until 1999 this community had been able to manage these face needs, but in the 2000 examination meeting one member of staff disrupted the facework conventions. Arguing his position and saving his own face was more important than maintaining the face needs of this collegial community.

As I argued earlier in relation to the Humanities case, a second issue at stake is what constitutes legitimate evidence for the interpretation of performance. There is in this discourse a contestation between two modes of knowledge: one which privileges the more external, distant, uninvolved view, and the other which privileges the internal perspective, the one closest to the performance. I argue that the former mode of knowledge is privileged in the 2000 departmental examination meeting. This privileging is to some extent structurally constituted by the discourse and the exchange structure of the meeting. The privilege of the externals was further strengthened as a result of the community's need to demarcate turf, to set boundaries, to label what and who was 'out of line'.

And finally, how were these differences resolved? The contestation which emerged in the moderation meeting was resolved in two stages. In the context of the meeting, resolution was achieved by privileging the external examiner's perspective. The effect of this was to diminish the authority of the internal examiners in general, but specifically to discredit one particular internal examiner. Following the event, procedures were put in place to ensure that final authority rested with the department. However, fundamental issues about what constitutes legitimate evidence for the assessment of final year projects remained unresolved.

In the next chapter I turn to the task of pulling out the major insights which emerge from the case study analysis – insights which offer a deeper and richer understanding of assessment as a social practice.

Chapter 6: Discussion and Conclusions

It is this *double truth*, objective and subjective, which constitutes the whole truth of the social world. (Bourdieu in Wacquant and Bourdieu 1992, p. 255)

6.1 Introduction

This study set out to explore a problem in practice, namely, how academics evaluate the soundness of their assessment-based interpretations. I explored this problem within the context of final year undergraduate projects in two different communities of practice, one in the Humanities faculty and the other in the Engineering faculty at SAU. With respect to their validation practices I was interested in the following issues:

- Firstly, academic communities of practice, by definition, share some common ground. What is the basis of this ‘common ground’? In other words, what explains the manner in which academic assessors, with minimal co-ordination or explicit articulation of ‘norms’, are able to get on with the numerous classificatory acts which are required of them as members of a particular disciplinary field?
- Secondly, I was specifically interested in the phenomenon of multiple interpretations of the same task. What explanation is there for these differences in interpretation? What can dissensus within a community of interpreters expose about the complexities of assessment as an interpretive process?
- Thirdly, reaching consensus between assessors about what value to award an instance of student performance is not always simple to achieve. The negotiation of interpretive differences can be a site of struggle, and this struggle implies that there are certain stakes in the resolution. Thus I inquired, in the context of collegial dialogue, what is at stake in the resolution of differences?
- Finally I asked, if consensus is not a given, how then are differences resolved? Specifically, how is power exercised in the resolution of these differences?

Consistent with the goals of critical discourse analysis (and critical theory more generally, see Calhoun 1995), my goal was to open up these validation practices for scrutiny. In each case I used specific ‘moments of crisis’ (Fairclough 1992) in the life of these communities as springboards to de-naturalize their practices. I used the methods of

ethnography and critical discourse analysis to interrupt the every day common sense of institutional life (Luke 1995-1996). This de-naturalization entails illuminating unsuspected connections between what is often perceived, in assessment, to be an antagonistic relationship between objective and subjective modes of knowledge. What my analysis exposed was the 'objectivity of the subjective' (Bourdieu 1990), that the objective and subjective are in a mutually constitutive relationship. Ultimately the goal of this opening up is not simply to disrupt but to offer alternative conceptualizations of these very practices. Chouliariki and Fairclough (1999) note that neither CDA nor any other forms of critical social science are in the business of prescribing alternative practices but are there to help clear the ground for those invested in the practices, by clarifying obstacles *to* change and possibilities *for* change.

In this chapter I summarize the insights which I gained from this study by responding to the four sets of questions which have guided this study. I then conclude by drawing out the specific ways in which this study extends our understanding of assessment as a social practice. These conclusions have implications for our task as academic communities of interpreters, and I explore some of these implications in the final chapter.

6.2 Discussion

I took as a premise for this study what Bernstein (1983) refers to as the "recovery of the hermeneutical dimension" (p. 171), that is, a growing recognition of the role of interpretation at every stage of scientific inquiry in both the natural and social sciences. I noted that even psychometrics (as presented by Messick and Cronbach) has had to reject the "permanent division between observation and theory" (Bernstein 1983, p. 172). Accepting "the universality of hermeneutics" (Flyvbjerg 2001, p. 28), the crucial question is, on what basis do we evaluate the soundness of our interpretations? I argued in response to this question, that different approaches to validity inquiry will have different goals and thus prioritize different criteria. The goal of psychometrics is the construction of generalizations across individuals and contexts. Only stripped of their particularity can interpretations be meaningful across individuals, groups and contexts. The challenge of this scientific inquiry is to delineate the limits of generalizability (Moss et al 2002). In

contrast, interpretive science is interested in particularity and would argue that social life is best understood in context, because that context is constitutive of its meaning. The goal of interpretive and socio-theoretical approaches to validation is to understand the actions of individuals *in* context, “to understand what humans mean and intend by what they say and do, and to locate those understandings within the historical, cultural, institutional and immediate contexts that shape them” (Moss et al 2002, p. 7).

As the theoretical framework for this investigation I proposed a social theoretical approach, based on Bourdieu’s theory of practice, which I argued would illumine validity inquiry in productive ways. In particular it would illumine the constitutive role of context, the generative exploration of differences, and the inevitable relations of power. As I argued before, these concerns do feature in psychometric approaches to assessment, but they are not foregrounded sufficiently. In contrast, from a social theoretical perspective, these features are privileged as central to an understanding of assessment as an interpretive act, as well as central to the task of validation. In other words, from a social theoretical approach the task of evaluating the soundness of assessment interpretations cannot be undertaken without consideration of context, difference and power.

I now turn to a discussion of my central questions.

6.2.1 Academics’ ‘feel for the game’

My first set of questions was: What is the basis of academic assessors’ ‘common ground’? How is it that they are able to get on with the numerous classificatory acts which are required of them as members of a particular disciplinary field?

From a social theoretical perspective, I have argued that we cannot make sense of validating practices outside of the logic of the field. In the context of this study, this means that academic validating practices only make sense within the logic of the academic profession, and within the sub-field of disciplinary communities of practice. Bourdieu argues that it is the field that determines the kinds of capital which are valued,

and I have argued that for higher education these are principally knowledge and autonomy. These forms of capital in turn define institutional and individual positions within the field, positions which are either advantageous or disadvantageous in relation to capital accumulation. Thus institutions which have accumulated vast amounts of capital (measured, for example, through scientific ratings, publication counts, and research funding) are advantageously positioned in the hierarchy of institutional prestige. The same applies to individual positioning within the field; academics' positioning within the field is, to a large extent, determined on the basis of their accumulation of specialist-based expertise. Furthermore, since capital is a form of power, it is largely on the basis of this expertise that professional authority is determined.

Not only does capital define institutional and individual positioning, but it also specifies the "principles of vision and division", that is, "the legitimate principles of legitimation" (Bourdieu 1996, p. 265). These are the principles which inform academic classificatory acts: what is perceived to be legitimate knowledge, the legitimate criteria for assessing this knowledge, the legitimate assessors of this knowledge, and the legitimate validators of these assessments. Fields are inevitably sites of struggle, although the degree of contestation within and from outside the field over these positions and principles will vary. Despite the turbulent changes occurring in higher education more globally and in South African higher education locally, I have argued that because of its relatively privileged position in the higher education hierarchy, SAU has been able to determinedly resist calls for change. In particular, it has been able, to date, to resist calls for increasing accountability which challenge the legitimacy of traditional modes of validation practice. An example of the latter is the deeply entrenched external examination system which serves to maintain those values which the field prizes most.

As members of the field and members of disciplinary communities, academics acquire practical mastery, or what Bourdieu refers to as a 'feel for the game'. With very few explicit codes, rules or criteria, seasoned members of the field know how to get on with the myriad of classificatory tasks which are required of them. This explains the appeal to intuition which became a dominant theme in the interviews as I probed for the

implicit and explicit criteria which academics draw on in their interpretive processes, that is, their notions of assessment as a matter of 'taste', as a 'qualitative judgment', as 'art', a 'stamp of approval', an 'academic feeling'. But the basis of this intuition is not subjectivity in its common sense understanding, that is "'merely' a matter of personal opinion, taste or bias and consequently idiosyncratic" (Bernstein 1983, p. 11). My study found these judgments rather to be "socialized subjectivity" (Bourdieu in Bourdieu and Wacquant 1992, p. 126). Academics' intuitions are "internalized externality" (Bourdieu 1996, p. 29), that is, the internalization of the objective social conditions of the field and the sub-field in which they find themselves. It is a knowledge which develops through the experience of working through thousands of concrete cases in an area of expertise over time (Flyvbjerg 2001). This is what grants them practical mastery; this is what explains how "without any conscious reference to a norm" they are able to "harmonize" their practices (Bourdieu 1990, p. 58). As the Engineering course convener noted, "No one has to tell you anything, you learn. You are in a place" (Dav 00-2).

I noted the challenges this poses for newcomers to the field who are required to perform these classificatory acts without having had time to develop the requisite professional vision (Goodwin 1994). By their own accounts, they flounder. This floundering does not go unnoticed by the more senior experienced staff, but explicit strategies for the development of such a professional vision either are not a priority, or are deemed superfluous; as one interviewee noted, "It's unnecessary...they'll learn quickly on the job" (Geo 00-1). According to one staff member's perspective, it would seem that "coming into line" with the community's 'ways of seeing' is essentially a matter of time. He recounted how in his first year of marking his marks were too high, the following year his marks were too low. He anticipated in his third year that he would be more "in line" with the others (Dav 00-2). This was the pattern he was told to expect by senior colleagues.

Unfortunately for others, the process may not be so straightforward. More deliberate strategies are required by the community to bring outliers into line, and communities of practice have a repertoire of strategies by which they preserve some degree of closure,

namely, strategies for demarcating the territory, for delineating the boundaries, and positioning individuals accordingly. The process of dealing with outliers can be painful for both communities and individuals. To avoid the pain communities may choose to ignore the differences and the conflict which these differences might evoke. The Humanities HOD admitted, "I think we actually know (our practices) are extremely problematical but (we) suppress (them) because of painful experiences we have had in trying to deal with the problems (e-mail, 22/9/99). With respect to individual responses to conflict, there are a variety of responses. When confronted, individuals may opt to give in. Another strategy which they may adopt is to continue to fulfil the obligations of their job but to divest from committed engagement. Yet another strategy is that individuals may choose to leave the community altogether.

If we were to leave the story here, we would be left with a highly determined account of academic professional judgment, that is, professional judgment as essentially the outcome of social conditioning, in which assessors are either 'in-line', 'out of line' or in the process of 'coming into line'. But assessors' accounts of their interpretive processes revealed another context which also profoundly constitutes the interpretive process. This is the immediate context of the assessment task and all that this entails. Their accounts revealed a complexity of contingencies which they bring to bear in the assessment of student performance. This was illustrated in the Humanities case study through the analysis of Mary's validity argument. There a range of factors were exposed which influenced her interpretation of a particular student's performance: her relationship with the student and the class, the research process and her particular role as supervisor, the broader socio-political context and the inevitable compromises this posed. This complexity of contingencies was also illustrated in the Engineering case study through an analysis of Robert's validity argument in which he argued for a particular interpretation based on the complexity of the task, the student's independence, as well as the student's ability to explain and defend himself. A significant reference point for these interpretations is the immediate context, in particular the students' efforts, commitment, engagement, and interests; the supervisors' interests, intentions, and expectations; as well as the development of a research project with all its possibilities and risks. As one

member of the Engineering staff noted with reference to the thesis, "How much is the supervisor and how much is the student? If it's something I really care about, I'm involved" (Dav 00-1). This is the complex and inescapable subjective terrain of assessment.

Thus the study has shed some insight on supervisors' interpretive matrices; it has illuminated something of the environment out of which interpretations are born. Assessors' interpretative matrices are constituted, in part, by the objective conditions of the field and of the community of practice. These are objective because they are to a large extent independent of the individual assessor; they are conditions which apply as a result of being a member of the field. But at the same time, assessors' matrices are constituted by the particular context of the assessment event. This is a highly subjective terrain, that is, it is significantly dependent on the assessor and the particular assessment circumstances. As the Engineering HOD candidly noted, "A lot of the mark is going to be a quality judgment and it's going to depend on the characteristics of the supervisor" (Ken 00-1).

The tension between these two contexts, these two modes of knowledge was acknowledged by many interviewees in a variety of different ways. I tried to capture the essence of this tension in the distinction between the mark as a measure of product and the mark as a measure of process. On one hand, the mark is a measure of product against some tacitly agreed notions which a particular community (within the wider context of the field) holds about 'good' research. On the other hand, the mark is a measure of process, a communicative exchange between a supervisor and a student. These functions create tensions because they privilege different contextual considerations, different forms of evidence. In its function as a measure of product, assessment may foreground products (e.g. the research proposal) and background processes (e.g. the drafts, the class discussions, one-to-one time with the student). In its function as a measure of process, assessment may privilege evidence from the student (e.g. their effort, progress, and attitude) and the immediate, local context of the classroom (e.g. the student's ability in relation to the class). I argued that this "massive and permanent tension" (Jos 98-1) is

one which has to be held as fundamentally constitutive of assessment-based interpretations.

6.2.2 Academic interpretive matrices

I turn now to my second set of questions. With regard to the phenomenon of multiple interpretations of the same task, what explanations are there for these differences? What might these differences expose about the rich complexity of assessment as an interpretive process?

As argued above, Bourdieu's 'logic of practice' helps us to make sense of the homology between the academic 'game' and academics' 'feel for the game', that is, academics as a result of inhabiting the same field, do share a common professional vision. This vision (or what Taylor 1987 and Goodwin 1994 both refer to as intersubjective meanings) is not the same as consensus. It is rather a pre-condition for both consensus and dissensus. As Taylor (1987) argues it is *because* of this set of intersubjective meanings that differences of interpretation emerge.

In interviews with staff from both communities I explored their own explanations for the differences which emerge between assessors' interpretations of the same performance. These explanations brought further insight into the constitution of assessors' interpretive matrices. I stress once again that interpretive matrices are highly complex constructs and the particular features which emerged in my study were strongly influenced by the context of the data and my own particular interests. I am therefore not suggesting that my analysis is a comprehensive picture of features that influence assessors' interpretations, nor am I assuming that these features are generalizable across contexts. Further investigation would need to be conducted in other communities of practice to determine the extent to which the features which have emerged from my particular sites are generalizable to other sites.

Academics' accounts exposed three features which, in part, constitute assessors' interpretive matrices: specialization, experience and involvement. With reference to

specialization, the study revealed that assessors will bring particular disciplinary and specialist orientations to the assessment act which will influence their interpretations. It appears that different disciplinary specializations tackle particular research problems from different methodological and epistemological standpoints. It also emerged that assessment may serve as one of the primary means by which communities of practice implicitly signal which standpoints they value and which they do not. In other words, assessment becomes the battle ground for disciplinary border disputes. With reference to experience, the accounts from both novices and more experienced staff testified to the process by which assessors become socialized over time into a community of practice's 'ways of seeing'. As one experienced member of staff noted, "It's relatively easy to decide whether a project deserves a first, second or third class pass. He conceded however, "But not the first time" (Ken 00-1).

Finally with reference to involvement, the interviews as well as the analysis of collegial discourse suggests that different assessor roles (whether, for example, an assessor is a supervisor or not) will inevitably give priority to different assessment functions, different assessment contexts, and different kinds of evidence. This was illustrated through the Humanities case study in the differences which emerged between Keith and Mary's validity arguments. Keith, who is not a supervisor of the student project, draws on evidence primarily from the project and his point of reference is the context of the course and the kind of graduates the department values. Mary's interpretation, as a supervisor, is premised on a different evidential base: what she knows about the student, the class, the research process and her particular role as a supervisor, as well as the broader social-political context. I noted similar differences of interpretation in the Engineering moderation meeting between Robert and Victor, on the one hand, and the external examiners on the other. From the perspective of different assessor roles, different interpretations of student performance are justified on the basis of different kinds of evidence.

I noted how traditional approaches to assessment have viewed differences of interpretation as a threat to the validation process, and in some cases a great deal of

resources are invested in devising procedures which will ensure the elimination of these differences. To value and work towards consistency of interpretation is a desirable goal. But, as Moss and Schutz (2001) argue, there may also be some value in suspending the goal of consensus, if only temporarily. As my study has revealed, an exploration of these differences can enhance our understanding of what assessment is. We are afforded a deeper appreciation of the complex interpretive environment, that is, the multiple constitutive contexts out of which our deeply intuitive judgments emerge. This deeper appreciation will also sharpen our critical stance. Moss and Schutz (2001) argue, "Dissensus can lay the groundwork for critical evaluation and guard against taken-for-granted beliefs and practices that might dominate our thinking" (p. 65).

6.2.3 Academic stakes

My third question was: In the context of collegial dialogue, what is at stake in the resolution of differences? As noted above, a consequence of these different matrices is that assessors will not necessarily arrive at the same interpretations. Mary passed the project and Keith failed it; Victor awarded a first class pass and the external recommended a lower second. In neither case, were they able to reach consensus. The analysis of the validity arguments suggested some of the underlying reasons for these differences. But I was also interested in what was at stake in the negotiation of these differences.

I argued that the collegial discourse of the moderation meetings is fulfilling two functions simultaneously. It is functioning relationally, that is, to preserve or challenge interpersonal relations between participants. It is also functioning ideationally, that is, it serves to sustain or challenge systems of belief and knowledge held by members of this group. In relation to the former function, I argued that one of the stakes in these moderation meetings is collegiality. In the negotiation of differences of interpretation, there is the potential risk of threatening individuals' face needs, that is, a person's need to be well thought of and not to be imposed upon. Equally important is collective face, that is, a group's need to be well-thought of by another group whose opinions they respect, for example, the internal examiners' need to be found reputable in the eyes of the external

examiners and vice versa. In a context where professional identity and status are primarily based on specialist expertise, a challenge to an interpretation (whether an individual's or a collective's) can be perceived as a challenge to professional credibility. Thus colleagues weigh carefully the risks involved in such a challenge. While these academics may hold to their interpretations with a fair degree of certainty, they do so conscious of the potential threat that such differences may pose to individual and collective face needs. As one member of the Humanities staff noted, a threat to one's face may have very real negative sanctions. Academics may even at times, for the sake of collegiality, seemingly concede to a position that they in fact do not fully embrace. I argue, therefore, that collegiality is one stake in the negotiation of differences.

In relation to the ideational function of this discourse, I propose that another stake in these meetings is these communities' perceptions that they do share a common professional vision. Early in my study, I found that on several occasions members of staff in both communities felt the need to assure me (and perhaps themselves) that there was a common professional vision. The Humanities course convener, for example, argued, "Most of our staff have a common concept of (a particular disciplinary form of) insight...it might be different to the Afrikaans lecturers at (an Afrikaans-speaking university) but we (at SAU) have a common concept" (Dan 98-1). A member of the Engineering staff noted, "Statistics might call me a liar but the sense I get is that in fact most first and second markers are very close. Sometimes it is spot on, so it means there is a common understanding of what a thesis is about" (Car 00-1). I would suggest that academic communities are heavily invested in a belief that they do share a common ground. Piper (1994) explains the logic of this belief. He argues that the sharing of esoteric knowledge is the "cement" which binds specialist-based professions. Thus agreement between assessors is an affirmation of mutual membership in a disciplinary community. It is a "signal that they have a common professional understanding of quality as manifest in their subject" (p. 80). Thus, Piper argues, there is a deep-seated vested interest in this perception of agreement since it touches academics' very identities, identities which are both professional and personal.

But then how do these very same communities make sense of the dissensus which arises out of the moderation meetings? I argue that underlying this discourse may lurk some form of Cartesian anxiety (as discussed in Chapter 3). At the heart of this anxiety is the belief that as assessors they are faced with a choice between two alternative forms of rationality. One alternative, what Bernstein (1983) refers to as 'objectivism', is the belief that there are 'standards', there are 'fixed points' against which performance can be measured. The other alternative, what he refers to as 'relativism' is a resignation to the reality that there is no 'overarching framework' to which we can appeal, there are no 'fixed points'; there are only 'your' standards and there are 'my' standards. Assessors may well wonder, if they cannot reach consensus, what does this suggest about the forms of rationality that inform their discipline? A form of rationality in which 'anything goes' will result in, as Mary argues, "students being at the mercy of our idiosyncrasies" (MW 250*). Underlying the tension which characterizes these negotiations is the haunting spectre of relativism, that point where "nothing is fixed, where we can neither touch bottom nor support ourselves on the surface" (Bernstein 1983, p. 18). At stake, I argue, are these communities' perceptions of themselves as rational communities of interpreters.

While the moderation meetings may constitute this anxiety, I argued that in their actual professional judgments the members of these communities do not hold either of these options as viable. They have rejected the claims of objectivism, and are equally sceptical of relativism; it is not the case that 'anything goes'. The rationalism that informs their interpretations is something beyond either of these positions. Their judgments are constituted by the tension between the macro and the micro contexts, between objective and subjective modes of knowledge and between the various kinds of evidence which these contexts and modes of knowledge require. Indeed their very own practices are a testimony to a different kind of rationality which is "beyond objectivism and relativism" (Bernstein 1983), one which eschews the antinomy of objectivity and subjectivity (Bourdieu 1990). In describing the kind of rationality which characterizes expert judgment, Dreyfus and Dreyfus (1986 cited in Flyvbjerg 2001) propose the concept of 'arationality' (in contrast to irrationality). They argue that in the West rationality has become synonymous with analytical thinking, that is, the conscious separation of whole

into parts and the application of context-independent rules or criteria. Their study of expert judgment suggests that professionals' judgments do not always conform to this Western notion of rationality; it is rather a context-dependent, experience-based, and situational judgment¹. I would suggest that academics' professional judgment is an illustration of 'arational' judgment, what Flyvbjerg refers to as 'phronesis'. As noted in Chapter 3, this is a form of judgment which is pragmatic, variable, context-dependent, based on practical value-rationality. Part of the reason why moderation meetings are sites of unresolved and unproductive conflict is because communities of practice have failed to recognize and value this form of rationality. I return below (6.3) to elaborate on features of the rationality which constitute academic professional judgment.

6.2.4 The exercise of academic power

As noted above, at stake in the negotiation of difference is both collegial relations amongst the participants, as well as the community's perception of itself as a rational community of interpreters. The final question which I explored was how these differences were resolved. In responding to this question, I put the issue of power 'on the table'. But as Flyvbjerg argues, it is a particular understanding of power; it is power understood as the force of relations². Thus the important question is not, 'Who has power?' or 'Where is power located?', but 'How is power exercised?'.

In Humanities, I argued that the differences of interpretation were not resolved; they were acknowledged as problematic by the HOD, but for a variety of reasons it was not in this community's interest to commit itself to working through these differences. This task was perceived (by some anyway) to be very difficult, if not impossible. With reference to the Hum300 course I asked, "What would it take to arrive at consistency in the marking? Is it possible?" Elsie responded: "Honestly, it's not.... with research we have different values, we have different opinions, we have different expectations, so it's very difficult to maintain consistency" (Els 99-2). The HOD noted, "...perhaps we need more marking workshops. But I wonder...some of our differences are fairly deep" (e-mail, 22/9/99).

¹ These findings are consistent with the work of Paul Hager and David Beckett (Hager 2000; Beckett and Hager 2000) who argue for the crucial role of context in professional work-place practical judgment.

² Flyvbjerg (2001) is drawing on a Foucaultian understanding of power.

In Engineering there were two different stages of resolution; there was resolution in the context of the moderation event itself and resolution which arose out of the events which followed the meeting. In the context of the meeting, resolution was achieved by privileging a particular perspective, that of the external examiner. The effect of this was to diminish the authority of the internal perspective, in general, but specifically to discredit one particular internal examiner. Thus it would seem that the best way to preserve collegial relations (particularly between the external examiners and the SAU examiners), as well as to uphold the community's belief in itself as rational, was to castigate the one who was 'out of line'. This is indeed a solution that communities may choose to adopt when faced with conflict. (I, however, propose an alternative approach in the final chapter).

Outside the event, the process of resolution took a different turn. Once removed from the tense dynamics of the meeting, some members of the community were able to reflect more critically on what had happened in the meeting. The issue that was explicitly put on the table was: in the case of a dispute between an internal and external examiner who was to have the final say. Some argued in the moderation meeting that the external examiners were arbiters of internal disputes, and thus were the final authority. But the collective position which eventually emerged was that, in fact, final authority should rest with the department. In order to give effect to this decision, an alternative procedure was recommended for future practice, one which placed final authority with the department. Thus a procedural resolution was achieved, but as one of the members of staff argued, there was as yet still no clear policy for how a mark should be awarded (e-mail, 27/11/00). This lack of clarity pointed to fundamental educational issues, specifically what constitutes legitimate evidence for the assessment of the final year projects, which remained unresolved.

6.3 Conclusions

In conclusion, this study offers two contributions to a small but growing body of research which takes a sociological perspective on assessment. The first contribution is a

theoretical and methodological approach to the study of assessment practices. The second contribution is that, by exposing some of the complexity of the interpretive environment out of which academic professional judgment emerges, we achieve a deeper understanding of assessment as a social practice. This in turn illumines some of the features of the kind of rationality which constitute academic professional judgment. I turn now to elaborate briefly on each of these contributions.

I noted at the end of Chapter 2 that educational assessment finds itself at the nexus of complex and highly contested expectations for social and political change. In response, scholars from more interpretivist and socio-cultural traditions have been critical of the role of assessment as a technology of educational reform and control. Underlying these agendas are “conceptual mythologies” (Davis 1998, p. 1) about assessment based on particular notions of science which these scholars would reject. They argue that assessment is a politically and conceptually complex social practice, and the theories and methodologies for exploring this complexity are undeveloped. Thus there is an urgent call for theoretical and methodological approaches that enable us to better understand the ways that assessment works within the local context (Moss 1996).

Drawing on social theory, on the one hand, and the methods of critical discourse analysis and ethnography on the other, I have constructed a theory and method of inquiry which illumines aspects of assessment as an interpretive process which have been obscured by traditional psychometric approaches. With respect to other socio-cultural and interpretivist studies, I feel that my study has demonstrated the particular value of analyzing assessment discourse using the tools of CDA. There is still much research to be done in exposing relations of power which are pervasive in assessment discourse, not only among fellow assessors (my focus) but as importantly between assessors and assessed.

Luke (1995-1996) argues that one of the challenges in educational research has been to bridge macro and micro analysis of educational practice. His argument is that much of critical education research has focused on one or the other; but few have been able to

show “how large-scale social discourses are systematically (or for that matter, unsystematically) manifest in everyday talk and writing in local sites” (p. 11). Consistent with the goals of social theory and Bourdieu’s theory in particular, I have attempted to construct a theoretical and methodological approach which links macro and micro analysis, that is, the structural social conditions and the particular acts of agents. I have, however, privileged the acts of agents and given less attention to the broader social conditions. I felt that this emphasis is an important corrective to traditional approaches to assessment which have privileged the analysis of regularities and are less interested in how assessment works locally. But as Moss et al (2002) argue, both emphases are necessary. More work is needed to develop socio-theoretical methods of analysis which link the macro and the micro, “to grasp particularity within generality and generality within particularity” (Bourdieu in Bourdieu and Wacquant 1992, p. 75). I believe that the particular theoretical and methodological approach which I have constructed is a contribution to further research in this area.

The second contribution of this study has been to enrich our understanding of the complex interpretive environment out of which academic professional judgment emerges. By ‘drilling down’ into differences of interpretation, I have illuminated the multiple contexts which constitute academics’ judgments. These multiple contexts include the macro-social conditions of the field and the way in which these conditions constitute the “principles of vision and division” (Bourdieu 1996, p. 265) that inform academic classificatory systems; the meso-level context of disciplinary (and sub-disciplinary) communities of practice with their particular epistemological orientations; and finally the micro-level context of assessors’ interpretive matrices, which are significantly constituted by their disciplinary interests, professional experience and levels of involvement. My study has shown how ‘context matters’ by exposing the multiple layers of contingency that shape assessor’s interpretive matrices in predictable and unpredictable ways.

In addition to illuminating the multiple contexts which constitute academics’ judgments, this study also took up Bernstein’s challenge that our move beyond objectivism and relativism requires a re-conceptualization of human rationality (as argued

in Chapter 3). As I examined more closely academics' interpretive processes and the ways in which they negotiate the tensions between these multiple contexts, I gained insight into patterns of rationality which underlie academics' professional judgment: a rationality which I have argued is neither objectivist nor relativist. I noted the work of Flyvbjerg (2001) who has taken up Bernstein's challenge but my overall impression is that in terms of the task of re-conceptualizing an alternative, more has been said about what this rationality is *not* (by way of critiques of objectivism and relativism, modernism and post-modernism) than about what it *is*. I offer therefore, in conclusion, some tentative proposals of what some of the features of such a rationality might be.

If as Bourdieu (1992) argues social reality is a "double truth" (p. 255) then human rationality might be usefully characterized as a double reading, as an iterative movement between two modes of knowledge which comprise the objective and the subjective. On one hand, there is the objective reading, that is, the attempt to "grasp from the outside" (Wacquant in Bourdieu and Wacquant 1992, p. 8), to observe, measure and map reality independently of the representations of those who live in it (both the assessed and the assessor). This is the task privileged by psychometrics, the task of ascertaining context-independent regularities. This is an important task; student performances do exhibit regularities and these must be understood. These regularities in turn powerfully shape the classificatory schemes which are re-applied in the interpretation of this objective reality. This is the manner in which the field sustains itself. As one of the interviewees argued, "The university is a self-sustaining institution. The belief that the university matters as an important place, that it is the site where knowledge is produced, is constructed by the university. It constructs a lot of things which allow it to continue its existence" (Jos 99-1). As I have argued these objective regularities also explain how without reference to explicit criteria, standards or benchmarks, academics are able to get on with the myriad of classificatory acts required of them as members of their profession. Their intuitive judgments are internalizations of the objective regularities of the field they inhabit. Thus human rationality, as it is characterized by academic professional judgment, entails an objectivist reading.

The other mode of knowledge which comprises this rationality is the subjective reading. This is the reading which is deeply invested with the self, the reading which acknowledges professional judgment as inescapably (in part anyway) an embodiment of the assessor. This is an insight which assessors are acutely conscious of as illustrated by many interviewees' testimonies to how deeply implicated supervisors are in their students' performances. This subjective reading is also relational, in that it is a communicative exchange between the assessor and the assessed. It thus inevitably has to acknowledge and contend with power, that is, the force of relations. This subjective reading is situational, in that not only are judgments invested with the assessor but they are also invested with the particularities of the assessment event. This subjective reading is also profoundly pragmatic, in that, in their judgments, academics are cognizant of the multiple functions which their assessments are playing. They are also extremely sensitive to the consequences of their assessments both for themselves as professionals as well as students. Therefore far from being *mere* personal opinion or an *arbitrary* 'taste' or 'gut-feel', this subjective reading is a socially constituted, practical mastery. This is an aspect of human rationality which has been greatly undervalued, and yet it is essential to an understanding of assessment as an interpretive process and to professional judgment more generally.

In a nutshell, this study contributes to an understanding of professional judgment – and the forms of rationality which underlie it – as a double reading, as an iterative movement. As Bourdieu argues, “It is this *double truth*, objective and subjective, which constitutes the whole truth of the social world” (Bourdieu in Bourdieu and Wacquant, 1992, p. 255) (*italics in the original*).

These insights into the nature of academic professional judgment and its underlying forms of rationality are not simply theoretical issues; they have implications for how we conduct our task as an academic community of interpreters. I turn therefore in the final chapter to explore some of the implications of this study for academic validation practices.

Chapter 7: Implications for practice

What had become manifest is that the movement beyond objectivism and relativism is not just a theoretical problem but a practical task.
(Bernstein, 1983, p. 230)

7.1 A focus on value

One of the insights which has emerged from this study is the crucial function which assessment serves as one of the primary means by which communities of practice implicitly communicate (to each other as well as to students) what they value as a community. This includes, as I noted before, the values of knowledge and autonomy and the implications of these values for what is considered to be legitimate in terms of the forms of knowledge, the assessors of this knowledge, and the validators of these assessments.

In both case studies I noted how each of the communities stopped short of the task of articulating this value-basis and its impact on the assessment of the final year projects. Although the Engineering community of practice succeeded in bringing some resolution to the particular issue of where final authority lies for assessment decisions, several members of staff pointed out that there were still no clear guidelines on how marks are to be awarded (see 5.4.3.3). I interpret this not as a call for tighter marking memoranda (although this might emerge as one of the outcomes) but rather a call for more opportunities for dialogue about, as Joseph put it, "what we believe we are doing and why" (Jos 99-1).

These communities of practice are not exceptional in this failing; it characterizes the higher education community as a whole. We are like fish which do not feel the weight of the water and thus take the world for granted (Bourdieu in Bourdieu and Wacquant, 1992, p. 127). We have tended to believe that the soundness of our interpretations is self-evident. As a community our primary substantial investment in validation has been the external examination system, and varying degrees of validation that come from professional bodies. But for a substantial number of our academic programmes it has been validation enough to have our external peers' annual comments about the extent to which, for example, the examination was "good", it was "fairly marked", and the "overall

course results are fair”¹. Various other validation mechanisms, for example, marking workshops, departmental examination meetings, faculty board meetings, are all essentially technologies of reliability where the main goal is to make sure there are no inconsistencies around a groups of markers or across a particular student’s set of marks. Even on occasions when there are contestations about a particular assessment, we still rarely confront the very meanings of the marks which are being contested.

The problem is not these reliability technologies. They have an important role to play in a validation system. The Engineering department is a notable example of a community that has invested considerable resources in reliability, and these practices are commendable. The problem is that far fewer resources have been invested in addressing other issues which are fundamental not only to our assessment practices but to our academic programmes as a whole. The central issue I would suggest is what we as communities of practice value. What is a ‘good’ engineering research project? What is a ‘good’ humanities research project? On the basis of what criteria shall we judge them? These are difficult questions, but our task is made somewhat more manageable if we fixate not on finding answers (i.e. what can concisely be entered into a faculty handbook) but rather on committing ourselves to a process. So the question I turn to in this final chapter is: In what ways can this study assist us in this on-going validation process?

Before turning to this question, I need however to clarify who I mean by ‘us’, to whose practice are these implications addressed. Ultimately this is a question about who is responsible for change. First and foremost, ‘us’ are members of the communities of practice, that is, academics who are responsible for the design, administration, supervision, and marking of assessment tasks, as well as the design and administration of the broader assessment system. We² are the ones who have the greatest invested interests in assessment systems which give rigorous and systematic attention to validation. If we

¹ These are some examples of the issues which SAU external examinations are asked to comment on. They are taken from the Report by the External Examiner to the Head of Department which all SAU external examiners are meant to complete.

² I say ‘we’ because as well as being an educational developer, I am also a member of a community of practice responsible for the delivery and assessment of an academic programme. These recommendations are as much intended for my own practice as they are for others.

are not invested in this process, then no amount of pushing and probing, no 'carrot and stick' set of incentives from outside the community will bring about meaningful change. In the context of a discussion about the problems of university transformation, Clark (1998) argues that it is the operational units (what I refer to as 'communities of practice') that are central to any significant change; "Change in colleges and universities comes when it happens in the trenches; what faculty and students do is what an institution becomes. It does not happen because a committee or a president asserts a new idea" (Leslie cited in Clark 1998, p. 145). Thus while pressure *for* change may come from outside (as it currently is in higher education), change itself can only come from within.

The other constituency comprising the 'us' are outsiders whose role it is to support these communities of practice in a variety of different ways. These include individuals within the faculty office, for example, the Dean, who is accountable for the overall development of the quality of teaching and learning, and other faculty office positions who are responsible for the day-to-day management of these processes. It also includes higher education development staff and others who are responsible for various forms of institutional development. If communities of practice responsible for assessment are invested in the strengthening of their validation systems then there are many ways in which these outsiders can support them in this task. but the prime responsibility for change rests with the community of practice.

I now turn to explore some features of such a validation process.

7.2 Implications for validation

7.2.1 Validation as a community process

One of the central themes of this study has been the recognition of assessment as an interpretive act. If we accept this, then it follows that the validation of our assessment-based interpretations is a community process. While this is clearly an emphasis within more socio-cultural approaches to assessment, it is a point which Cronbach (1989) and Messick (1989) also acknowledge. Cronbach argues that acceptance of the outcomes of a particular validation inquiry gains support within the relevant community and "the process is social as much as rational...it is inherently a community process" (p. 164).

Bernstein (1976) has argued that striving for objectivity is a community process. In contrast to objectivism (as discussed in Chapter 3), Bernstein argues that objectivity is essential to responsible intellectual inquiry. By 'objectivity' Bernstein means that in any domain of human inquiry there are "intersubjective standards of rationality or norms of inquiry by which we attempt to distinguish personal bias, superstition, or false beliefs from objective claims" (p. 111). There is however no "single univocal set of criteria for distinguishing what is objective from what is not" (ibid.); there will be significant disagreement about what the criteria should be and how they should be applied. The crucial issue, he argues, is "the existence of a community of inquirers who are able, willing and committed to engage in the argumentation" (ibid.). Thus the kind of rationality which emerges in the move beyond objectivism and relativism is one which is dialogical, a rationality in which there is choice, deliberation, interpretation, judicious weighing and application of 'universal criteria', and even rational disagreement about which criteria are relevant and most important" (Bernstein 1983, p. 172). I recall once again Nietzsche's quote at the opening of Chapter 2, "the *more* eyes, different eyes, we use to observe one thing, the more complete will our 'concept' of this thing, our 'objectivity' be" (Nietzsche cited in Bourdieu 1990, p. 28) (*italics in the original*).

Bernstein (1983) poses the question however, "What...is to be done in a situation in which there is a breakdown of such community, and where the very conditions of social life have the consequences of furthering such a breakdown?" (p. 226). These are apt questions for higher education where many academic communities of practice are in a fragile state. Indeed some have argued that the notion of 'academic community' is under threat (Henkel 2000). In Chapter 5 I noted some of the structural conditions which contribute to this fragmentation, namely, growth in number and diversity of student intake, increasing demands from the labour market for particular kinds of graduates, massive explosion and fragmentation of knowledge, destabilized employment conditions, and shrinking state subsidization. One of the Humanities staff members noted, "There's been a change in the department. We used to have Collegiality with a capital C, where we met more often, had tea. There's been a real breakdown...it's been very fragmented for a whole lot of reasons" (Dan 00-1). The extent of social fragmentation within the academic

institution is likely to have a significant effect on the extent to which genuine validation dialogue can occur. This also points to a potentially unique role that assessment can play in helping to develop community. The Engineering course convener contrasted the intensive interaction amongst staff over the assessment of the final year projects to the rest of the year, “The rest of the year we never meet... but for this (the final year projects) you have to (meet)...this is the point where people exchange ideas and talk about what their expectations are...it forces you to face up to each other” (Dav 00-3). I inquired whether there were no other opportunities for dialogue. “Other than that...there’s tea”, but he noted “(some) hardly ever come to tea” (Dav 00-3).

Bernstein (1983) argues that none of us need to be reminded of how fragile our community life really is, and that the threat of extinction is not an “abstract possibility but most imminent” (p. 229). In the current higher education climate of “disquieting turmoil that has no end in sight” (Clark 1998, p. xiii), I propose that a commitment to on-going evaluation of not only our assessment interpretations, but our assessment systems as a whole, can provide an extremely challenging and rewarding opportunity for the development of what are increasingly fragmented communities of practice.

7.2.2 Validation which prioritizes community values

Another central theme of this study has been to characterize the kind of rationality that informs academic professional judgment where neither objectivism nor relativism is a viable alternative. The rational basis of our professional judgment is neither in relation to nor in defence of either of these positions. Once freed of this Cartesian anxiety, we can begin to articulate an alternative kind of rationality, a rationality which is contextual, experiential and, perhaps most importantly, value-based. I suggest that while this alternative rationality is what implicitly informs our interpretive judgments, we struggle as communities to articulate and legitimate it. Without a language to describe the rationality which has informed our intuitions we become easy prey to our individual and collective insecurities as communities of interpreters (not to mention the growing threat of litigation).

I would suggest that a starting point for our validation dialogue is to prioritize, as I noted earlier, a discussion about values³. In characterizing this 'arationality' or *phronesis*, Flyvbjerg argues, "(It) thus concerns the analysis of value, i.e. what is good and bad for a person, as a point of departure for action...it focuses on what is variable, on that which cannot be encapsulated by universal rules...it requires an interaction between the general and the concrete; it requires consideration, judgment and choice" (p. 57). Flyvbjerg suggest that a dialogue based on *phronesis* needs to begin with Aristotle's three classic questions: Where are we going? Is this desirable? What should we be doing? Flyvbjerg argues, however, that Aristotle's and most modern day scholars' explications of *phronesis* do not take seriously the issue of power. Thus Flyvbjerg adds a fourth question, "Who gains and who loses and by which mechanisms of power?" (p. 60).

Thus, at the heart of validation is 'value' (as Kvale 1993 notes, 'value' being the root word of validity). And yet, as communities this is where we have been most inarticulate. In critique of his own department, one of the members of the Humanities department noted, "There's a lot of presumption and assumption...we are so used to constructing the least problematic nothingness that what everyone believes never gets stated. So I think there would be great value in us saying what we believe we are doing, and why we're trying to do it and why we think it's right or important or problematic...at least to know what each other thinks" (Jos 99-1).

One fairly standard approach to making our values more explicit is the design of marking memoranda. In Chapter 2 I noted the common practice (supported by educational developers such as myself) of encouraging participants in marking workshops to design a marking memorandum on the assumption that this set of criteria will be useful to students and assist the markers in adopting a consistent framework.

³ One of the critiques which I have of Bourdieu's theory of social practice, and of social theory more broadly, is the self-referentiality of the 'double reality' (see 3.4.1). I find this a potentially limiting theory for a discussion of values. Further research into the ethical basis for assessment-based interpretations will require a different set of theoretical resources. For example, I note developments in the field of systems theory and cybernetics which argue for 'a third order cybernetics' as a way of moving 'beyond objectivism and relativism'. This 'third order' holds out the possibility of being 'ethical'. I am grateful to Rosemary Kearney for pointing me to these developments and their great potential for a value-based approach to validation practices.

Broad (2000) offers an insightful critique of marking memoranda (or 'rubrics', as he calls them). He argues that firstly, contrary to expectation they rarely eliminate disagreements. (In my experience they often compound disagreement since assessors disagree not only about the marks about also about the memorandum). But more importantly, Broad argues, these rubrics support the mistaken notion that assessment consists of identifying characteristics in the text and assigning them relative values. Rubrics describe textual features but are completely silent about the "prejudices and foreknowledge" which are inevitable and valuable in interpreting a text (Moss cited in Broad 2000, p. 247). Ultimately, he argues, rubrics are misleading. They are completely silent about the value system which underlies the interpretive acts.

I would propose, therefore, as communities of interpreters that we need to move beyond the task of designing marking memoranda which focus on qualities of text. While our discussion may start there, we need to explore more deeply the implicit values which underlie those textual qualities. This discussion is of course not to be confined to colleagues; it must also engage students. It must start, however, as a commitment on the part of the disciplinary community to make more explicit the value base which underlies their professional judgment.

7.2.3 Validation which respects difference

Once a community begins to engage in dialogue about the values that inform its assessment-based interpretations, inevitably differences in value perspectives will arise. I would suggest two principles which need to guide our validation process. Firstly, as I have argued throughout this study, differences can be a resource which strengthens our validation practices. Secondly, some value differences may be irreconcilable. Thus while consensus may be an ideal to strive for, the reality is that it is not always possible to achieve. I will deal with each of these principles in turn.

One of the central themes of this study has been the importance of exploring differences of interpretation as a means of gaining deeper insight into assessment as an interpretive process. This would suggest that rather than a threat to validation, dissensus

is a resource (Moss and Schutz 2001); conflict can be productive. As the Engineering course convener noted, "I think the conflicts that we have...they are very important...they are helping all of us...if everything was great, you know, no arguments, no fights, we'd never grow. But now we allow people to be independent and we encourage them to give their views on things and work out things, then all of us get challenged" (Dan 01-01). As McCarthy argues, "There is no surer path to awareness of unspoken preconceptions and prejudgements than communicative encounters with others who do not share them" (McCarthy cited in Moss 1998, p. 55). Neither should the foregrounding of differences of value need to be perceived as a threat to collegiality. It will however require a willingness to relinquish (some of) our rights of autonomy which we prize as individuals and as a profession. It also requires re-building levels of trust where they have been severely eroded within departmental communities.

Secondly we need to accept the possibility of irreconcilable differences. The reality of pluralism which characterizes our communities poses a serious challenge for the achievement of consensus. One might well argue that is it all well and good to respect differences of opinions, but at the end of the day only one mark will be entered in the spreadsheet. I offer what I recognize is not a fully satisfactory response to this dilemma. Moss and Schutz (2001) distinguish between consensus and agreement. Consensus assumes that all parties reach an understanding that all interpret in the same way⁴. Agreement, as they define it, is where all parties accept a particular conclusion in a particular context, but what is agreed upon may actually be (and they argue, to some extent always is) interpreted differently by each person (p. 59). Agreement is the practical reality that a judgment has to be made based on the best evidence that we have before us, but it is not consensus. The advantage of this distinction is that it identifies the final decision for what it is – a compromise. And, if nothing else, community life is about learning to live with compromise.

Someone might ask, but what about the students? I recall one of my interviewees' confessions that in relation to students' perceptions of their assessments he often found

⁴ They argue that this is a Habermasian definition of consensus.

himself in collusion with what he knew to be the myth of objectivity. He argued, “This whole evaluation thing...(we’re) kind of trapped. I sometimes have tried to pop the bubble of it, but you’ve almost got to collude in it and keep the belief going” (Kei 99-2). This collusion is, to some extent, something which we are all trapped in as educational assessors. Recalling Bourdieu’s argument that it is a “*double truth*, objective and subjective, which constitutes the whole truth of the social world” (Bourdieu in Bourdieu and Wacquant, 1992, p. 255) (italics in the original), I wonder whether we do our students a favour by colluding with the myth of a partial truth rather than the reality of a double truth. Challenging as it may be, I would suggest that as communities of interpreters we need to open up our assessment practices to students; much greater transparency is required of us. The spin-off is that by exposing students to and modelling for them the contextual complexity of professional judgment we prepare them for the kinds of rational thinking which their future professional contexts will require of them. This might be one of the most crucial life-skills we have to offer our students.

7.2.4 Validation which acknowledges power

I propose as a final practical consideration that we put firmly on the table the issue of power. This could start with the acknowledgement that our collegial dialogues are not in reality dialogues among equals. Bourdieu’s theory of practice illumines significant differentials of power, based on the dissemination of particular kinds of capital which are valued within the field. In relation to this capital we are not all equally endowed, (e.g. we are not all experts, we are not all equally experienced, and we are not equally articulate). This would suggest that in all kinds of ways our dialogues are structured to advantageously position certain perspectives over others. This structural privilege will ensure that, while we might all contribute to the dialogue, those who are positioned advantageously, will be able to effect change in ways which those who are less advantageously positioned will not be capable of. This is not necessarily a bad thing. The issue is how those who are better positioned to effect change *use* that power. These are Flyvbjerg’s (2001) questions, ‘Who wins and who loses?’ and ‘Whose interests are served?’.

There is a sense, however, in which those endowed with less capital (and thus power to effect) have a different kind of positional advantage. They are best positioned to disrupt the practices simply through the posing of naïve questions, such as the new member of staff who asked of his HOD, ‘What is the official policy on how we should mark?’ Thus newcomers and outsiders have a unique capacity for disruption, and this is a role which must not be underestimated. This is not to say that the established members of the community cannot also be disruptive, but the very fact that they have become established, that they have acquired a ‘feel for the game’, means that they are less likely to question ‘the game’.

Thus the challenge for communities committed to the validation task is to re-structure our discourse (e.g. rethinking the kinds of meetings we have and how they are conducted) in such a way so that those who – by nature of their expertise, their experience, their rank, their gender, their race, their language – are in less advantaged positions are re-positioned to more fully contribute. This is a particularly challenging responsibility for those in leadership positions.

As a final consideration of the implications of an alternative rationality for academic validation practices, I close with a quote from Bernstein (1992). This seems fitting given that his work has so much influenced my intellectual journey. Though he is addressing primarily the issue of philosophical plurality, I believe his points are relevant to the plurality of assessment interpretations. He writes, “There can be no escape from pluralism – a plurality of traditions, perspectives, philosophic traditions” (p. 329). He goes on:

Such a pluralistic ethos places new responsibility upon each of us. For it means taking our own fallibility seriously – resolving that however much we are committed to our own styles of thinking, we are willing to listen to others without denying or suppressing the otherness of the other ... What makes this task so difficult and unstable is the growing realization that there are no uncontested rules or procedures... But because there are no uncontested decision procedures for

adjudicate(ion)..., it is always a *task* to seek out commonalities and points of difference and conflict. The achievement of a 'we' – where 'we' are locked in argument with others – is fragile and temporary achievement that can always be ruptured by unexpected contingencies. Conflict and disagreement are unavoidable in our pluralistic situation. There is little reason to believe that 'we'... will ever achieve any substantive permanent consensus, and there are many good reasons for questioning the desirability of such a consensus. What matters, however, is how we respond to conflict. The response...is a dialogical response where we genuinely seek to achieve a mutual reciprocal understanding – an understanding that does not preclude disagreement” (p. 336-337) (italics in original).

Appendix One

Data on Humanities Staff

[All information is for 1998 unless otherwise specified]

Name	Age	Sex	Nationality	Race	Language	Rank	Dept. position	Yrs. in Dept
Charles	?	M	Tanzanian	Black	Swahili/ English	Visiting lecturer	Visiting lecturer in 1998 Option convener in 1999	0-5
Dan	53	M	South African	White	English	Ass. Prof	Course convener in 1998	16-20
Elsie	?	F	Lesotho	Black	Sotho/ English	Lecturer	Option convener in 1998-1999	0-5
Frank	37	M	South Africa	White	English/ Afrikaans	Lecturer	Option convener in 1998-1999	0-5
John	?	M	South African	White	English	Ass. Prof.	Head of Department in 1999	?
Jordan	37	M	South African	White	English	Sr. Lecturer	Lecturer on Honours course	0-5
Joseph	48	M	South African	White	English	Lecturer	Option convener in 1998-1999	11-15
Keith	55	M	South Africa	White	English	Ass Prof.	Option convener in 1998 Course convener in 1999 Head of Dept in 1999	26-30
Larry	?	M	South African	White	English	Ass Prof?	Option convener in 1998-1999	?
Mary	63	F	British	White	English	Sr. Lecturer	Option convener in 1998	0-5
Oliver	45	M	South African	White	English	Sr. Lecturer	Option convener in 1998-1999	0-5
Zora	37	F	South African	Coloured	English	Lecturer	Option convener in 1999	0-5

Appendix Two

Excerpt from the Humanities marking workshop

Code

/ interruption

(*) unclear word

Text in parentheses is either my additions for clarity or it is non-verbal forms of communication.

DAN: /why and then a lot of a lot the discussions comes out that way, so maybe (Keith) shall we go around this, do you want to start?

KEITH: Yes alright uh I I thought they were both terrible pieces of work and I, you know, I used my old marking schedule, you see. So for the one I gave 41% and the other one 40 %/

5 **DAN:** Which one 41?

KEITH: The the

DAN: Not that it's that much different (Dan laughs).

KEITH: Well I gave 41% for uh the A one which is this uh I suppose it's Larry's (Project A title) (moan from someone) and the other one on (Project B title) I got 40 for uhhh (laughter from someone). Ja look, this this one doesn't, I mean is very weak methodologically/

10

DAN: Which the first A?

KEITH: uh you know just trots out, you know, what what uh certain respondents said. Uh you don't know even know who's talking and why the people were picked on and uh then it just kind of ends without any analysis or anything like that but uh you know my marks are low because if these were my students they would have known they would have had to do some kind of assessment. data presentation, would have to do quite a lot of methodology. Uh they would have had to include, you know, the one doesn't have any (***) questionnaire interview schedule in the appendix. Uh I think it has one reference, uh? There's absolutely no uh in the introduction there's no reference. It sort of says

15

20 "there's a lot of work being done in this field internationally etc...etc..." but doesn't refer to a single you know source (***). It doesn't locate the topic so very weak. And uh this one jo! (reference to B) this is terrible piece of work.

MARY: I passed it generously (laughter in background).

KEITH: You passed it?

25 **MARY:** Yep!

KEITH: Generously?

MARY: Ja

(lots of overvoice)

DAN: Let's come to you Mary. (turning back to Keith) What are your criticisms?

30 **KEITH:** Well I mean uh this student just is uh going through the motions uh but doesn't understand what uh you know what he or she is doing uh uh (flipping through report) sort of misreads tables, doesn't know how to actually, you know do anything with the data but it's there. Uh you know contradicts what's in the tables, and also just there's no analysis at the end it just ends, boom, sort of saving grace is that she's read (reference to
35 some article used) and done some, you know, literature work reading etc. So that that's a reasonable section at the start, the review and the introduction. But it just goes nowhere and then it just ends. Bingo! There's no discussion, no analysis, in the theoretical sense. Um quite a nice uh reference list (chuckling from someone) and then it's got an appended questionnaire which is actually a group one so you can credit the student too
40 much for this. So that's why they got plug!

DAN: OK alright um, Mary?

MARY: Uh well (****) I'm sorry to say. I gave this one 67%/

DAN: This is this is the B?

MARY: This is the one that I did (Project B title). Yep and um wow if you're going to give that
45 one 40 all the rest of mine should have got 20 is all I can say for last year. Um let's see I (the student's name) I know this, really clever girl, really clever hard-working, nice. (Mary's anxious laughter and laughter from others). I can't get her out of my mind (chuckle from someone).

OLIVER: How much prompting, did you prompt her on the analysis?

50 **MARY:** Did I prompt her on the analysis?

OLIVER: Did you say (*)

MARY: No no we all put the data ja we just put the data in and that was it/

DAN: Then you let them/

MARY: /and then I show them how to use stats graphic package and they can do what they like
55 and um/

DAN: And then you leave it up to them, then you don't get a draft in and tell them to redo it?

MARY: Um you know I really can't remember that, I know that we were running at a different pace from the other ones because we were sort of held up by each other. I mean there were there are there are lots of mistakes in here (**). This is wrong that's wrong but uh
60 taking. I take the whole bunch of them and I think, well, 20 % might fail but no more.

This is one of the better ones. I rank them. I chuck some marks at them. So uh ja I can, if you want me to jack up my standards, I can but /

KEITH: No no, that's not what we're here about.

MARY: No no that's what I'm saying, you know I just think if if the students if the students are
65 doing so badly, that's our fault. We don't fail them because in their third year they are
doing badly. We have to say, this is an average group of students. They've sat here for
three years. They've been through our courses and you know up to 25% may fail, maybe
less. That's a very high wastage rate into third year. So uh there we are. I just say this is
the material we've got. We just put some figures onto it and uh (clears throat) off we go.

70 **OLIVER:** Certainly not all our fault (laughter from the group).

DAN: No, I think that is, I'm goin', I think that is important. Oliver, do you, what are your
marks?

Appendix Three

Data on Engineering Staff

[All information is for 2000 unless otherwise specified]

Name	Age	Sex	Nationality	Race	Home Language	Rank	Dept. position	Yrs. in Dept
SAU Academic staff								
Allan	46	M	South African	White	English	Sr Lecturer	Educational Dev. Officer	0-5
Anthony	35	M	South African	White	English	Ass. Prof.	Supervisor	0-5
Carl	50	M	South African	White	English	Ass. Prof.	Supervisor	11-15
Thomas	53	M	South Africa	White	English	Ass. Prof.	Supervisor	0-5
David	35	M	South African	Black	English	Sr. Lecturer	Supervisor Course Convener	0-5
Donald	?	M	South African	White	English	Prof.	Supervisor	31-35
George	63	M	South African	White	Afrikaans	Sr. Lecturer	Supervisor	31-35
Peter	64	M	South African	White	Afrikaans	Prof	Supervisor	11-15
Kenneth	50	M	South African	White	English	Prof.	Supervisor HOD	11-15
Kevin	33	M	South African	White	English	Sr. Lecturer	Supervisor	0-5
Kurt	29	M	South African	Indian	English	Lecturer	Supervisor	0-5
Paul	57	M	South African/Angolan	White	English	Sr. Lecturer	Supervisor	11-15
Richard	59	M	South African	White	English	Sr. Lecturer	Supervisor	11-15
Robert	56	M	British/SA	White	English	Prof	Supervisor Ex-HOD	16-20
Stephen	60	M	South African	White	English	Prof.	Supervisor	21-25
Victor	?	M	Zimbabwean	Black	?	?	Supervisor	0-5
Andre	?	M	?	White	?	?	?	?
James	?	M	South African	White	English	Prof.	Supervisor	31-35?
External Examiners								
Ella	?	F	South African	White	Afrikaans	Prof.	External examiner	n/a
Philip	43	M	South African	Black	English	Prof.	External examiner	n/a
Nicholas	40	M	South Africa	White	English	Prof.	External Examiner	n/a

Appendix 4

Excerpt from Engineering departmental examination meeting: Dispute A

Code

(X) To protect anonymity of speakers, references to specialist discourse have been omitted and replace with an X.

(*) unclear word

K Kenneth

D Donald

R Robert

A Andre

N Nicholas

K Mr V (the student), the supervisor gave it 88, second marker gave it 72, third marker gave it 69, and the consensus mark was 75, and external examiners recommend a 70.

D (**)

K You wished to make a remark?

5 D It was about the (X), it was vague on the (*). It's a multi-disciplinary subject and he seemed to get around and sort out information from a number of different sources but you're recommending 70 (**)

10 R Could I say that I gave it 72, and it's actually very useful to speak to the student and to see how the student approached the problem. I was convinced it should be a first after discussing the project with the student and hearing the student's story of how he went through this project. It was entirely self-initiated, had almost no support from his supervisor whatsoever, and and from the report, for instance, Prof. R. gave it 69, and I think if I came into it cold, and I read the report I would have given it a 69 or 70, but I was really very impressed with the fact that the student had worked entirely on his own, and I was convinced to raise the mark from 72 to 75, and I think that that it's important to take that into account.

20 A This brings out an important aspect is how much the candidate is able to communicate what he's doing to the outside world and how important that is. We've had in past many excellent guys but as part of an engineer you must be able to communicate to the outside world. That you can't do this, I think this must weigh against that, but I think both needs to be taken into consideration.

25 K I think looking at the write up itself, the thesis, is not as good as it should be, the references are not and the bibliography is not properly documented, and on that account I gave it (*) but taking everything into account, yes, he did quite a lot of work.

30 N I think there are a few reasons not to recommend a first. One of them is that some of the parts of the project didn't work. There was an (X) that was supposed to operate every six hours and he couldn't get that to work, a fairly simple (X) process of (X). In addition he couldn't (X), and really for a distinction candidate you'd like to have these things anticipated in advance and then in the project report itself, really there wasn't great multi-disciplinary detail, for example, (X). There was no additional investigation as to the process. So I think the 70% is a reasonable mark.

K 70?

X 70.

Appendix 5

Excerpt from Engineering departmental examination meeting: Dispute B

Code

In order to protect the anonymity of those involved in the research I have omitted specialist discourse which would give away the disciplinary identity of those involved. Where it has been necessary to do this, I have in parentheses and italics included a brief description of what has been omitted.

- (•) pause, one (•) per second
- [] the vertically aligned bracketed text is simultaneous speech
- ↑ rising intonation
- ↓ falling intonation
- (*) unclear word
- under emphasis
- / interruption
- K Kenneth
- V Victor
- E Ella
- P Philip
- D Daniel
- X unknown

- K Mr R (*the student*), the supervisor gave it 80, the second marker gave it 65, and the externals recommended 70.
- V No ↓
- K Dr. V ↓
- 5 V so um what happened was yesterday we sat down myself and uh [*the second marker*], when we did a lot of (*) and we agreed that (*) fair compromise would be around 76. Honestly speaking I don't see the reason why this student should be denied a first class, if the externals can justify that ↓ ••• If they can tell me where they think he missed it ••
- 10 V [please] ↓
- E Wel[l, we]didn't hear that presentation ••
- V No, no [the presentation actually deter]mines one of the best/
- 15 E [Ours is just based on the report]
- K We are basing this largely [on the report] so ↓
- X [on the report]
- 20 P Well looking at the report its a well written report which/ explains the theory fairly fairly well. the theory of um (x) that's actually quite well, and that constitutes about 80% of the thesis. The other 20% describes the work that was actually done and the major part with that work really revolves around understanding the (*a particular software package*), once one understands the (*package*), it becomes fairly straight-

- 25 forward, to be able to configure different network topologies, and from those network typologies determine the (****). So/ it seems it was a lot of literature survey done, the work that was done was actually done quite well, but/ not meriting a first.
- V Uh can I just please answer this one↓ I think this guy actually spent a lot of time. In the first place I think we've got to understand that (*the package*) is quite a very difficult package to master. Plus basically what he was looking after he had validate to make sure that the (****) is within the (*speed*) which is the whatever the maximum which is acceptable that is for real-time (**) over any network. So that is what he was trying to investigate. And he tried that with several configurations including the whatever (*name of something*) which is based on the market protocol levels. He also had to try different configurations especially between Mauritius and across Africa and so forth and in most cases he was actually exceeded the (*speed*) which is the acceptable delay. So in this case I thought he did exactly what was expected of him, so in this case I mean what is wrong? Really if you are to compare with some of the theses that are said to be whatever with whatever that are said whatever in whatever (*), don't you think there is too much discrepancy here? The student has done a lot of hard work/ but then you find a way of saying, no, I don't think he has done the whatever/ no I don't think that I uh agree with you, not at all. I personally would think that a 76/ or so/ would be the fairest mark for that stu[dent.]
- 30
- 35
- 40
- K [Could]
- 45 we could we ask the meeting to rule on it, are we going to accept the mark of 70 from the external examiners ↑
- D Ja, let me this thesis had a third marker Prof. __ and I think he it's [72]
- X [ja] I think the report was unbalanced as mentioned. Because uh although the results are uh reported I think are four pages, there were no in-depth discussion of the results and I think that is what you would expect from a first [pass].
- 50
- V [Yes], but as far as I know what he intended to do there was to check whether in that the (**) exceeds the expected whatever the threshold which is (*the speed*), yes or no. In as much as let's say for instance he wanted to establish let's say a real time connection across a (**) it depends (*****). That's what you'll see in (**) agrees with that. If it does agree with that. that's fine. Why would he start putting up junk. filling up the whole whatever with junk. I see that most of the students here what they do is they will actually get whatever problem and they just think (*)relevant, whatever problem from whatever american search and whatever put them in the report just to try to spice it up when in fact it's of no relevance at all. But here is a student who is trying whatever to report whatever his work, I mean present his work in a very mature manner, (*) I don't think that is fair.
- 55
- 60
- K Thank you Dr V. I wonder if we could ask the meeting to that we go with the mark of 70% (*) from the externals.
- 65 X (a few nonverbal noises of agreement)

REFERENCES CITED

- Angoff, W. 1988. Validity: An evolving concept. In *Test Validity*, edited by H. Wainer and H. Braum, pp. 19-30. Hillsdale, New Jersey: Lawrence Erlbaum.
- Barritt, L., P. Stock, and F. Clark. 1986. Researching Practice: Evaluating Assessment Essays. *College, Composition and Communication*, 37, no. 3: pp. 315-27.
- Becher, T. 1989. *Academic tribes and territories: Intellectual inquiry and the cultures of disciplines*. Buckingham: Society of Research in Higher Education and Open University Press.
- . 1989. *Academic tribes and territories: Intellectual inquiry and the cultures of disciplines*. Buckingham: Society of Research in Higher Education and Open University Press.
- Becher, T., and P. Trowler. 2001. *Academic Tribes and Territories: Intellectual enquiry and the cultures of disciplines (2nd Edition)*. Society for Research into Higher Education and Open University Press.
- Beckett, D., and P. Hager. 2000. Making judgments as the basis for workplace learning: Towards an epistemology of practice. *International Journal of Lifelong Education*, 19, no. 4: pp. 300-11.
- Berlack, H. 2000. Cultural politics, the science of assessment and democratic renewal of public education. In *Assessment: Social practice and social product*, edited by A. Filer, pp. 189-207. London: Routledge and Falmer.
- Berlack, H., F. Newmann, E. Adams, T. Archbald, and T. Romberg. 1992. *Toward a New Science of Educational Testing and Assessment*. New York: State University of New York Press.
- Bernstein, R. 1976. *The restructuring of social and political theory*. New York: Harcourt Brace Jovanovich.
- . 1983. *Beyond objectivism and relativism: Science, hermeneutics, and praxis*. Philadelphia: University of Pennsylvania.
- . 1991. *The new constellation: The ethical-political horizons of modernity/postmodernity*. Cambridge: Polity Press.
- Bhaskar, R. 1979. *The possibility of naturalism*. Hemel Hempstead: Harvester Wheatsheaf.
- Biggs, J. 1996. Assessing learning quality: reconciling institutional, staff and educational demands. *Assessment and Evaluation in Higher Education*, 21, no. 1: pp. 5-15.
- Blommaert, J. 2001. Context is/as critique. *Critique of Anthropology*, 21, no. 1, March: pp. 13-32.
- Blommaert, J., and C. Bulcaen. 2000. Critical Discourse Analysis. *Annual Review of Anthropology*, 29: pp. 447-66.
- Bloor, T., and M. Bloor. 1995. *The functional analysis of english: A Hallidayan approach*. London: Arnold.
- Bohman, J. 1991. *New philosophy of social science*. Cambridge: Polity Press.
- Bourdieu, P. 1988. *Homo Academicus*. Stanford, California: Stanford University Press.
- . 1990. *The logic of practice*. Stanford, CA: Stanford University Press.
- . 1996. *The state nobility: Elite schools in the field of power*. Cambridge: Polity Press.
- . 1998. *Practical Reason: On theory of action*. Stanford, CA: Stanford University Press.

- Bourdieu, P., and L. Wacquant. 1992. *An invitation to reflexive sociology*. Cambridge: Polity Press.
- Broad, B. 2000. Pulling your hair out: Crises of standardization in communal writing assessment. *Research in the Teaching of English*, 35 (November): pp. 213-60.
- Broadfoot, P. 1996. *Education, assessment and society: A sociological analysis*. Buckingham: Open University Press.
- . 1999. Empowerment or performativity? English assessment policy in the late twentieth century. Paper presented as part of the Assessment Reform Group Symposium on Assessment Policy presented at the British Educational Research Association Annual Conference, University of Sussex at Brighton.
- . 2000. Preface. In *Assessment: Social practice and social product*, edited by A. Filer, p. ix-xii. London: Routledge and Falmer.
- Broadfoot, P., and A. Pollard. 2000. The changing discourse of assessment policy: The case of English primary education. In *Assessment: Social practice and social product*, edited by A. Filer, pp. 11-26. London: Routledge and Falmer.
- Brown, P., and S. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Brown, S., and P. Knight. 1994. *Assessing Learners in Higher Education*. London: Kogan Page.
- Bunting, I. 2001. The higher education landscape under apartheid. In *Transformation in higher education: Global pressures and local realities in South Africa*, edited by N. Cloete, R. Fehnel, P. Maaseen, T. Moja, H. Perold, and T. Gibbon, pp. 58-86. Lansdowne, South Africa: Juta Press.
- Calhoun, C. 1995. *Critical Social Theory: Culture, history, and challenge of difference*. Oxford: Blackwell.
- Chouliariki, L., and N. Fairclough. 1999. *Discourse in Late Modernity*. Edinburgh: Edinburgh University Press.
- Clark, B. 1983. *The Higher Education system*. Berkeley, CA: University of California Press.
- . 1987a. *The academic life: Small worlds, different worlds*. Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.
- . 1987b. *The academic profession: National, disciplinary and institutional settings*. Berkeley: University of California Press.
- . 1998. *Creating entrepreneurial universities: Organizational pathways of transformation*. Paris, France: IAU Press.
- Cloete, N., and P. Maaseen. 2002. The limits of policy. In *Transformation in higher education: Global pressures and local realities in South Africa*, edited by N. Cloete, P. Fehnel, T. Maaseen, H. Moja, H. Perold, and T. Gibbon, pp. 447-90. Lansdowne, South Africa: Juta Press.
- Cronbach, L. 1971. Test Validation. In *Educational Measurement (2nd Edition)*, edited by R. Thorndike, pp. 443-507. Washington: American Council on Education.
- . 1988. Five perspectives on the validity argument. In *Test Validity*, edited by H. Weiner and H. Braun, pp. 3-17. Hillsdale, NJ: Lawrence Erlbaum Associates.
- . 1989. Construct validation after thirty years. In *Intelligence: Measurement, theory, and public policy: Proceedings of a symposium in honor of Lloyd G. Humphreys*, edited by R. Linn, pp. 146-71. Urbana: University of Illinois Press.
- Davis, A. 1998. Special Issue: The Limits of Educational Assessment. *The Journal of the Philosophy of Education*, 32, no. 1, March: pp. 1-150.

- D'Andrea, V., D. Gosling, and S. Shay. 2001. Implementing outcomes-based approach in curriculum design in higher education: Comparisons between the United Kingdom and South Africa. Co-presented paper presented at the Society for Research in Higher Education Conference, Cape Town, South Africa.
- Engestrom, Y. 1999. Activity theory and individual and social transformation. In *Perspectives on activity theory*, edited by Y. Engestrom, R. Miettinen, and R. Punamaki, pp. 19-38. Cambridge: Cambridge University Press.
- Fairclough, N. 1989. *Language and Power*. London: Longman.
- . 1992. *Discourse and Social Change*. Cambridge: Polity Press.
- . 1993. Critical Discourse Analysis and the Marketization of Public Discourse. *Discourse and Society*, 4, no. 2: pp. 133-68.
- Filer, A. 2000. *Assessment: Social practice and social product*. London: Routledge and Falmer.
- Flyvbjerg, B. 2001. *Making social science matter: Why social science inquiry fails and how it can succeed again*. Cambridge: Cambridge University Press.
- Gee, J. 1996. *Social Linguistics and Literacies: Ideology in Discourses*. Basingstoke: Falmer Press.
- . 1999. *An introduction to discourse analysis: Theory and method*. London: Routledge.
- . Forthcoming. Critical discourse analysis. In *Multidisciplinary perspectives on literacy research*, edited by R. Beach, J.L. Green, and M.L. Kamil. Urbana, Illinois: National Council for Teachers of English.
- Gipps, C. 1991. *Beyond Testing: Towards a Theory of Educational Assessment*. London: Falmer Press.
- . 1995. *Assessment for the Millennium: Form, Function and Feedback*, Inaugural Lecture. University of London: Institute of Education.
- . 1999. Socio-cultural aspects of assessment. *Review of research in education/ AERA*, 24: pp. 355-92.
- Goodwin, C. 1994. Professional vision. *American anthropologist*, 96, no. 3: pp. 606-33.
- Hager, P. 2000. Know-how and workplace practical judgement. *Journal of Philosophy of Education*, 34, no. 2: pp. 281-96.
- Halliday, M.A.K., and R. Hasan. 1985. *Language, Context and Texts: Aspects of Language in a Social-Semiotic Perspective*. Geelong: Deakin University Press.
- Halliday, MAK. 1994. *An introduction to functional grammar Second Edition*. London: Edward Arnold.
- Hatch, J., C. Hill, and J. Hayes. 1993. When the Messenger is the Message. *Written Communication*, 10, no. 4: pp. 569-98.
- Henkel, M. 2000. *Academic identities and policy change in higher education*. London: Jessica Kingsley.
- Higher Education Quality Council. 2002. *Improving Teaching and Learning Project: Interim Guides to Good Practice*. Draft document.
- Huot, B. 1990a. The Literature of Direct Writing Assessment: Major Concerns and Prevailing Trends. *Review of Educational Research*, 60, no. 2: pp. 237-63.
- . 1990b. Reliability, Validity, and Holistic Scoring: What We Know and What We Need to Know. *College, Composition and Communication*, 41, no. 2: pp. 201-13.
- . 1996. Toward a New Theory of Writing Assessment. *College, Composition and Communication*, 47, no. 4: pp 549-66.

- Jawitz, J., R. Moore, and S. Shay. 2002. Management and Assessment of Final Year Projects in Engineering. *The International Journal of Engineering Education*, 18, no. 4: pp. 472-478.
- Kane, M. 2001. Current concerns in validity theory. *Journal of Educational Measurement*, 38, no. 4, Winter: pp. 319-42.
- Kasper, G. 1997. Linguistic Etiquette. In *The Handbook of sociolinguistics*, edited by F. Coulmas, pp. 374-85. Oxford: Basil Blackwell.
- Kraak, A, and N. Mahomed. 2001. Qualifications reform in higher education: An evaluation of the work of national standards bodies. In *Curriculum restructuring in higher education in post-apartheid South Africa*, edited by M. Breier, pp. 139-56. University of the Western Cape, South Africa: Education Policy Unit.
- Kress, G. 1989. *Linguistic Processes in Sociocultural Practice*. U.K.: Oxford Univeristy Press.
- Kvale, S. 1993. Examinations reexamined: Certification of student or certification of knowledge? In *Understanding practice: Perspectives on activity and context*, edited by S. Chaiklin and J. Lave, pp. 217-40. Cambridge: Cambridge University Press.
- Linn, R.L., E. Baker and S. Dunbar. Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher*, 20, no. 8: pp. 15-21.
- Linn, R.L. 1993. Educational Assessment: Expanded Expectations and Challenges. *Educational Evaluation and Policy Analysis*, 15, no. 1: pp. 1-16.
- Luke, A. 1995-1996. Text and Discourse in Education: An Introduction to Critical Discourse Analysis. In *Review of Research in Education 21*, edited by M. Apple, pp. 3-48. Washington D.C.: American Educational Research Association.
- Maaseen, P., and N. Cloete. 2002. Global reform trends in higher education. In *Transformation in higher education: Global pressures and local realities in South Africa*, edited by N. Cloete, R. Fehnel, P. Maaseen, T. Moja, H. Perold, and T. Gibbon, pp. 13-57. Landsdowne, South Africa: Juta Press.
- Messick, S. 1989. Meaning and Values in Text Validation: The Science and Ethics of Assessment. *Educational Researcher*, 18, no. 2: pp. 5-11.
- . 1995. Standards of Validity and the Validity of Standards in Performance Assessment. *Educational Measurement: Issues and Practice*, 14, no. 4: pp. 5-8.
- . 1998. Test validity: A matter of consequences. *Social Indicators Research*, 45: pp. 35-44.
- Miller, R. 1996. Mark my Words Part 1: Teachers. *South African Journal of Higher Education*, 10, no. 2: pp. 13-24.
- Mishler, E. 1990. Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Education Review*, 60, no. 4, November: pp. 415-42.
- Moss, P.A. 1992. Shifting Conceptions of Validity in Educational Measurement: Implications for Performance Assessment. *Review of Educational Research*, 62, no. 2: pp. 229-58.
- . 1994. Validity in High Stakes Writing Assessment: Problems and Possibilities. *Assessing Writing*, 1, no. 1: pp. 109-28.
- . 1996. Enlarging the Dialogue in Educational Measurement: Voices from Interpretive Research Traditions. *Educational Researcher*, 25, no. 1: pp. 20-28.
- . 1998a. The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, Summer: pp. 6-12.

- . 1998b. Rethinking Validity for the Assessment of Teaching. In *With portfolio in hand: Validating the new teacher professionalism*, edited by N. Lyons, pp. 202-19. New York: Teachers' Colleague Press.
- Moss, P.A., D Pullin, J. Gee, and E. Haertel. 2002. The idea of testing: Expanding the foundations of educational assessment. Draft paper.
- Moss, P.A., and A. Schutz. 2001. Educational standards, assessment, and the search for consensus. *American Educational Research Journal*, 38, no. 1: pp. 37-70.
- Moss, P.A., A. Schutz, and K. Collins. 1998. An integrative approach to portfolio evaluation for teacher licensure. *Journal of Personnel Evaluation in Education*, 12, no. 2: pp. 139-61.
- Muller, J. 2000. *Reclaiming Knowledge: Social theory, curriculum and education policy*. London: Routledge.
- Nietzsche, F. 1969. On the Genealogy of Morals. Trans. W. Kaufman and R.J. Hollingdale. New York: Vintage Books.
- Nitko, A. 2001. *Educational assessment of students Third Edition*. Upper Saddle River, New Jersey: Merrill Prentice Hall.
- Pennycook, A. 1994. Incommensurable discourses? *Applied Linguistics*, 15, no. 2: pp. 115-38.
- Phelps, L. 1989. Images of Student Writing: A Deep Structure of Teacher Response. In *Writing and Response: Theory, Practice, Research*, edited by C. Anson. Urbana, IL: National Council for the Teacher's of English.
- Piper, D.W. 1994. *Are Professors Professional? The Organization of University Examinations*. Higher Education Policy Series, vol. 25. London: Jessica Kingsley Publishers.
- Postone, M., E. LiPuma, and C. Calhoun. 1993. Introduction: Bourdieu and critical theory. In *Bourdieu: Critical perspectives*, edited by C Calhoun, E. LiPuma, and M. Postone, pp. 1-13. Chicago: Chicago University Press.
- Rabinow, P., and W. Sullivan. 1987. The interpretive turn: A second look. In *Interpretive social science: A second look*, edited by P Rabinow and W. Sullivan, pp. 3-30. Berkeley: University of California Press.
- Resnick, L., and D. Resnick. 1992. Assessing the thinking curriculum: New tools for educational reform. In *Changing Assessments: Alternative views of aptitude, achievement and instruction*, edited by B Gillford and M. O'Connor, pp.37-75. London: Kluwer Academic Publisher.
- Schwegler, R.A. 1991. The Politics of Reading Student Papers. In *The Politics of Writing Instruction: Post-Secondary*, edited by R. and Trimbur Bullock, J. Portsmouth, N.H.: Heinemann.
- Shepard, L. 1993. *Review of Research in Education/AERA*, 19: pp. 405-50.
- Slembrouck, S. 2001. Explanation, interpretation and critique in the analysis of discourse. *Critique of Anthropology*, 21, no. 1, March: pp. 33-57.
- Starfield, S. 2001. 'I'll go with the group': Rethinking 'discourse community' in EAP. In *Research perspectives on English for Academic Purposes*, edited by J. Flowerdew and M. Peacock, pp. 132-47. Cambridge: Cambridge University Press.
- Sutherland, L. 1999. Moderating marks or negotiating consensus? *Academic Development*, 4, no. 1: pp. 167-78.
- Taylor, C. 1987. Interpretation and the Sciences of Man. In *Interpretive Social Science: A second look*, edited by P. Rabinow and W. Sullivan, pp.33-81. Berkeley: University of California Press.

- Ting-Toomey, S. 1994. Face and facework: An introduction. In *The challenge of facework*, edited by S. Ting-Toomey, pp. 1-14. New York: State University of New York Press.
- Titscher, S., M. Meyer, R. Wodak, and E. Vetter. 2000. *Methods of text and discourse analysis*. London: Sage Publications.
- Trowler, P. 2001. Captured by the discourse? The socially constitutive power of new higher education discourse. *Organization*, 8, no. 2: pp. 183-201.
- van Dijk, T.A. 1993. Editor's foreword to Critical Discourse Analysis. *Discourse and Society*, 4, no. 2: pp. 131-32.
- Verschueren, J. 2001. Predicaments of criticism. *Critique of Anthropology*, 21, no. 1, March: pp. 59-81.
- Wenger, E. 1998. *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.
- Widdowson, H.G. 1998. The Theory and Practice of Critical Discourse Analysis. *Applied Linguistics*, 19, no. 1: pp. 136-51.
- . 2000. On the limitations of linguistics applied. *Applied Linguistics*, 21, no. 1: pp. 3-25.
- Yancey, K.B., and B. Huot. 1996. One Book: Three Reviews. *Assessing Writing*, 3, no. 2: pp. 197-201.