Issues in interpreting student feedback statistical data

PUBLISHER

© Dean Polworth

REPOSITORY RECORD

Polworth, Dean. 2019. “Issues in Interpreting Student Feedback Statistical Data”. figshare. https://hdl.handle.net/2134/35214.

BLDSC no:- DX229573

Issues in Interpreting Student Feedback Statistical Data

by

Dean Polworth

A Doctoral Thesis
Submitted in partial fulfilment of the requirements for the award of the degree of
Doctor of Philosophy of Loughborough University

April 2001

Dedication

To my mother and father

Acknowledgements

I would like to thank all those individuals who have provided me with support, opportunity and encouragement during my research. As joint supervisor Dr. David Coates provided me with guidance throughout the course of my research and Professor Adrian Wilkinson offered me advice and encouragement when I most needed it.

I would especially like to thank Professor Malcolm King, a man of immense integrity and honour, from whom I have learnt so much.

Abstract

Student feedback ratings are becoming an increasingly utilised means of measuring the quality of the student learning experience in U.K. universities. Despite this there has been little published statistical analysis on student feedback ratings using U.K. data.

The research explores whether lecturers should have confidence in the validity of the student feedback ratings they receive. Guidance in the presentation and interpretation of the data is offered. This is intended to facilitate a more sophisticated understanding of the data, allowing decisions based on it to be made on a more informed basis. The research used the student feedback data collected on all taught modules (both undergraduate and postgraduate) in the Business School at Loughborough University over two academic years (October 1996 - June 1998). This consisted of 305 modules and 13813 individual student feedback forms.

The thesis contributes to the literature in the following ways:

1. Through illustrating the existence of heterogeneous groups of students in many Business School modules, which reflect the presence of different learning styles being utilised by Business School students, and discussing the consequences of this for:
   a) the use of factor analysis on student feedback data;
   b) the appropriateness of reporting the results of student feedback in the form of class averages.

2. Through illustrating the effects on student feedback ratings specific to modules taught by more than one lecturer. Two variables not previously reported in the student feedback literature are shown to influence the ratings that lecturers receive, namely:
   a) the proportion of lecturing hours undertaken by a lecturer on a particular module;
   b) the ratings of the lecturer(s) with whom a lecturer teaches alongside on a particular module.

3. Through examining the impact of external factors on the ratings lecturers receive. Regression analysis is used to model the influence of a set of nine predictor variables on student feedback ratings. Lecturers' ratings are shown to be significantly influenced by the level of the module, the class size and the subject area of the module. Characteristics of the lecturer, namely, the lecturer's age, rank and experience are shown to significantly influence lecturers' ratings for some aspects of lecturing.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION AND OVERVIEW

CHAPTER 2: THE DIVERSITY OF THE HIGHER EDUCATION ENVIRONMENT

CHAPTER 3: A REVIEW OF THE STUDENT FEEDBACK LITERATURE

CHAPTER 4: DATA GATHERING

CHAPTER 5: METHODS OF DATA ANALYSIS

CHAPTER 6: SYSTEMATIC INTERRATER VARIANCE

CHAPTER 7: CLUSTER STRUCTURES IN THE DATA

CHAPTER 8: THE EFFECT OF POSITION

CHAPTER 9: CONTRAST EFFECTS IN STUDENT FEEDBACK RATINGS

CHAPTER 10: A GENERAL MODEL OF LECTURERS' RATINGS

CHAPTER 11: CONCLUSIONS AND IMPLICATIONS

APPENDICIES

LIST OF TABLES

LIST OF FIGURES

Chapter 1
INTRODUCTION AND OVERVIEW

1.1 Introduction

The last 40 years has witnessed a considerable change in higher education. Higher education has considerably expanded with a substantial increase in the number of students entering higher education. This expansion has lead to an increased diversity in the nature of institutions offering undergraduate education and an increased diversity in the students entering higher education. The rapid expansion of higher education has led to increasing pressures on government funding. Governments have responded by desiring more formalised mechanisms for monitoring the product for which considerable sums of government money is spent. This research is an investigation into one of these mechanisms for monitoring the quality of the student experience in higher education, namely, student feedback data collected in the Business School at Loughborough University.

This chapter sets the scene for the research by placing student feedback in its historical context. It also describes the outline of the thesis and the contributions that the thesis intends to make to the student feedback literature. Section 1.2 outlines the early research on student feedback data dating back to the 1920's and the reasons behind the increased use of student feedback in the 1970's. The section also outlines the differences between the U.K. and the U.S. in terms of the acceptance of student feedback data as a method of measuring the quality of modules and the skills of lecturers. The reasons for the increased usage of student feedback in U.K. universities in the 1990's are discussed. Section 1.3 presents the aims of the research, outlining the contribution that achieving each aim will make to the published research on student feedback. Finally in section 1.4, the structure of the thesis is presented with the content of each of the chapters briefly described.

1.2 The development of student feedback

1.2.1 Early research into student feedback data

Research into student feedback data dates back to the 1920's. The University of Washington, the University of Texas, Purdue University and Harvard all introduced programmes to systematically collect students' views of the teaching they received in the mid 1920's. The pioneer of the system at Purdue University, H. H. Remmers has been described as the 'father of research into students' evaluations of teaching effectiveness' by Herbert Marsh, himself a leading contemporary light in the student feedback field (Marsh, 1987). Remmers' early work on analysing the data from student feedback included analysing the reliability, validity, halo effects and potential biases in the ratings (Remmers and Brandenberg, 1927). Remmers went on to analyse the relationship between grades and student feedback ratings (Remmers, 1928) and the degree to which multiple traits of lecturing effectiveness can be distinguished from each other and differentiated in importance (Stalnaker and Remmers, 1928). Remmers was also the first researcher to publish a factor analysis of student feedback data (Smalzried and Remmers, 1943).

1.2.2 Reasons for the increase in research on student feedback data

The early 1970's saw a large increase in the collection of student feedback data in the U.S. This growth in available data spawned a plethora of studies that purported to analyse it. There were three main reasons for this increase in the collection of student feedback data:

a) The 1960's saw a significant change in the make up of the U.S. undergraduate student body. Wider access to university education implied that both the diversity of students had increased and that the amount of public funds being spent on university education had also increased. As a result, lecturers had to become more responsive to the demands that the students were placing on them and devote an increasing proportion of their time to lecturing. At the same time, the public were demanding a greater degree of monitoring of the quality of the product that ever increasing amounts of public money was being spent on (Murray, 1984).

b) Improvements in technology allowed student feedback forms to be read automatically by machine. This greatly reduced the amount of labour involved in creating data sets and therefore made larger scale data collection exercises more feasible. As will be discussed in chapter 4, the machines that can automatically read the student feedback forms, called optical mark readers, are not infallible and therefore do not eliminate the need for original forms to be manually checked to 'complete' data sets.

c) Improvements in computer technology made it possible to analyse large data sets quickly and relatively easily and this allowed researchers to employ more rigorous and potentially revealing statistical tests on the data. Prior to these developments in computer technology, a technique that is based on correlations between variables would have been very time consuming on anything other than the smallest data set.

The improvement in technology and the growing interest in collecting students' views of the education that they were receiving lead to a considerable increase in the use of student feedback data. There are now very few U.S. universities that do not employ student feedback as a measure of lecturing evaluation and there are now more than 1500 references dealing with research on student feedback (Cashin, 1995).

1.2.3 Student feedback research in Europe

The systematic collection and analysis of student feedback data in the U.K., has lagged considerably behind the collection and analysis of student feedback data in the U.S. Over the last decade however, there has been a growing interest in collecting student feedback data in the U.K. The U.K. is not alone in Western Europe in showing a growing interest in collecting student feedback data. In Holland the process of collecting student feedback data is widely accepted and promoted by the Government. In 1990 the Dutch Minister for Education stated that it was important to systematically gather students' opinions about the teaching they receive (Willems et al., 1992 - quoted in Husbands, 1996). France and Germany have not followed the U.K. and Dutch example of introducing systematic student feedback systems. In France there is assessment of lecturing by questionnaires completed by students, but such practices are not routine (Husbands and Frosh, 1993). In Germany there is little consensus regarding how quality in university departments should be measured and consequently there has been less progress in implementation (Husbands and Frosh, 1993). Husbands and Frosh describe this process as 'a patchwork of various federal and regional initiatives' (Husbands and Frosh, 1993, p. 99). The major theme in assessment in German universities is how best to revise courses to accelerate students' progress in completing their courses (Husbands and Frosh, 1993).

In the U.K. there has been concern about the purposes for which the resulting data might be used (Husbands, 1997). Compared to the U.S. there has been much less focus in the U.K. literature on the appropriate methods of analysing the data and the issue of possible biases in the data (Husbands, 1996). Recent empirical statistical studies by Husbands (1996, 1997) have added to the knowledge of U.K. student feedback data but there are few other studies that rigorously analyse the data derived from student feedback gathered in the U.K.

1.2.4 The introduction of quality assurance in U.K. Higher Education

The use of student feedback data as a measure of the quality of the teaching and learning environment increased in the U.K. during the 1980's. The Academic Audit Unit (AAU) of the Committee of Vice-Chancellors and Principals (CVCP) was created in 1990. It laid emphasis on the need for effective methods of student feedback in the universities (Silver, 1992). The 1992 Further and Higher Education Act enshrined the principle of formalised monitoring of teaching quality as a feature of U.K. higher education. As a result, the AAU was subsumed into a larger organisation the Higher Education Quality Council (HEQC). The HEQC's parameters embraced the higher education system as a whole and became the central focus for efforts in maintaining and enhancing quality in higher education (Webb, 1994).

Government policy resulted to some extent in the redefinition of students as customers or clients and therefore in an increased interest in ascertaining their views (Silver, 1992). The Higher Education Funding Councils began to regard students' views of the education they received as a significant element of quality assessment in universities (Green, et al., 1994).

In reaction to the increased emphasis on monitoring quality in U.K. universities, the Quality Assessment Unit (QAU) at Loughborough University was set up in 1994.
The QAU has two main functions:

- to implement the University's internal teaching assessment scheme;
- to provide departments involved in QAA/OFSTED assessments with assistance in their submissions and preparations for visits.

1.3 Aims of the research

As will be shown in detail in chapter 2, the last 40 years has been a period of extensive change for U.K. higher education. The number of students enrolled in university degree courses has considerably increased. Alongside the increase in the size of higher education has come a greater degree of diversity in the students present in lecture theatres. The proportion of female students, 'mature' students, students from working class backgrounds and students from the ethnic minorities have all increased as has the number of institutions offering undergraduate education.

As also shown in the next chapter one result of the increase in the number of students studying in U.K. universities is that the need for government financial support for higher education has increased. This has led to an increasing desire from governments to monitor the quality of higher education, aiming to ensure that state funding is put to good use.

This research aims to consider whether the student feedback system in place in the Business School at Loughborough University is congruent with the increasing diversity common in higher education. In particular the research questions whether the assumptions prevalent in the student feedback literature are congruent with the actual behaviour and expectations of students as manifested in their chosen learning style and the behaviour and expectations of lecturers as manifested through their teaching style.

The need to consider the appropriateness of student feedback as a mechanism of measuring the quality of the student learning experiences and its usefulness in the potential improvement of modules and lecturing is important given the manner in which the introduction of formalised mechanisms of measuring quality in higher education has been received in some quarters.

The Higher Education Quality Council reported in 1994 that the increased systemisation of the collection of student feedback data had lead to 'tension between control on the one hand and ownership on the other'. This had been noted in many of the 'Academic Audits' that the HEQC had conducted in U.K. universities.

> 'Attempts to regularise practices [on either a departmental or campus wide basis] had met with some resistance [particularly when the] new standardised

> system had been superimposed on and merged with existing procedures' (Higher Education Quality Council, 1994, 29).

Some Business School based researchers have been more candid in their criticism of the collection of student feedback data. Grey and Mitev (1995) contend that systems for collecting student feedback data are not intended to improve the students' learning experiences, but rather reflect a more cynical desire of managers to impose increased control over the education process:

> 'attempts on teaching quality...are an attempt to introduce bureaucratic management controls into universities, and within this student evaluations play a key role' (Grey and Mitev, 1995, 82).

Grey and Mitev further contend that:

> 'teaching quality assessment exercises have an insidious conforming and confirming effect on commonsensical and technicist approaches to teaching management students' (Grey and Mitev, 1995, 83).

Grey and Mitev explicitly describe their argument as 'a polemic', but it is clear that if such views are widespread amongst lecturing personnel one of the principal aims of collecting student feedback, namely the potential use that lecturers can make of the data as a diagnostic guide to actively improve their lecturing, will not be realised.

Much of the hostile reaction towards the collecting of student feedback data has been offered without recourse to analysis of the resulting data. This is a reflection of the fact that despite the increased use of student feedback in U.K. universities in the 1990's, there has been little published statistical analysis of student feedback data collected within the U.K.

For student feedback to serve the positive objectives of improving the quality of the student's learning experiences, through giving lecturers a tool from which to monitor their own teaching, lecturers need to have confidence in the resulting data. The data needs to be seen to be valid and reliable. Lecturers are often surprised to receive ratings that are noticeably different from the ratings that they were expecting; lecturers generally expect the ratings that they receive to show a degree of consistency. Substantial variations in the ratings that lecturers receive may be interpreted as revealing weaknesses in the data, particularly if the lecturer has taught the same module in broadly the same way, using the same prepared material, in consecutive years. This is closely linked to one of the observations that lecturers often make about their ratings which is that the changes that they make to a module and their teaching in light of student feedback ratings are not appreciated by the students the following year. Some lecturers may feel that this year on year inconsistency in the students' ratings shows that student attitudes are essentially fickle. This may be seen as a reason to doubt the worth of student feedback data, attempting to improve the module and teaching in light of student feedback ratings being akin to trying to hit a moving target.

This research will examine reasons for variations in lecturers' ratings, by investigating the influence that the interaction between students' learning styles and lecturers' teaching styles on the module have on lecturers' student feedback ratings. Students utilising different learning styles will approach their studies with different emphasis and correspondingly favour a different style of teaching from the lecturer. Different learning styles present amongst students in the class would provide a basis for students taking a different perspective towards evaluating the lecturer. Variations in lecturers' ratings for a module in consecutive years could be explained by a change in which learning style is dominant in the class in any particular year. This would provide a logical alternative explanation for changing ratings that does not imply that students are fickle in the ratings they give lecturers. The research will also address the issue of whether the students are rating lecturers absolutely (i.e. independently of student experiences of other lecturers) or relatively (i.e. in direct comparison to student experiences of other lecturers).

This research seeks to explore whether lecturers should have confidence in the ratings that they receive from students. It also aims to offer guidance in the presentation and interpretation of the data, drawing out features in the data that need to be taken into account when considering what action individual lecturers and managers should take in light of the data. This should allow decisions based on the results of student feedback data to be made on a more informed basis. In doing so the research is able to draw upon the student feedback data collected in the Business School at Loughborough University in the academic years 1996-1997 and 1997-1998. The student feedback data at Loughborough University is suitable for such analysis given the fact that the student feedback form used has been developed over a period of years (see section 4.2) and the amount of student feedback data held by in the Business School is quite considerable. Given the potential benefits of using the results of student feedback data for the improvement of student learning experiences together with the increasing usage of student feedback data in personnel decisions (e.g. promotion), it is clearly important to gain as deep an understanding of the data as possible.

Bearing the foregoing considerations in mind, the research sets out with three principal aims:

1. to examine the extent to which there are groups of students in Business School modules who psychologically perceive the quality of the module and the skills of the lecturer differently as a result of employing different learning styles and to consider the consequences of this for student feedback data;
2. to examine whether the psychological concept of contrast effects is applicable to student feedback data, whereby lecturers' student feedback ratings are influenced by the ratings of the other lecturer(s) teaching on a particular module;
3. to examine the impact of external factors on the ratings lecturers receive.

The first of these aims is to question one of the most prevalent assumptions in the published student feedback literature, namely that it is valid to treat the class as a single coherent homogeneous entity. The research will explore whether the presence of different student learning styles in a class invalidates the assumption that a class can be treated as a homogeneous entity. The research will explore the consequences that heterogeneous groups of students in a module have for:

- the use of factor analysis on student feedback data;
- the appropriateness of reporting the results of student feedback in terms of class averages.

Both of these practices are very common in the published student feedback literature. They are two separate issues, linked by the underlying assumption that it is appropriate to treat the students in a class as a single coherent homogeneous entity. Both the use of factor analysis and the reliance on class average ratings as a measure of student satisfaction would be questionable if heterogeneous groups (or clusters) of students were to be regularly found in the student feedback data for individual modules, yet a systematic analysis of the presence and implications of heterogeneous groups of students in modules who systematically view the quality of the module and skills of the lecturer differently has not featured in the student feedback literature.

The second of these aims is to examine whether the psychological concept of contrast effects by which, in modules taught by more than one lecturer, the ratings that a lecturer receives are influenced by the ratings of those with whom they lecture. The Business School student feedback data provides an excellent opportunity to examine the consequences of shared lecturing responsibilities for lecturers' ratings, given that many modules in the Business School are taught by more than one lecturer. In other words, the analysis seeks to measure whether the ratings that lecturers receive are absolute (in the sense of being independent of comparative judgements) or relative. Examination of these issues has not featured in the student feedback literature.

The third aim is to examine which factors significantly influence the ratings lecturers receive. In modelling lecturers' ratings against a range of predictor variables that could influence the ratings lecturers receive, the research is able to make use of a large overall data set. The data set consists of the student feedback data collected on every module taught in the Business School over two academic years at both undergraduate and taught course postgraduate level. There is therefore no element of module selection in the data used to model lecturers' ratings. In total the data consists of 305 modules and 13814 individual student feedback forms.

Given that most of the empirical analysis of student feedback data is U.S. in origin, using data collected in U.S. universities, this research aims to contribute to the published student feedback literature though offering an empirical analysis of the variables that influence student feedback ratings in a U.K. university and specifically in a Business School context. The research also intends to comment on the validity of the data at measuring the quality of the student's learning experiences. It is hoped that the analysis will prove to be of use as a guide for individual lecturers and departmental managers in interpreting student feedback data, allowing decisions made in light of the data to be made on a more informed basis.

1.4 Outline of the thesis

The thesis is organised into 11 chapters. The content of the remaining chapters is outlined below.

Chapter 2 discusses the changing nature of higher education over the last 40 years. The chapter outlines the expansion of higher education focusing on the considerable increase in diversity now common in U.K. universities. The need for increased levels of state financial support for higher education over this period and the consequent desire by Governments to more closely monitor the quality of education provided by universities is also considered. Various views of the purposes of monitoring and evaluating higher education are considered. The dynamics of the imposition of an evaluation system are also discussed. Different learning styles utilised by students and different teaching styles employed by lecturers are discussed in detail and consideration is given to the implications of these differences for the student feedback ratings lecturers receive.

Chapter 3 reviews the main themes in the published research on student feedback data. The main purposes of collecting student feedback data are discussed as well as evidence of how useful student feedback data is at meeting these purposes. Published research that examines the reliability and validity of student feedback data are outlined as is the published research on a set of fourteen variables that could influence or bias the student feedback ratings lecturers receive. Reasons for the current lull in the published research on student data are proposed. A pervasive underlying assumption common to most of the published student feedback literature is highlighted. Reasons for questioning the appropriateness of this assumption in light of the changing nature of higher education are discussed and some of the implications for student feedback data should this assumption be shown to be flawed (which are explored in more depth in later chapters) are considered.

Chapter 4 describes the data gathering process at Loughborough University. The student feedback form used in the Business School at Loughborough University is presented and discussed. The lecturer-specific questions are compared to those on three alternative student feedback questionnaires that have featured in published research on student feedback. The mechanism of collecting the data is outlined together with the method by which the student feedback forms are read by machine and the method by which the data files sent to departments are constructed. An analysis of the extent to which the machine fails to read responses is presented.

Chapter 5 describes how the data files received by the Business School were transformed into the data files used in this research. Detailed discussion of the additional descriptive variables added to the data sets is presented as well as the reasons for having versions of the data sets at both individual form and class average levels. The three main multivariate statistical techniques used in the research, cluster analysis, factor analysis and multiple regression are discussed in detail. Emphasis is placed on explaining the reasons for selecting the chosen methods and outlining why alternative methods that could have been used were not chosen.

Chapter 6 is the first of two chapters that examines the implications for student feedback data of considering the class as a single coherent homogeneous entity when there are distinct groups of students in the class who are utilising different learning styles. In chapter 6 the focus is on the appropriateness of using factor analysis as a tool to analyse student feedback data when there are groups of students who systematically view the quality of the module and the skills of the lecturer(s) differently as a result of utilising different learning styles which lead them to desire different emphasis in the teaching offered by the lecturer. Both Business School student feedback data and simulated data with known cluster structures are used to illustrate that evidence of factor structures in student feedback data may often be illusory, being the product of correlations between cases reflecting the presence of a cluster structure (i.e. heterogeneous groups of students in the class).

Chapter 7 is the second chapter examining the implications of heterogeneous groups being present in the class. In chapter 7 the focus is on examining whether the potential problems that heterogeneous groups can cause in the analysis and presentation of student feedback data are actual problems requiring a more subtle understanding of the data by lecturers and managers. The prevalence of which heterogeneous groups (or clusters of students) are present in student feedback data is explored through the analysis of a sample of 65 Business School modules. Four main methods were used to determine whether a cluster structure was present. These were the visual inspection of the dendrogram, the examination of the cluster means, contrasting the between and within-cluster correlations and the application of the Calanski and Harabasz stopping rule. The chapter further explores the implications of the presence of heterogeneous groups for the practice of reporting the results of student feedback in the form of class averages. The implications for the interpretation of student feedback data when heterogeneous groups are present are discussed.

Chapter 8 explores the extent to which the position that the lecturer appears on the student feedback form influences the ratings they receive. Initial evidence suggested that lecturers received higher ratings when they appeared as lecturer 1 compared to when they appeared as lecturer 2. Lecturers tended to receive their lowest ratings when they appeared as lecturer 3. The insights resulting from a set of interviews suggested that an alternative variable, the proportion of lecturing hours undertaken by a lecturer on a particular module, might explain some of the variation in lecturers' ratings. Regression analysis showed that the proportion of lecturing hours undertaken by a lecturer on a particular module did significantly influence the ratings lecturers' received. Once this variable had been taken into account, the position that the lecturer occupied on the student feedback form was no longer a significant predictor of student feedback ratings. The analysis highlights the importance of understanding the implications for student feedback ratings of the interpersonal relationship between the lecturer and the student. The significance of building rapport and developing trust with the students is discussed, as are the consequences of this finding for the consistency of lecturers' student feedback ratings over time.

Chapter 9 explores whether the psychological concept of contrast effects influences the ratings a lecturer receives on a particular module, whereby a lecturer's ratings are influenced by the ratings of those with whom they share lecturing responsibilities on a particular module. Two predictor variables that could be used to measure the degree to which lecturers' ratings are relative are described. The ability of the two variables to predict student feedback ratings are considered separately. The analysis shows that there is a significant contrast effect in lecturers' student feedback ratings. The effects of additional module-specific predictor variables are also considered and the managerial implications of the findings are discussed.

Chapter 10 presents a variety of multiple regression analyses of lecturers' ratings against a set of nine predictor variables. These predictor variables are shown to influence student feedback ratings either in the published literature or in chapters 8 and 9 of this research. Four potential predictor variables that could measure contrast effects in student feedback ratings are discussed. Lecturers' ratings are modelled using each of the variables measuring contrast effects separately, allowing comparisons to be made. The models are run separately on each of the lecturer-specific questions (i.e. each of the six lecturer-specific questions are used separately as a dependent variable), allowing comparisons of the extent to which the predictor variables differ in their impact on lecturers' ratings for the different questions. The implications for managerial decision making of these results are considered in light of some of the criticisms levelled against student feedback raised in chapter 2.

Chapter 11 recapitulates on the main findings of the research and considers the contribution that they make to the published student feedback research. The implications for student feedback data of the presence of different learning styles and different teaching styles found in the Business School data are considered. Consideration is also given to the relevance that the wording of the questions has for both the style of teaching encouraged in the Business School and the implications for the way in which lecturers are evaluated. Emphasis is also placed on the guidance that the findings can offer lecturers and departmental managers regarding the most appropriate interpretation of, and reaction to, student feedback data. A series of recommendations for managerial practice that arise from the research are outlined as well as suggestions for policies regarding the organisation of the student feedback system. The future role of student feedback in U.K. higher education is also considered. The limitations of the research are discussed and some recommendations for further research are offered.

Chapter 2
THE DIVERSITY OF THE HIGHER EDUCATION ENVIRONMENT

2.1 Introduction

This chapter considers how the expansion of higher education, particularly over the last forty years, has considerably increased the diversity of higher education. The chapter also considers how this expansion has led to the need for increased state support and monitoring of higher education partly as a consequence of the increasing amounts of government funding required to finance higher education and partly as a method of ensuring the quality of higher education in a time of rapid change.

The chapter begins with an outline of the increased diversity in the backgrounds of students entering higher education, this is discussed in section 2.2. It is shown that the last 15 years has witnessed a considerable increase in the proportion of women and 'mature' students in full time undergraduate education, as well as increases in the numbers of students from working class backgrounds and the ethnic minorities. It is also shown that widening participation remains a key objective of higher education.

Some implications of the expansion of higher education and its increased diversity are discussed in section 2.3. Also in section 2.3 various views of the purposes of monitoring and evaluating higher education are considered. It is argued that Teaching Quality Assessments (TQA) were introduced with the positive aim of shifting attention away from academics' research output (which had dominated thinking in many universities, largely through the importance and high profile of the Research Assessment Exercise (RAE)) towards encouraging excellence in teaching.

In section 2.4 the dynamics of the imposition of an evaluation system are discussed. It is argued that evaluation does not occur in a vacuum and the evaluation process is likely to change the behaviour that it aims to measure, as lecturers adapt their behaviour to what they consider to be in their best interest given the nature of the evaluation process to which they are being subjected.

Diversity in both the learning styles of students' and the teaching styles of lecturers' further colour the aims of those involved in higher education. Several different student learning styles are presented in section 2.5. The literature presented suggests that students will adopt different learning styles for different types of modules. It is argued that these differences may affect how a student evaluates the performance of the lecturer on the module. Differences in lecturers' teaching styles are presented in section 2.6. It is shown that not only do lecturers differ in style but also in their aims and objectives when they enter the lecture theatre. Research is presented that argues that lecturing styles change over the course of a lecturer's career. Finally, some conclusions are presented in section 2.7.

2.2 Higher Education in Historical Context

University education in the U.K. has a long and distinguished history. Oxford University, the oldest English-speaking university in the world has offered teaching in some form since 1096. Teaching at Oxford developed rapidly from 1167 as a result of the monarch, Henry II, banning English students from attending the University of Paris. The University of Cambridge records groups of scholars congregating in the town for the purpose of study as far back as 1209 and its first college Peterhouse was founded by the Bishop of Ely in 1284.

The development of higher education from its origins in the Middle Ages to its current form entering the 21st century has not followed a pattern of steady growth. For several centuries only these two universities existed in England. The colleges in Oxford and Cambridge were sovereign, self-governing communities of fellows (Cobban, 1988, 124). This allowed the universities to act autonomously both in their managerial practices and in the design of their curricula, they were completely independent of external jurisdiction (Maassen, 1997, 113).

Two hundred years ago there were only six universities in Britain - Oxford, Cambridge and four Scottish universities - Edinburgh, Glasgow, St. Andrews and Aberdeen. Between them these six universities enrolled fewer than 5,000 students (Scott, 1995). Scott comments:

> 'in no sense did [these six universities] form a system. Not until the mid-nineteenth century were there any public policy interventions to shape, or reform, what today would be regarded as higher education' (Scott, 1995, 11).

A century ago there were still only 14 universities in Britain and the number of students was only 20,000 (which represented just over 1% of the current total).

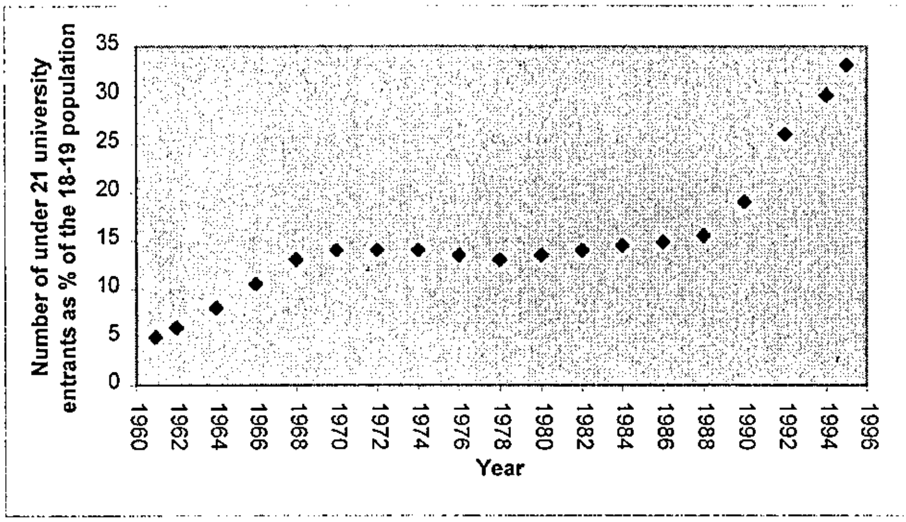
Varying reasons have been offered to explain the slow development of higher education in the U.K. Maassen (1997) suggests that a natural conservatism, primarily a product of the existence of only two universities for such a long period, ultimately held back the development of university education:

> 'both the strong emphasis on the classical curriculum and the control of the church over higher education prevented Oxford and Cambridge from modernising and adapting to the changing world around them' (Maassen, 1997, 113).

Trowler (1998) argues that a suspicion of the possible consequences of educating large numbers of working class continued to pervade discussions about the expansion of higher education throughout the nineteenth century (Trowler, 1998).

The numbers of students entering higher education remained low well into the second half of the 20th century. In 1938 2% of 18 year olds attended university. University education was male dominated, with only 0.5% of 18 year old females attending university (Blackburn and Jarman, 1993). As shown in Figure 2.1, by 1960 the proportion of 18-20 year olds entering university had only risen a further 2% to around 5%.

Figure 2.1 Proportion of under 21 year olds entering higher education

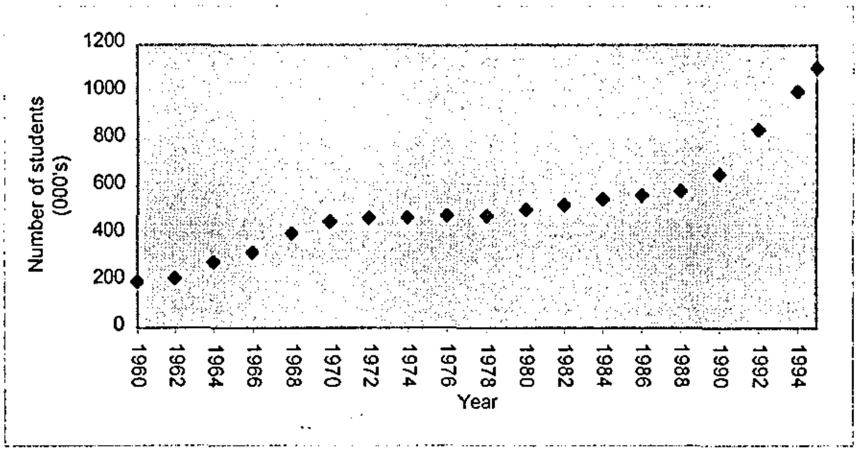


(Source: The Dearing Report, 1997).

Figure 2.1 clearly shows that there was a steep rise in the proportion of 18-20 year olds entering university in the second half of the 1960's. The catalyst for this change was the Robbins Report in 1963. The Robbins Report on higher education recommended the expansion of university provision in order to 'tap the talent' that was being excluded from university education. Particular emphasis was placed on the need to increase the diversity of the student body entering higher education, particularly by encouraging a greater proportion of students from working class families to attend university (Trowler, 1998). Figure 2.2 shows that the post Robbins expansion in the numbers of students entering higher education was swift.

An examination of Figure 2.1 highlights that there was a fall in the proportion of school leavers entering higher education in the late 1970's. The overall number of students in higher education however, did not fall to the same extent in this period.

Figure 2.2 Full time U.K. students in higher education in the U.K. (000's)



(Source: The Dearing Report, 1997).

Figure 2.2 shows the number of full time students in U.K. higher education from 1960 to 1995. Figure 2.2 clearly shows that there have been two periods of rapid growth over the last forty years, in the late 1960's and from 1988 through to 1995, with a sustained pause through the 1970's and early 1980's.

In 1970 there were 618,000 students in higher education. By 1995 this number had nearly trebled, reaching 1,720,000 (Trowler, 1998). Of these 1,100,000 were studying full time. This expansion has had a considerable effect on higher education.

The changes from the original nature of higher education to its present condition, both in terms of type of activities undertaken by academics and the management process they work within is summarised by Willmott (1995):

> 'universities have been transformed from up-market finishing schools catering exclusively for the progeny of the church and the aristocracy, where instruction centred around a craft or tutorial mode of delivery, into education factories based predominantly upon a lecture mode of instruction. Academics have been converted from self-employed and entirely self-regulating providers of educational services working in the colleges of ancient universities into salaried employees whose activities are increasingly regulated by external agencies' (Willmott, 1995, p.1005).

With the increase in student numbers has come an increase in the diversity of students entering higher education. Whereas the numbers in all categories of students have increased during the 1990's, postgraduate numbers have grown the fastest (Dearing Report, 1997, 3.7). Postgraduates now represent 14%[1] of U.K. higher education students compared to 6% in 1962/3.

Table 2.1 shows that there have been considerable changes in the mix of students at undergraduate level in U.K. universities since the mid 1980's.

Table 2.1 Undergraduate student intake by various categories as a proportion of the undergraduate population

	1986	1995
Women	42	52
Aged over 21	15	29
Socio-economic groups IIIm-V	23	28
A level entrants	86	74
Ethnic minorities	11	13

(Source: The Dearing Report, 1997).

Table 2.1 shows that there has been a considerable increase in the proportion of women on undergraduate degree programmes[2]. It is worth noting that the percentage of women undergraduates in the U.K. has doubled since the Robbins Report (Dearing, 1997, 7.6). There has also been a considerable increase in the proportion of 'mature' students entering undergraduate education[3]. The increase in the number of 'mature' entrants in higher education has reduced the predominance of 'A level' examinations as being the precursor to higher education. One of the direct consequences of the increase in 'mature' students is that university students have an increasing range of life experiences prior to entering higher education. There has also been an increase in the proportion of university entrants from a working class background[4] and from the ethnic minorities.

---

[1] Figures for 1995/6, excluding the Open University (quoted in Dearing, 1997).

[2] Although women remain under-represented in science and engineering disciplines (Dearing, 1997).

[3] There tends to be a greater proportion of 'mature students' in the 'new' universities (Dearing, 1997).

[4] Although men from socio-economic groups IV and V are particularly unlikely to participate (Dearing, 1997).

Dearing is keen to promote the continuation of widening participation in higher education. This is reflected in one of the Dearing recommendations:

> 'We recommend to the Government and the Funding Bodies that, when allocating funds for the expansion of higher education, they give priority to those institutions which demonstrate a commitment to widening participation' (Dearing, 1997, Recommendation 2).

2.3 Increased control over the product of higher education

2.3.1 The link between state funding and institutional autonomy

The change in the nature of universities from self-governing institutions to the more regulated, government-monitored institutions they are today is directly related to the change from universities being self-financing to being the beneficiaries of increasing government funding.

The separation of universities from the modern state was first challenged in 1850 when a Royal Commission was established to review the operation of Oxford and Cambridge colleges (Willmott, 1995). More extensive government support for U.K. higher education occurred shortly after the First World War, when the financial positions of the 'new' red-brick universities (e.g. Birmingham, Manchester and Bristol) had deteriorated to such an extent an institutionalised form of public funding was sought. The funding was secured through the establishment of the University Grants Committee (UGC).

The next occasion for a reconsideration of government expenditure on higher education resulted from the boom in the numbers of students entering university as a result of the Robbins Report (1963). These post Robbins changes 'stimulated a closer examination of funding arrangements' (Willmott, 1995, p.1005) that gradually induced a more direct desire by governments to intervene in higher education. The direct consequence of this was that the 'buffer role' played by the UGC was gradually eroded, declining in the 1970's and faced its ultimate demise in the mid-1980's

(Willmott, 1995, p.1006). The UGC, which had functioned for over half a century was disbanded for the precisely the same reason as it had been inaugurated - the need for state funding of higher education had grown considerably.

The biggest challenge to the funding of higher education has come over the last 10 years with an unprecedented increase in the numbers of students entering university (see Figure 2.2). This has lead to governments becoming increasingly interested in monitoring academic institutions. In the 1990's the quality of higher education became one of the most important policy issues of governments across Western Europe (Maassen, 1997). Maassen (1997) explains these developments as being the direct consequence of the large increase in student numbers, suggesting that the public funding of higher education had reached its 'upper-limits'.

The Dearing Committee reported that the expansion in student numbers entering higher education had been a desirable response to:

> 'the needs and aspirations of individuals, contributing to the health of society, and as an economic necessity...there should be maximum participation in initial higher education by young and mature students and in lifetime learning by adults, having regard to the needs of individuals, the nation and the future labour market' (Dearing, 1997, 3.10).

The expansion in student numbers and the consequent increase in the levels of government funding required to finance the expansion has led to a need for greater efficiency in universities in an attempt to reduce unit costs. Explicit in the terms of reference of the Dearing Committee was that the Dearing recommendations must be couched within the constraints of the Government's spending priorities and affordability. Consideration needed to be taken of how:

> *'value for money and cost-effectiveness should be obtained in the use of resources'* (Dearing, 1997, 15.2, emphasis in original).

Dearing recognised the efficiency savings that have been made in higher education:

> 'increasing workloads and outputs at a time of declining unit resources have been a feature across the system' (Dearing, 1997, 14.16).

These efficiency savings have been most keenly felt in terms of lower staff-student ratios, larger class sizes and less contact time for students. The magnitude of cost savings of reducing staff-student ratios are clear when one appreciates that staffing is the largest single cost to institutions accounting for about 58% of all expenditure (Dearing, 1997, 15.11).

Much of the increased cost of higher education is likely to be passed on to students. Tuition fees are already in place for undergraduates in higher education and Dearing believes that the costs to students in higher education are likely to rise:

> 'the level of investment needed in a learning society is such that we see a need for those who benefit from education and training after the age of 18 to bear a greater share of the costs' (Dearing, 1997, 1.21).

2.3.2 The move towards increased government monitoring of higher education

Increasing diversity in higher education is not restricted to the nature of students. The range of institutions carrying the title 'university' has also increased, largely as the result of the Further and Higher Education Act (1992) that abolished the binary divide between universities and polytechnics and also allowed various other institutions that had not previously been polytechnics to use the title university. Within this wide range of higher education institutions there are different missions and different emphases on the balance between teaching and research. Furthermore, different universities have different aims and styles in education provision and this further colours the diversity of U.K. higher education.

The increase in diversity within higher education has been linked to the 'marketisation' of higher education (Scott, 1999). This increasing diversity comes as a result of universities being expected to play a more direct part in aiding the needs of the economy, through the provision of more applied degree programmes that offer teaching of practical skills and knowledge relevant to specific occupations. Furthermore the last 50 years has seen the increase in corporate investment in universities. This investment is often accompanied by the expectation that the curriculum will mirror the needs of business.

The link between higher education and industry is promoted by the Government. The Dearing Committee had as one of its terms of reference:

> 'learning should be increasingly responsive to employment needs and include the development of general skills, widely valued in employment'.

Another key feature of the 1992 Further and Higher Education Act was that it enshrined the principle of formalised monitoring of teaching quality as a feature of higher education. As a consequence of this new-found focus on the monitoring of quality, it has become common in U.K. higher education to view students as customers (e.g. Hill, 1995; Thorne and Cuthbert, 1996) and following on directly from this lecturers as service providers. The use of terminology characteristic of the business world is fundamental and precipitates other changes that are both fundamental and far-reaching to the nature of professional life in universities.

One of the implications of higher education being shrouded in commercial discourse is its susceptibility to face the same pressures for increased public accountability and monitoring that has become pervasive in other state sector industries in the U.K. over the past two decades.

The growth of government audit bodies (e.g. the National Audit Office and the Audit Commission) since they were established in the early 1980's has been striking (Power, 1997). The introduction of the 'Citizens Charter' followed by an extension of this initial charter to charters that covered most aspects of national life were one of the more noticeable features of the early 1990's.

There is nothing unnatural about the concept of monitoring. Accountability and account giving are elements of what makes people rational (Douglas, 1992). It is however important when a monitoring system is introduced for those affected by it to believe that the system will be ultimately beneficial to the organisation. Given the obligatory nature of the increased government monitoring of higher education in the 1990's through imposed evaluation exercises, it becomes pertinent to consider what the intended effects of these evaluation exercises are.

Two rather contrasting possible motives for the introduction of systematised evaluation practices are posed by Neave (1988) who argues that the question is whether these measures:

> 'are intended to build consensus around those options that evaluation may reveal or whether the purpose of evaluation is to bend a recalcitrant academia to what the government deems to be "the new reality"' (Neave, 1988, 16).

Clearly these opposing possible motives differ in their potential benefit to those most immediately effected by the quality of provision in universities, namely the students. Students stand to benefit little if evaluation is merely a political game used by the incumbent government for political motives, particularly if this practice alienates the professionals that students most depend upon for their university education, namely their lecturers.

Trow (1993) has no doubt which of Neave's two potential motives best reflects Government intention:

> 'the British government is motivated more by a desire to control the academic community than by its quest for top quality higher education' (Trow, 1993).

Trow's argument is based on the premise that the increasing monitoring of higher education exists to discipline rather than to inform academics:

> 'a "hard" managerialism has displaced trust and elevates institutional and system management to a dominant position in higher education' (Trow, 1993).

Becher (1999) characterises quality issues as:

> 'a political phenomenon, designed to contain what might otherwise be represented as the unbridled power of professional groups. In this light, quality could indeed be labelled as a creature of political fashion - "the flavour of the day": and, like all fashions, expect to enjoy only a limited life' (Becher, 1999, 235).

Not all academics however concur with this view. From a set of interviews with university academics Henkel (1997) found for some academics:

> 'the assessment exercises enhanced academics' sense of being part of a collective enterprise with colleagues. It made them more aware of the programmes as a whole and of other people's teaching. This enabled a more coherent departmental approach. Sometimes it generated more sharing between colleagues of their approaches to teaching and what they were trying to achieve' (Henkel, 1997, 19).

The negative portrayal of teaching quality assessment (TQA) expressed by Becher (1999) and Trow (1993) places political control as the central motivation for the introduction of TQA. However this portrayal of TQA does not adequately reflect the significance of one of the fundamental reasons for the introduction of teaching quality assessments: namely, to address the balance of priority given within academic departments to research on the one hand and teaching on the other. Dearing points out that the Research Assessment Exercise (RAE) which predates TQA:

> 'has deflected attention away from learning and teaching towards research. An analysis of the impact of the 1992 RAE in higher education institutions in England suggests that it has devalued teaching' (Dearing, 1997, 8.9).

Rather than aiming to control the behaviour of lecturers in higher education, TQA was designed to have a positive effect on teaching and learning. The higher profile of teaching and learning was intended to liberate lecturers to engage in more innovative teaching practices, within an environment that fostered and encouraged high quality teaching. Dearing explicitly states:

> 'our national consultation suggested that, if the quality of students' educational experience is to be maintained or improved, innovative teaching strategies which promote students' learning - many of which are already in place - will have to become widespread' (Dearing, 1997, 8.10).

For Henkel (1997) the most important issue of teaching quality assessments is the question:

> 'whether public accountability can be reconciled with academic autonomy. Public accountability means that the choice of values, objectives, agenda and priorities for which institutions might be accountable is a matter of public interest and decision' (Henkel, 1997, 12).

This emphasises the importance of shared values and desired outcomes between academics and those funding higher education:

> 'the terms of the contract under which accountability operates will depend on the goals, the interests, the degree of mutual trust and shared values, and the relative power of the parties concerned' (Henkel, 1997, 12).

The need to ensure that the values held by those funding higher education are compatible with a viable approach to delivering higher education, becomes increasingly important given the fact that government support for higher education has led to a change of emphasis within universities:

> 'higher education in the U.K. has been seen, at least in the traditional university sector, as a unique and separate sector of the polity with strong external boundaries conserving its rights to self regulation. The reforms of the 1980's and the 1990's imply a different relationship between higher education and the state and constitute a threat to the institutional norms and structures

> that have sustained a particular conception of academic worlds' (Henkel, 1997, 12).

2.3.3 The increasing implementation of 'managerialism' in higher education

Power (1997) points out that the introduction of systematic evaluation of quality in higher education represents a shift from localised forms of self evaluation, which effectively allowed individual lecturers to be the manager of their own teaching, to increasingly standardised measures of quality.

The introduction of teaching quality assessments are portrayed by some researchers as encouraging the influence of managerialism in higher education:

> 'institutional managers, both academic and administrative, are key actors in teaching quality assessments. Not only do they intervene directly in them, they have also established structures linked to them and to academic audit which increase and centralize regulation of academic work' (Henkel, 1997, 21).

Another feature distinctive of teaching quality assessment is the language of quality itself. Focus on the concept of quality has come from government and from the world of management. Academics in contrast, have tended to talk in terms of academic standards (Reynolds, 1986).

Henkel argues that the implementation of standardised measures of assessing the quality of teaching, in which the teaching of each lecturer is evaluated on the same basis, has adapted the nature of academic work practices:

> 'internalist models of academic values, practice and the organization of knowledge have, at least until recently, remained remarkably influential. They depict academics as living in worlds of their own making, where the dominant influence on their organisation and on the formation of their values and

> assumptions is the nature of knowledge or the cognitive practices developed for the production of knowledge' (Henkel, 1997, 14-15).

Henkel continues:

> 'at the core of traditional academic organisation is a highly institutionalised individualism. Academics in all fields are encouraged to strive above all to demonstrate freedom, originality and intellectual leadership - often achieved by the forging of personal networks across institutional, national and, increasingly, disciplinary boundaries' (Henkel, 1997, 15).

The inference to be drawn from criticism of standardised measures of teaching effectiveness is that by changing the traditional practices of academics the introduction of standardised measures of measuring quality will damage the very product it is aiming to measure and ideally enhance. This reasoning does not reflect the predominant reason for the increasing need for active, good quality management in higher education. The increase in numbers of students in higher education and the increase in diversity of the students, as well as the economic necessity to encourage cost efficiency, makes increasingly active management a necessary and appropriate element of contemporary higher education. Dearing is particularly clear on the importance for effective management:

> 'the effectiveness of any organisation depends in the long term upon the effectiveness of its management and the arrangements for its governance. This applies particularly during periods of change and especially to higher education institutions in the years ahead' (Dearing, 1997, 15.3).

2.4 The unpredictable dynamics of evaluation

The introduction of a quality or performance measurement system in an organisation must be viewed as a form of intervention into the workings of the organisation. This is reflected in the aim of TQA to change teaching practices for the better. Therefore

TQA is not aiming to simply observe and then report what is happening in academic departments, but rather to adapt teaching practices to achieve improvement.

This reflects Power's conception of the effects of auditing:

> 'auditing works by actively creating the external organisational environment in which it operates...audit is never purely neutral in its operations; it will operationalise accountability relations in distinctive ways not all of which may be desired or intended' (Power, 1997, 13).

The implication for student feedback is that lecturers are likely to adapt their teaching to what they believe is in their interests given the nature of the evaluation system to which they are being subjected. To use Power's terminology:

> 'new motivational structures emerge as auditees develop strategies to cope with being audited' (Power, 1997, 13).

Power argues that the most deleterious manifestations of an evaluation system can take two forms 'decoupling' and 'colonization'. Each of these will be discussed in turn.

Decoupling occurs in situations in which the evaluation process becomes remote from the processes within the organisation that it is attempting to measure. Power (1997) uses the term 'compartmentalised' to describe the way in which the evaluation process and those conducting it become detached from those they evaluate. The ultimate consequence of this is that the evaluators and the evaluation system itself exists in:

> 'a world to itself self-referentially creating auditable images of performance' (Power, 1997, 95).

There is therefore a danger that the evaluation process will function merely to justify its own existence, rather than to be instrumental in actively aiding improvement in the performance it principally exists to monitor.

The monitoring of higher education is explicitly designed to minimise any 'decoupling' effects. The Dearing Report explicitly recognises that different institutions have different aims and objectives and that this diversity should be allowed to remain:

> 'institutions of higher education do not and will not fit into simple categories, they do and will emphasise different elements in their chosen purposes and activities, they are and will be diverse. Those which already have established a world reputation should be able to retain their distinctive characters. There should be no pressure on them to change their characters' (Dearing, 1997, 1.6).

Furthermore Dearing states:

> 'we encourage diversity of missions between institutions and their autonomy. This permeates many of our ...recommendations' (Dearing, 1997, 7.22).

Colonization in contrast has a rather different effect within an organisation. Colonization occurs when the evaluation process so completely enmeshes the organisation that the aims of the evaluation process become so prominent in the organisation that they become:

> 'a dominant reference point for organisational activity...the organisation becomes colonised by the audit process which disseminates and implants the values which underlie and support its information demands' (Power, 1997, 95).

In these circumstances the audit process is likely to be both unfair and undesirable as its side effects may actually undermine performance.

The lesson from Power's work is that it becomes imperative for managers implementing evaluation systems to ensure that 'the manner in which the audit process interacts with the audited domain' is congruent with the aims and objectives of the evaluation process. The evaluation process must have, and been seen to have, a clear rationale and legitimate underpinnings, to avoid the undesirable situation whereby the:

> 'audit cloaks its fundamental epistemological obscurity in a wide range of procedures and routines' (Power, 1997, 91).

Changing behaviour or attitudes of lecturers is not a bad thing in itself; indeed it could be a legitimate and openly stated objective of the student feedback system. What is important is the need to ensure *that any changes* to lecturers' behaviour and teaching patterns *imply an improvement* in their teaching and the quality of the modules of which they have responsibility. Furthermore the evaluation system needs to promote (and actively measure) good teaching practice and be responsive to the educational needs of students. This was highlighted in one of the recommendations of the Dearing Report:

> 'We recommend that, with immediate effect, all institutions of higher education give high priority to developing and implementing learning and teaching strategies which focus on the promotion of students' learning' (Dearing, 1997, Recommendation 8).

2.5 Appreciation of differences in students' learning styles

Meyer and Parsons (1989) found that there were two main orientations to students' leaning styles consistently evident in the different population samples they analysed. Meyer and Parsons termed these approaches the:

- *meaning orientation*, which is characterised by the student relating ideas, using evidence in creating arguments, showing comprehension and the student having an intrinsic motivation;
- *reproducing orientation*, which is characterised by 'syllabus-boundedness', the student being fearful of failure, having disorganised study methods, displaying negative attitudes to their studies and the student acting with improvidence.

One of the most significant attempts to categorise students learning styles was the 'Approaches to Study Inventory' (ASI) devised by Entwistle (Entwistle, 1981; Entwistle and Ramsden, 1983).

This instrument originated through interviews with students and was developed through a psychometric questionnaire. Factor analysis on the data derived from the inventory offered various combinations of factors that explained students learning. Entwistle (1994) simplified these factors into three categories, each category represents a distinct approach that a student can take in their learning. The three categories are:

a) deep approach;
b) surface approach;
c) strategic approach.

Matron and Saljo (1976) had previously introduced a pair of terms similar to two of the terms discussed by Entwistle. Matron and Saljo used the terms 'surface level' and 'deep level' referring to them as 'different levels of processing'. The main characteristics of each of approaches will be considered in turn.

Characteristics of a 'deep' approach to learning

The intention of a student taking a 'deep' approach to learning is to understand the material. The attitude of a student following a 'deep' approach towards the course will tend to be manifested in an active interest in the course content. The student's study style will generally involve them relating ideas to previous knowledge and experience and attempting to develop an understanding of the underlying principles and assumptions that underlie the material they have been presented with. The student will tend to adopt a critical approach to their reading, checking evidence and relating it to conclusions and examining logic and argument cautiously and critically.

Characteristics of a 'surface' approach to learning

The intention of a student taking a 'surface' approach to learning is to cope with the content of the module and the tasks set. The attitude of a student following a 'surface' approach towards the course will tend to be manifested in the student studying with little reflection and often making little effort to appreciate the module as a whole (or as part of their wider degree) and rather approach the material on a module as discrete and unrelated pieces of knowledge. The student's study style will generally involve them memorising facts and procedures as a matter of routine which may make it difficult for the student to make sense of new ideas. The student will tend to adopt an anxious approach to their reading, often feeling undue pressure and worry about their work.

Characteristics of a 'strategic' approach to learning

The intention of a student taking a 'strategic' approach to learning is to achieve the highest possible grade on tasks set before them. The attitude of a student following a 'strategic' approach towards the course will tend to be manifested in pragmatism. The student's study style will generally involve them being acutely aware of the assessment requirements and criteria and the student will often be a little Machiavellian in tending to gear their work to the perceived preferences of lecturers. The student will tend to adopt an consistent and organised approach to their reading, finding the most appropriate conditions and materials for studying and being effective in managing their time to achieve the highest grades they are capable of.

Research studies have found no systematic differences between men and women in terms of learning styles. Men and woman should:

> 'not be regarded as distinct populations with regard to their approaches to studying and learning' (Richardson, 1993, 10).

Kolb (1981) has suggested that students develop a preference for studying in a particular way. Kolb describes this as the students' 'preferred learning style'. This preferred style however is a tendency rather than an absolute. A student will adapt their learning style to a given situation if they consider that their preferred learning style is not suitable for a particular module. The learning style utilised in any particular module will reflect the individual student's abilities, the particular subject and its requirements and their previous learning history.

This assessment is supported by Nulty and Barrett:

> 'students select studying behaviours from their repertoire on the basis of the specific learning context the student finds themselves in' (Nulty and Barrett, 1996, 333).

This is likely to impact in the prevalence of different learning styles in both the students' year of study and academic discipline.

The prevalence of different learning styles is likely to change over the course of a degree programme. Research has detected changes in students' learning styles as they develop during their undergraduate studies. A 'surface' approach to learning is more common in first year undergraduates compared to students in later years of their degree programme. Many first year undergraduates put little effort into understanding what they are trying to learn, they rely on reproducing what is required for assessment (Entwistle, 1994).

This prevalence in a 'surface' approach to learning by first year undergraduates may be explained by experiences at school which often put too much emphasis on 'spoon-feeding' towards specific examination requirements (Wall, et al., 1991).

This places the focus on one of the main challenges facing students entering university. The nature of education style between what the students have experienced at school whilst studying for their 'A-levels' and the expectations on them when they enter the university environment are likely to differ. Students need to transform their approach to academic study from closely following the largely teacher-driven process of 'A-levels', in which the teacher is likely to lead them through all the pertinent material. Students need to transform their approach to the more student-driven process necessary for a student to achieve success in higher education, in which the onus is on the student to develop their own understanding based on their reading away from the confines of the lecture theatre. The speed with which students make this adjustment is likely to influence how comfortable they feel towards their academic work in their first undergraduate year.

Research has also shown that there are differences in students' learning style based on academic discipline. This is a reflection of the nature of different disciplines. Klob (1981) argues that different academic disciplines occupy 'different epistemological positions', and that different disciplines traditionally adopt a particular 'mode or discourse' in the educational process.

Therefore a particular learning style is likely to be more appropriate for some disciplines than others. For example students studying medicine are required to master the complexity of medical principles and medical practices in as detailed way as possible. This will require a degree of analytical skill, but predominantly it will require an ability to memorise and recall a large body of information. In contrast students in the social sciences are likely to be required to engage in critical thinking, questioning the assumptions and theories on which an argument is based.

Cullen (1994) has categorised academic disciplines into categories based on both the degree to which they are pure or applied and the degree to which they are 'hard' or 'soft'. These are presented in Table 2.2.

Table 2.2 A categorisation of academic disciplines

	Applied	**Pure**
Soft	• Marketing • Education • Political Science • Public Policy	• History • Philosophy • Sociology
Hard	• Applied Economics • Computing • Law	• Economics • Statistics • Mathematics

(Adapted from Cullen et al., 1994).

A 'hard' discipline will generally require the undergraduate student to master, in as greater depth as possible, a body of existing knowledge relevant to the discipline. In contrast a 'soft' discipline will generally require the undergraduate student to develop a critical understanding of the subject material and integrate their own understanding and experiences into their study. The potential implication of these distinctions for student feedback ratings is discussed in section 2.6.3.

There will be an interaction between the lecturer's teaching style and the student's learning style. Entwistle and Tait (1990) argue that students are likely to rate the lecturers higher that teach in a manner which they perceive as most assisting their chosen approach to learning. Gibbs, et al., (1996) offer an example of this:

> 'inexperienced students studying introductory courses may well rate highly those courses which are educationally unsound and which foster a passive surface approach while rating poorly those courses which demand an active deep approach' (Gibbs, Lucas and Simonite, 1996).

The student's perception of the aims of education is also likely to affect their approach to learning. Students who conceive of learning in a module as a quantitative increase in knowledge, or as memorising, are unlikely to be those who adopt a 'deep' approach to learning in that module (Marton and Saljo, 1984).

A student's approach to learning is likely to influence how they consider the quality of lecturing. In other words, what students believe to be good teaching depends on the sophistication of their conceptions of learning (Van Rossum and Taylor, 1987). Differences in lecturers' teaching styles will be discussed in section 2.6.

2.6 Appreciation of differences in lecturers' teaching styles

2.6.1 Alternative models of teaching

It is generally accepted that there are a variety of skills that constitute good teaching. Thomas (1993) for example, lists the following set of criteria describing excellence in teaching. An excellent teacher:

- Communicates subject matter effectively;
- Communicates an enthusiastic interest in the field of study;
- Stimulates thinking and develops understanding;
- Challenges the student's intellect;
- Uses assessment methods with reward understanding;
- Provides regular and useful feedback on students' course work;
- Take a personal interest in students and is willing to help.

Despite the range of dimensions that reflect excellent teaching, researchers of education categorise teaching into two distinct types. These two different types of teaching are often referred to as different paradigms. The use of the word paradigm emphasises that the difference between the two types of teaching are not simply alternative styles, but rather represent fundamentally different approaches to teaching that are underpinned by a different philosophy regarding the ideal purposes of teaching. Although each of the teaching traits outlined by Thomas (1993) are likely to be reflected in an excellent teacher, the emphasis on the importance of each of these traits will differ between lecturers employing different teaching styles. Different authors have used different names to describe these two paradigms; for example, 'instruction paradigm' and 'learning paradigm' (Barr and Tagg, 1995) or as 'knowledge transmission' and 'learning facilitation' (Kember and Gow, 1994).

Whilst these two alternative paradigms have been given different names they are each based on a distinction between teaching methods that fundamentally aim to:

- impart knowledge to students (i.e. instruction paradigm);
- act as facilitator to aid the student develop their own understanding (i.e. learning paradigm).

Kolitch and Dean (1999) contrast the 'transmission model of teaching' (Freire, 1970) and the 'engaged-critical model of teaching' (Weiler, 1988). In the transmission model, the purpose of teaching is seen as imparting information and the learning is the 'taking in' or absorbing this information (Hendry and King, 1994). In the transmission model the absorbing of large quantities of knowledge is more important than the construction and development of knowledge. Students who are only familiar with transmission learning can be characterised as perceiving the lecturer as a knowledgeable authority for whose expertise they are paying (Shor, 1996). In contrast, in the engaged-critical model students prior perspectives and experiences are taken as a starting point for their intellectual development. Knowledge is achieved through a reciprocal relationship between the lecturer and the student. Kolitch and Dean sum up the essence of the engaged-critical model:

> 'critical thinking, student expression and the development of respect among all participants are promoted in the engaged-critical classroom' (Kolitch and Dean, 1999, 31).

Dearing does not directly refer to the alternative teaching models by name, although it is clear that the type of teaching favoured by Dearing is closer to the engaged-critical model than the transmission model.

The Dearing Report quotes directly from Boyer, 1990:

> 'great teachers create a common ground of intellectual commitment. They stimulate active, not passive, learning and encourage students to be critical, creative thinkers, with capacity to go on learning after their college days' (Boyer, 1990, 24).

Dearing is clear that this vision:

> 'puts students at the centre of the learning and teaching process and places new challenges and demands upon teachers' (Dearing, 1997, 8.4).

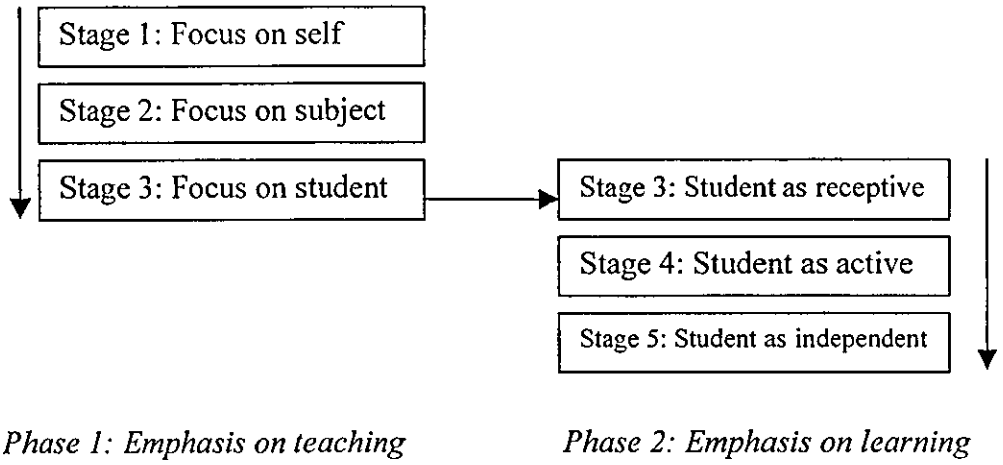
The Dearing vision is based on the need to equip students for the increasingly changeable world of employment they are likely to face and the corresponding need for 'lifelong' learning. Dearing stresses that emphasis needs to be placed on helping students developing their learning styles. Although Dearing does not specifically promote a 'deep' learning style, Dearing reports that:

> 'employers emphasised to us in their evidence the importance of high level analytical skills. The development of such skills characterises higher education, and should continue to be one of its primary purposes' (Dearing, 1997, 9.4).

2.6.2 The development of lecturers' teaching styles

Kugel (1993) argues that lecturers generally pass through a sequence of stages in their development as lecturers. This sequence is presented as Figure 2.3.

Figure 2.3 Developing styles of university lecturers



(Adapted from Kugel, 1993)

Kugel's model has two distinct phases. In the first phase the lecturer concentrates on his or her own performance, therefore the emphasis is on teaching. In the second phase the lecturers attention moves away from themselves towards the student. In this phase the lecturer concentrates effort in attempting to encourage the student to develop his or her own abilities.

The first phase develops through the lecturer focusing on their own performance as they attempt to overcome the initial anxieties and difficulties they face when they first begin to lecture students. Once this 'abject terror' (Kugel, 1993, 317) begins to diminish, lecturers turn their attention, Kugel argues, to a focus on the subject material of the discipline. Lecturers often need to develop their material and may 'worry about their mastery of the material they are presenting' (Kugel, 1993, 318). The next stage in the development of a typical lecturer is a change in focus from the lecturer's own performance to the needs of the student. To begin with lecturers tend to consider their students as being receptive to the information they convey, therefore the lecturer is working to a 'transmission model of teaching'. Lecturers, argues Kugel, remember that as an undergraduate their lecturers were active and they (as students) were often receptive to the material presented to them. Soon however, these same lecturers realise that in order to master their subjects they (as students) had to extend their knowledge beyond that presented in the lecture theatre and make associations that were not made explicit by the lecturer.

At this point lecturers are likely to consider:

> 'teaching more like coaching. The students' minds [being] less like pails to be filled than muscles to be strengthened by exercise' (Kugel, 1997, 322).

Over time a lecturer's perception of student learning changes. Lecturers begin to perceive of learning as an activity that occurs inside students' minds and the best function that they can perform is to:

> 'work actively as facilitators of their students' learning. As the professors' views of *how* to teach change, their views of *what* to teach may also change. They may decide that it is more important that students learn *how* to think than learn *what* to think' (Kugel, 1993, 323, emphasis in original).

Kugel qualifies his argument by stressing that not every lecturer will pass through each of the stages outlined in Figure 2.3, nor will lecturers necessarily proceed through the stages in a linear fashion as they develop their lecturing skills.

Kugel offers little in the way of empirical evidence to support his argument. The analysis, however, has not been challenged in the higher education or educational psychology literature. If Kugel's conceptualisation of the development of lecturing styles were to be generally accurate it would have important implications for attempts to measure the quality of teaching as Kugel's research implies that lecturers are not only going to differ in their styles *but also their aims* when they teach students.

2.6.3 Consequences and concerns for alternative teaching styles for student feedback

High quality teaching in higher education is generally identified with the promotion of effective learning opportunities for students (Broder and Dorfman, 1994). The consequence of this for student feedback is that to adequately evaluate the quality of teaching one must gauge the lecturer's performance in terms of the extent to which their teaching aided the students in their learning.

One of the problems that this raises for the measurement of effective teaching using a department wide student feedback form is that different lecturers are aiming to achieve different things in their lectures. Furthermore, undergraduate teaching in an environment such as the Business School at Loughborough University is likely to cover a wide range of disciplines that broadly represent all four discipline categories presented in Table 2.2. Furthermore the different degree programmes within a single department may have specific aims and objectives that differ from one another. This raises the question of whether lecturers in different subject areas in a department such as the Business School are likely to receive systematically different ratings as a result of the differing nature of the material they teach. The effect of subject area on student feedback ratings will be modelled in chapter 10. This diversity emphasises the need for departmental managers to be aware of the teaching attributes they most want to encourage and be clear that the student feedback system encourages and measures these.

There is also a danger that lecturers will not move beyond the first stage of Kugel's development process, simply continuing to:

> 'polish their teaching skills. Some [lecturers] develop [their teaching skills] so well that they join the ranks of the campus's most popular teachers. Their

> students enjoy their classes although a few may wonder how much they are really learning' (Kugel, 1993, 318).

Trow (1993) points to contrasting orientations to teaching in terms of delivery style, the degree of challenge posed to students, the extent of scholarly exploration and creative thinking which co-exist within any academic department and whose diversity defeats any single model of assessment.

This emphasises the need for care to be taken in the design of the form used to collect the student feedback. This is further highlighted by Kolitch and Dean (1999) who argue that there is concern that teachers' scores on rating forms are influenced by the goals and teaching strategies of individual instructors, which may not match the conceptions of teaching and learning portrayed on the typical evaluation instrument.

2.6.4 Contrast effects

In addition to the potential effects on student feedback of different teaching styles there is also the issue of the interaction between lecturers on modules in which the lecturing responsibilities are shared. In the psychology literature the analysis of the extent to which evaluations of performance are relative to either the performance of others or to previous performance of a particular individual are termed contrast effects. In particular, research that specifically attempts to explain the effects of others' performance on evaluations of a specific individual are referred to as context theories, these are distinct to consistency theories that explore the effects of an individual's previous performance on evaluation of his or her present performance (Buckley, Villanova and Benson, 1989).

Where this research has been applied to education evaluation it has taken the form of laboratory quasi-experimentation. In one such study Murphy et al. (1985) constructed an experiment in which two groups of undergraduate students were shown videotaped lectures and were asked to rate the performance of the lecturer. Each group was shown three lectures all delivered by the same lecturer. One group was shown two low quality lectures followed by an average quality lecture, the last of which they were asked to evaluate. The other group was shown the same average quality lecture, which they were asked to evaluate, but for this group of students the average quality lecture was preceded by two high quality lectures. The group of students that were presented with the two low quality lectures rated the average quality lecture significantly higher than the group who had watched the two high quality lectures.

This research is interesting as it suggests that students' evaluations of lecturers are a relative, rather than an absolute, evaluation of performance. This research however, does not provide a basis for the understanding of the extent to which the performance of one lecturer affects the ratings given to another lecturer with whom they share lecturing responsibilities on a particular module. Furthermore, being laboratory experiments, the research risks lacking generalisability to evaluations that take place outside the laboratory - in this case lecture theatres (i.e. the research potentially lacks ecological validity). The extent to which lecturers are rated relatively to the performance of those lecturers with whom they share lecturing responsibilities on a particular module is examined in chapter 9.

2.7 Conclusions

This chapter has discussed some of the changes in higher education over the last 40 years that have arisen primarily as a consequence of the significant expansion of university education in the U.K. With this expansion has come an increase in diversity both of the nature of students entering higher education and the nature of academic institutions offering university education. Alongside this expansion there has been an increase in the monitoring and evaluation of higher education. The obligatory monitoring of teaching quality in higher education was introduced in the 1992 Further and Higher Education Act.

The literature presented in this chapter implies that any mechanism to measure the quality of modules and the performance of lecturers in U.K. universities must be congruent with both:

- the changing and diversifying nature of higher education in the U.K.;

- the increasing development in the understanding of active learning environments.

In particular, any system to measure the quality of modules and the performance of lecturers needs to appreciate that there are different styles of students' learning and different styles of lecturers' teaching.

Research presented in section 2.5 outlined three principal styles of student learning namely, a 'deep' approach, a 'surface' approach and a 'strategic' approach. Students following different styles will consider the purpose of their studies in different ways and desire different types of teaching from lecturers. Similarly, the research presented in section 2.6 shows that lecturers vary in their styles of teaching. Two principal styles of lecturing were outlined, the 'instruction paradigm' whereby the principal aim of the lecturer is to impart knowledge to students and the 'learning paradigm' whereby the principal aim of the lecturer is to act as facilitator to aid the student develop their own understanding. Furthermore, research by Kugel (1993) suggests that both lecturers' teaching styles and their aims may change over the course of their career. The implications of these different learning and teaching styles (discussed in sections 2.5 and 2.6) for the validity of the underlying assumptions that are prevalent in the student feedback literature will be considered in the next chapter after a review of the general student feedback literature.

Chapter 3

A REVIEW OF THE STUDENT FEEDBACK LITERATURE

3.1 Introduction

This chapter explores the main themes in the literature on research into student feedback. The chapter also explores whether the common assumptions present in the student feedback literature are congruent with the contemporary nature of higher education given the extent to which higher education has increased in diversity in recent years. Section 3.2 presents an overview of the student feedback literature. It focuses on the key trends in the literature spanning the last thirty years. Reasons for the current lull in published research in the area are also suggested. Section 3.3 outlines four of the main purposes of collecting student feedback data. The extent to which student feedback data has been shown to serve each of these purposes is considered in turn. The section includes coverage of the usefulness of student feedback in improving a lecturer's teaching, the degree to which student feedback is used in personnel decisions, the degree of openness with the resulting data and the value of student feedback data in developing the curriculum.

Section 3.4 presents the literature on the extent to which student feedback data has been shown to be stable over time. Section 3.5 reviews the literature that has taken a construct validity approach to student feedback data. The reported relationships of two alternative measures of teaching effectiveness, student achievement and lecturers' self-evaluations, are also presented. Both of these measures are shown to significantly correlate with student feedback data. Section 3.6 presents the research findings on fourteen variables that have been proposed as potential biases on student feedback data. Section 3.7 outlines the ongoing debate in the student feedback literature on whether student feedback should measure multiple measures of teaching quality or whether a single (or small set of) question(s) can adequately represent the quality of a module and the performance of the lecturer(s) that teach it.

In section 3.8 assumptions prevalent in the student feedback literature are discussed. The appropriateness of these assumptions are questioned on the basis of the diversity of students entering higher education and the diversity of students' learning and lecturers' teaching styles discussed in chapter 2. Finally section 3.9 presents a summary of the student feedback literature.

3.2 Overview of the student feedback literature

As discussed in section 1.2.2, there was a rapid increase in the number of published articles on student feedback data in the mid 1970's. Most of these published articles were U.S. in origin and most of them were quantitative in nature. Many of the articles published in the mid 1970's attempted to quantify the usefulness of student feedback data. Some of these were able to utilise the data available from the larger databases. For example, Centra (1973, 1977) and Centra and Creech (1976) used the IDEA database to report on a range of issues including the effectiveness of student feedback in modifying instruction, the relationship between students, lecturers, module characteristics and student learning with the ratings received by lecturers.

These early attempts to examine the usefulness of student feedback data illustrate that studies of validity have been central to the student feedback literature for the last thirty years (see section 3.5). The second major strand in the literature that has been consistently present over the last thirty years has been a concentration on which variables could bias the results of the student feedback ratings received by lecturers (see section 3.6).

Marsh and Dunkin (1992) are highly critical of much of the student feedback literature that was published in the U.S. in the early / mid 1970's. Marsh and Dunkin claim that the main weakness of this literature was methodological. Many papers published during this period had experimental designs that were riddled with weaknesses and/or were analysed by inappropriate or inadequate statistical techniques (for example, Rodin and Rodin, 1972). Richard Schutz, who was the editor of the *American Educational Research Journal* during the latter part of the 1970's, commented that the prestigious educational research journals accepted far too many student feedback articles of questionable quality during this period (Marsh, 1987).

There was a reduction in the number of published articles on student feedback in the major educational research journals in the 1980's and these tend to be of higher quality compared to those published in the mid 1970's (Marsh, 1987). Since the late 1980's there has been a significant decline in the number of published articles on student feedback data. Few published studies significantly challenge the findings presented in Marsh's (1987) major literature review. Cashin, referring to his 1988 paper (IDEA Paper No. 20) writes:

> 'no major study published since then has substantively changed that paper's conclusions, but several studies or reviews of the literature provide modifications or further support for its conclusions' (Cashin, 1995).

One of the few exceptions to this is the continuing debate over the extent to which student feedback ratings can be considered multidimensional in contrast to a single question being able to adequately reflect the quality of the module or the performance of the lecturer (see section 3.7). The author of the Endeavor Instructional Rating Form, Peter Frey, suggests that one of the main reasons that explains why the number of studies on student feedback data have declined in recent years is that most quantitative based researchers in student feedback are convinced that the evidence supporting the validity of student feedback data is not in doubt and that the trends in the literature (as discussed in this chapter) are clear. Frey comments:

> 'most of the studies that purport to show problems with instructional ratings are based on poor instruments, poor data collection procedures or flawed design. Studies that have been designed well and conducted with solid methodology have uniformly demonstrated that students can discriminate good teaching from bad' (Frey, personal communication, 1997).

3.3 Purposes of collecting student feedback data

3.3.1 Main purposes of collecting student feedback data

According to Marsh (1987) there are four main purposes for collecting student feedback data:

1. To instruct individual lecturers about the effectiveness of their teaching from the viewpoint of their students, providing lecturers with knowledge that can ideally be used for the improvement of their teaching;
2. To act as a measure of a lecturer's effectiveness that can guide personnel decision making, playing a part in determining performance related pay, promotion and potential tenure;
3. To guide the students in the selection of modules;
4. As a measure of the quality of the module, to be used in module improvement and curriculum development.

Marsh also lists a fifth purpose of student feedback, to act as a guide to study the best teaching methods. This possible use of student feedback data has received little attention in the student feedback literature (Marsh, 1997). Evidence of the extent to which student feedback data has managed to satisfy these purposes is now presented for each of the four objectives in turn.

3.3.2 The effect of student feedback data in improving teaching

One of the key purposes of collecting student feedback data is to provide individual lecturers with an opportunity to find out how their teaching is perceived by the students on a range of teaching dimensions. The results for each lecturer are likely to indicate areas of teaching competence in which a lecturer is perceived as being relatively strong and other areas in which they are perceived as being weaker. For student feedback data to realise one of its major uses of guiding lecturers in improving their teaching, lecturers need to be receptive to the results of student feedback and use it as a tool for an active attempt at teaching improvement.

There have been many studies that have attempted to measure the degree to which student feedback influences teaching effectiveness. The usual design of these studies is to compare the results from student feedback taken in the middle of a module, to the results of feedback given by the same students using the same rating instrument at the end of the module. Cohen (1980) undertook a meta-analysis of all studies on this theme. He found that lecturers who received feedback mid-term significantly improved their ratings by around 1/3 of a standard deviation in the second round of feedback taken at the end of the module. The improvement in ratings tended to be considerably higher when student feedback was augmented with the active involvement of a teaching consultant who gave lecturers advice on strategies for improving their teaching based on the results of the student feedback (Overall and Marsh, 1979; Marsh and Roche, 1993; Hativa, 1995).

3.3.3 The role of student feedback data in personnel decision making

The use of student feedback data for personnel decisions, including promotion, pay rises and decisions on tenure (termed summative evaluation in the student feedback literature) is treated with a greater degree of suspicion by lecturers than when the same data is used simply to inform them of their students' judgements of their performance (formative evaluation). Most U.S. Business Schools now use student feedback data for personnel decision making, with 95% of Deans at 220 accredited American undergraduate schools making regular use of student feedback data as a source of information (Crumbley and Fliendner, 1995). One of the most popular commercially available U.S. rating instruments is IDEA which consists of two student feedback forms. The first form is a long (maximum 71 question) form, the data from which are intended to be used *primarily* for formative purposes. The second form is shorter (14 questions) and is intended to be used exclusively for summative purposes. Interestingly the students are told that the responses given on this form will be made available to university administrators and could influence their lecturer's chances of promotion, tenure or merit related pay. Research has shown that ratings tend to be higher if the stated purpose of the ratings (as printed on the rating forms) involves personnel decisions (Braskamp and Ory, 1994; Centra 1993; Feldman, 1979; Marsh and Dunkin, 1992). In other words, students tend to be more positive if they believe that individuals other than their lecturer will make use of the data.

Although there is a requirement for all lecturers seeking a new position, or tenure in U.S. universities to provide details of their student feedback results, the U.S. usage of student feedback results for personnel decisions may not be that distinct to the U.K. Frey comments:

> 'most universities [in the U.S.] do not want to emphasise good teaching as a criterion for promotion and salary raises because such a policy would reduce the income they receive from government sponsored research. Therefore, it is in the university's best interest to talk positively about the importance of good teaching for public relation purposes but to actually to make sure that it does not have much influence on hiring or promotion decisions' (Frey, personal communication, 1997).

3.3.4 Access to the results of student feedback data

The opportunity to use previous students' ratings of lecturers and modules to guide them in module selection is not available to all U.S. students. Some U.S. universities treat the results of student feedback as confidential, but many U.S. universities have a policy of complete openness about the results of student feedback, publishing the results in booklets available for purchase in university bookshops (Marsh, 1983). There is clear evidence that U.S. students are making extensive use of ratings in their module selections. UCLA students reported that the Professor/Module Evaluation Survey was the second most frequently read of the many student publications, following the daily campus newspaper (Marsh, 1983; 1987). Many U.S. universities now place summary results of student feedback ratings on the World Wide Web.

The introduction of student feedback into U.K. universities has not proceeded without controversy particularly in terms of the use to which the data should be put (Husbands, 1997). In the U.S. the principle of using the results of student feedback for summative purposes is well established and broadly accepted by lecturers. The application procedures for many U.S. universities will require an applicant for an academic post to provide a full account of their student feedback results (Marsh, 1987).

3.3.5 The effect of student feedback data in improving the curriculum

Marsh (1981, 1984a, 1987) provides evidence to suggest that student feedback data has limited value in directing module improvement and curriculum development. Marsh (1981) conducted a study aiming to separate out the impact that the lecturer has on the student feedback compared to that of the module, i.e. the degree to which ratings reflected factors specific to a module rather than a general reflection of the lecturer. Marsh arranged student feedback data from 1364 modules into sets of four such that each set contained ratings of the same lecturer teaching on the same module on two occasions, the same lecturer teaching on two different modules and the same module taught by two different lecturers.

The resulting table suggests that students are primarily rating the lecturer not the module itself.

Table 3.1 Correlations between different lecturer and module combinations for various dimensions of teaching performance

Factor	Same lecturer Same module	Same lecturer Different module	Different lecturer Same module	Different lecturer Different module
Learning / Value	.698	.563	.232	.069
Enthusiasm	.743	.613	.011	.028
Organisation/Clarity	.676	.540	- .023	- .063
Group Interaction	.699	.540	.291	.224
Individual rapport	.726	.542	.180	.146
Breadth of coverage	.727	.481	.117	.067
Examinations / Grading	.633	.512	.066	- .004
Assignments	.681	.428	.332	.112

Adapted from Marsh (1984a).

A consequence of this data is that results from student feedback provide us with little information about the module itself (independent of the lecturer). They therefore have limited use for measuring the quality of the module and cannot reliably be used as tools for curriculum development and module improvement (Marsh, 1987).

In one respect this finding is encouraging as it strengthens the validity of using a lecturer's student feedback data for personnel decisions. If the results of student feedback were subject to factors unique to a particular module, personnel officers would not be receiving information that reliably reflected a lecturer's performance. In other words, if student feedback ratings were generally module-specific student feedback data might not reflect a lecturer's ability and suitability for promotion, tenure and pay increases.

However this result is likely to be a function of the type of questions contained on the rating form. Marsh's analysis was based on results obtained using his own SEEQ student feedback form (Marsh, 1982). This instrument contains 31 questions referring to the lecturer and the module and a further ten questions that refer to the student. The 31 questions load onto nine factors, eight of which are reported in Table 2.1. The majority of these questions directly refer to the lecturer, which suggests that Marsh's conclusions could be more a function of his particular evaluation instrument rather than something universally true of student feedback data. On the Loughborough University student feedback form only a minority of the 18 questions overtly ask respondents to rate the lecturer.

3.4 Stability of ratings over time

Marsh points out that:

> 'some critics suggest that students cannot recognise effective teaching until after being called upon to apply course materials in further coursework after graduation. According to this argument, former students who evaluate courses with the added perspective of time will differ systematically from students who have just completed a course when evaluation teaching effectiveness' (Marsh, 1987, 276-277).

If this assessment is correct then one could hypothesise a statistically significant difference in the student feedback that lecturers receive from students whilst at university and at a later stage, for example after graduation (Marsh, 1984a). Overall and Marsh (1980) demonstrated that this is not the case. They conducted a longitudinal study in which they asked former students to retrospectively rate lecturers whom they had rated whilst a student. The minimum time differential between the two ratings was one year, (i.e. all respondents had been graduates for at least one year). Over 100 modules with paired ratings were available, the mean score for each lecturer was almost identical, the correlation between the two ratings of individual students (one at the time of tuition and the other some time latter) was staggeringly consistent ($r = 0.83$). Other studies (e.g. Firth, 1979) show very strong positive correlations between end of module ratings and subsequent ratings. These findings not only provide support for long term stability of student feedback ratings, but also provide support for their construct validity (Marsh, 1987).

3.5 Validity of student feedback data

3.5.1 The importance of validity research on student feedback data

Considerable research has been undertaken on the validity of student feedback ratings. This is interesting since student feedback ratings are by their nature difficult to validate since there is no single criterion of effective teaching (Marsh, 1987). One way of examining the validity of student feedback data is to take a construct validation approach (Cronbach, 1984). In this technique student feedback data would need to be significantly correlated with alternative indicators of effective teaching (Marsh, 1987). Two alternative criteria that have been proposed to measure effective teaching are student learning and instructor self-evaluations. Of these student learning is the most widely accepted criterion of teaching performance (Marsh, 1987). Both of these criteria have been compared to actual student feedback data and they are discussed in turn below.

3.5.2 The relationship between ratings and student learning

One of the main indicators of effective teaching is considered to be the 'degree to which an instructor facilitates student achievement' (McKeachie, 1979). For student feedback to be valid using this criterion it is necessary for there to be a clear positive relationship between the ratings students give their lecturer and the amount they learn as reflected in examination performance.

Cohen (1981) conducted a meta-analysis of 41 independent studies that reported data from 68 separate multi-section studies that analysed the relationship between the ratings students gave their lecturers and their subsequent performance in the module examination. The data presented by Cohen is based on the correlation between mean class achievement and mean class ratings for the lecturer. Cohen argues that for this type of analysis the lecturer must be used as the unit of analysis, not the individual student. If the individual student is used as the unit of analysis the data would describe whether students who learn more give higher ratings. Thus the result would be a description of the student not the lecturer and would fail to address the key issue of whether ratings differentiate to any degree among lecturers in terms of their contribution to student learning (Cohen, 1981).

Table 3.2 Correlations between various dimensions of student feedback ratings and student achievement

Rating dimension	Number of studies	Mean Correlation
Overall module	22	0.47
Overall Instructor	67	0.43
Skill	40	0.50
Rapport	28	0.31
Structure	27	0.47
Difficulty	24	-0.02
Interaction	14	0.22
Feedback	5	0.31
Evaluation	25	0.23
Student Progress	14	0.47

Reproduced from Cohen (1981).

The evidence in Table 3.2 indicates that there is a clear relationship between the two indexes of teaching effectiveness (i.e. ratings of teaching by students and student achievement). Cohen gives a detailed breakdown on the results for each category. Data allowing the correlation between ratings of the module (overall - the 'global' question) and achievement was available in 67 of the 68 studies that formed the meta-analysis. In 59 of these ratings of the module and student achievement correlated positively. In 31 of these the coefficient was statistically significant. Similarly, of the studies that asked the students specifically to rate the lecturer's skill the relationship between the class average of this and mean achievement was positively correlated in 37 out of 40 studies and statistically significant in 20 of these. These results provide clear evidence in support of the validity of student feedback.

3.5.3 The relationship between student evaluations and lecturer self-evaluation

A second measure of the construct validity of student feedback data is the relationship between students' ratings of lecturers and the same lecturer's evaluation of themselves. The correlation between the two sets of ratings tends to be statistically significant. Feldman (1989) cited 19 studies from the literature that dealt with this association and reported that the average correlation was 0.29. In a study of 329 modules using the standard SEEQ rating instrument (Marsh 1982), lecturers were asked to rate their own teaching on the same student feedback form as used by their students to rate their teaching. The results of each pair of rating forms were similar, the student-lecturer agreement on *each* of the nine dimensions were statistically significant, with the correlations ranging between 0.17 and 0.69 with the median correlation 0.45 (Marsh, Overall and Kesler, 1979). These results provide an independent second measure that reinforces the research on the association between ratings and achievement in support of the validity of student feedback data.

3.6 The search for biases in student feedback ratings

For student feedback ratings to have construct validity the ratings need to be related to variables that reflect effective teaching, but be relatively uncorrelated with variables that do not. Such variables can be considered biases (Marsh, 1987). There is no agreed definition of what actually constitutes a bias in the ratings. One approach is to define any impact on ratings of a variable not under the lecturer's control as a bias. This definition provides the unsatisfactory anomaly that any detectable grading leniency effect would not be considered a bias to the ratings as it would obviously be under the lecturer's control (Marsh, 1984a). Marsh advocates an alternative definition of bias that classifies a variable as a bias if it is not related to teaching effectiveness. Under this definition the effects of class size and student's interest in the module are not biases as they are related to teaching effectiveness.

The literature presented in chapter 2 provides plausible explanations for some potential biases on student feedback ratings. Kugel's (1993) research that argues that lecturers' teaching aims and objectives change over time could offer a rational explanation for student feedback ratings being influenced by lecturers' age, rank and experience. Similarly, the evidence from the educational psychology literature into different learning styles utilised by students (see section 2.5) showed that students' learning styles were likely to change over the course of their undergraduate studies and differ between different academic disciplines.

There have been a vast number of studies that have attempted to measure the effects of potential biases on student feedback ratings. Marsh has described the continuous stream of articles attempting to throw doubt on the validity of student feedback data as a 'witch hunt' (Marsh, 1987). The main findings of the literature are outlined below. There are very few variables that have consistently been shown to bias ratings, the impact of those that do bias ratings is usually small.

- **Academic rank of lecturer**

Of 33 studies analysed by Feldman (1983) the majority (21) reported no significant correlation between a lecturer's rank and their ratings. In ten of the studies, the lecturer's rank was significantly positively correlated with ratings, i.e. the higher the rank of the lecturer the higher their ratings. The one remaining study showed a significant negative correlation. Marsh (1987) reports that questions relating to lecturer knowledge and intellectual expressiveness tend to be positively related to academic rank whereas questions referring to encouragement of discussion, openness and concern for students tend to be negatively related to a lecturer's academic rank. In the SEEQ research once teaching assistants are excluded from the analysis the relationship between rank and ratings becomes very small; there is a clear trend for teaching assistants to receive significantly lower ratings than regular faculty.

- **Experience of lecturer**

Feldman (1983) analysed sixteen studies that showed a mixed pattern of results. Eight studies showed no significant correlation between the experience of the lecturer and the ratings they received. Of the seven studies where a statistically significant correlation was found, the majority (5) found significant evidence of a negative relationship between the experience of the lecturer and their student feedback ratings, i.e. the greater the experience of the lecturer the lower their ratings. The remaining study, Centra and Creech (1976) revealed a non-linear pattern. Specifically, from a sample of 10000 modules, they report an inverted J curve. Student feedback ratings for novice lecturers were low; the ratings climbed with experience reaching a plateau after about twelve years and then started to decline. Braskamp et al. (1985) also reported a non-linear pattern in which a lecturer's ratings increase over the first ten years of teaching and then decline thereafter.

- **Age of the lecturer**

Feldman (1983) analysed twelve studies and found that half of them reported a significant *inverse* relationship with ratings, i.e. the older the lecturer the lower the ratings. In the other six studies there was no significant correlation between the age of the lecturer and the ratings they received.

- **Sex of lecturer**

There is no significant relationship between the sex of the lecturer and the ratings they receive. Feldman (1993) analysed 39 studies that explored the link between the sex of the student and their judgement of teaching quality. Of the 28 studies that gave correlation coefficients, the average correlation between the lecturer's sex and their ratings was 0.02. On average sex differences explain 4/100 of 1% of the variance in overall ratings.

- **Research productivity of the lecturer**

There is no evidence of a significant relationship between the research productivity of the lecturer and the ratings that they receive (Centra, 1993). Feldman (1987) found that the average correlation between research productivity and student feedback ratings was only 0.12, indicating that performance on one element of the lecturing profession (research) cannot be used to predict performance on another (teaching).

- **Students' previous academic performance**

There is no evidence of a significant relationship between the student's level of performance in previous examinations on their degree programme and a lecturer's student feedback ratings (Feldman, 1976).

- **Class size**

Of about 30 studies analysed by Feldman (1978), one third found no significant relationship between class size and the student feedback ratings received by the lecturer. The remaining two thirds showed a significant negative correlation with ratings declining as the class size increased. However, although these correlations are statistically significant, they tend to be weak, typically ranging from -. 1 to -.3. In other words, variations in class size explain between 1 - 8 % of the variance in class ratings (Feldman, 1984). A more interesting possible relationship between class size and ratings is offered by Marsh, Overall and Kesler (1979). Through a detailed analysis of student feedback data they show that the weak negative correlations may 'mask' a non-linear relationship between class size and the ratings received by the lecturer. This takes a U shape, with relatively small and relatively large class sizes giving higher ratings and class sizes in the range 35-100 giving lower ratings (Centra and Creech, 1976). Marsh found that class size has the largest impact on questions relating to 'group interaction' and 'individual rapport'.

- **Module level**

There is evidence that modules taken later in a degree programme receive higher ratings compared to modules taken earlier in a degree programme. Similarly graduate level modules tend to attract ratings that are higher (on average) than undergraduate modules (Aleamoni, 1981; Braskamp and Ory, 1994; Feldman, 1978).

- **Level of the student**

There is some evidence that the level of the student (i.e. year of study) has no effect on ratings (McKeachie, 1979). This may seem paradoxical given that there is evidence of an effect due to module level. However, this is not contradictory if one remembers that the enrolment of some undergraduate university modules is drawn from students at varying stages of their degree programme. This is interesting as it suggests that ratings are a product of the lecturer and the module rather than the student.

- **Workload and difficulty of the module**

There is evidence illustrating a correlation between the variables of workload and difficulty with ratings. Rather counter intuitively the relationship is positive. The harder and more demanding the module the higher the ratings associated with it (Marsh and Dunkin, 1992; Centra, 1993).

- **Timing of the data gathering**

The research evidence illustrates that a lecturer's student feedback ratings tend to be very similar whether collected at the end of the module immediately prior to the examination or in the middle of the module (Feldman, 1979; Marsh and Overall, 1980).

- **Academic discipline**

There are considerable differences in the patterns of ratings associated with different types of discipline. Humanities and arts type modules tend to receive higher ratings than social science subjects, which in turn receive (on average) higher ratings than mathematical and science modules (Centra and Creech, 1976; Feldman, 1978; Marsh and Dunkin, 1992).

- **Whether student feedback forms are signed**

There is evidence that non-anonymous ratings tend to be higher (Braskamp and Ory, 1994; Centra 1993; Feldman, 1979; Marsh and Dunkin, 1992). It may be that students are concerned about possible undesirable implications if they voice criticism of their lecturer. For this reason it is recommended that students are instructed not to sign their evaluation forms (Cashin, 1995). This is not to say that the higher ratings are necessarily wrong but rather to ensure consistency.

- **Whether lecturer is present while forms are completed**

There is some evidence that if the lecturer is present in the room at the time that students are filling out the student feedback form the resulting evaluations are likely to be higher (Braskamp and Ory, 1994; Centra 1993; Feldman, 1979; Marsh and Dunkin, 1992). Cashin (1995) recommends that lecturers be instructed to leave the room whilst the student feedback forms are being completed and collected. Alternatively all lecturers could be instructed to be present. As with the issue of anonymity it is important to encourage consistency by standardising procedures (Marsh, 1987).

Although extensive, the literature analysing the effects of variables that could bias student feedback ratings is not exhaustive. For example, contrast effects between lecturers (see section 2.6.4), whereby a lecturer's student feedback ratings are influenced by the ratings of other lecturer(s) with whom they share lecturing responsibilities on a particular module could lead to another source of bias.

3.7 The multidimensionality of student feedback ratings

One of the current debates in the student feedback literature is the extent to which a single question can adequately summarise the performance of the lecturer or the quality of the module. There is general agreement in the student feedback literature that the skills of teaching are multifaceted (Abrami and d'Apollonia, 1991; Abrami, d'Apollonia and Rosenfield, 1996; Cashin and Downey, 1992; Cashin, Downey and Sixbury, 1994; Marsh, 1987, 1991a, 1994, 1995). A lecturer may not perform consistently well across a range of teaching skills; for example a lecturer may be well prepared but lack enthusiasm. The research into different teaching styles discussed in section 2.6 shows that lecturers enter the lecture theatre with different aims and objectives. For example, a lecturer following an 'instruction paradigm' is likely to place emphasis on the organisation of the module and the preparation of the module materials. In contrast, a lecturer following a 'learning paradigm' is likely to place emphasis on challenging students and stimulating them to pursue their own independent learning of the material.

In the literature a key distinction is made between appropriate measures of teaching effectiveness for formative evaluation (i.e. data made available to lecturers to guide them in the potential improvement in their teaching) and summative evaluation (i.e. data made available to departmental managers and personnel officers to be used for managerial purposes). There is general agreement that for formative evaluation data on several dimensions of teaching effectiveness need to be collected and reported. However there is disagreement over whether data on the different dimensions of teaching effectiveness need to be collected and reported for summative evaluation.

For the purposes of this research it is not necessary to draw such a distinction between the two potential uses of the data, but it is important to gauge whether it is possible to meaningfully report a lecturer's performance using a single general question of teaching effectiveness. This is because in chapters 8 and 9 the most general of the lecturer-specific questions (Q17) is used as the dependent variable. The decision to use Q17 as the dependent variable was taken after Q17 was found to highly positively correlate with the other five lecturer-specific questions. Furthermore, the regression equation modelling Q17 against the other lecturer-specific questions was shown to have a very high R^2 value. This analysis is presented in section 8.5.

It is possible to characterise three perspectives on the appropriateness of using data from global questions on student feedback forms. One perspective is that student feedback forms should consist exclusively of questions that measure specific elements of teaching effectiveness, with global questions excluded from the form (Frey, Leonard and Beatty, 1975; Frey, 1978). Another perspective is that only global questions are necessary as they can adequately reflect the students' views as to the quality of teaching (Abrami and d'Apollonia, 1991; Abrami, d'Apollonia and Rosenfield, 1996; Cashin and Downey, 1992; Cashin, Downey and Sixbury, 1994). This perspective is based on two principles. Firstly, the lack of even broad agreement in the literature as to the dimensions of effective teaching and the presence of different teaching styles in which lecturers differ in their objectives implies that no list of specific teaching traits can be said to be better than any other at measuring teaching effectiveness. Secondly, the three global questions on the IDEA student feedback form ('overall instructor effectiveness', 'overall course worth' and 'overall amount learned') have been shown to account for between 54% and 69% of the variance in a weighted composite of ten specific dimensions of teaching effectiveness (Cashin and Downey, 1992). Marsh (1994, 1995) provides a third perspective, arguing that a weighted average of both specific dimensions and global items should be used in preference to global questions alone. However there has been little advance in the literature as to an accepted and consistent method of weighting student feedback data. The analysis presented in this research takes a dual approach. Whereas the most general of the lecturer-specific questions is used as the dependent variable in chapters 8 and 9, in chapter 10 each of the lecturer-specific questions are separately modelled against student feedback ratings allowing the effects on student feedback ratings of different aspects of lecturing to be compared.

3.8 Assumptions in the student feedback literature

The research on student feedback data mentioned in this chapter so far portrays student feedback data in a positive light. Empirical studies have generally shown student feedback data to be valid and reliable. Research has also shown that few variables consistently bias ratings and the effects on student feedback data of those that do are generally small.

Much of the student feedback research however, is based on a key pervasive assumption that may not be fully justified given the nature of contemporary higher education as discussed in chapter 2. This is the assumption that it is appropriate to treat an individual class as a single coherent homogeneous entity. This assumption is often manifested in the common practice of averaging the ratings given by the students in the class for a particular feedback question into a single mean value which is perceived to represent the view of the entire class on that particular question.

The premise of treating the class as a single coherent entity is questionable given the increasing degree of diversity of students that one should now expect to find in any university lecture theatre. Given the diversity of students entering higher education it is important to explore the extent to which the practice of treating a class as a single entity (via the use of class average ratings) adequately reflects the reality inside lecture theatres.

The underlying assumption of treating the class as a single entity is reflected in the following explicit assumptions in the student feedback literature, namely that there is:

a) a 'true' score for each lecturer on a module, with variation of responses merely being random variation (Crichton and Doyle, 1975);

b) a direct relationship between teaching and learning (Timpson and Andrew, 1997).

The assumption that there is a true score for each lecturer on a module implies that there is a single objective rating for a lecturer for each of the questions on the student feedback form. This based on the premise that:

> 'there exists a true value on a given trait for the ratee which every rater, if he [or she] is not biased or unmotivated or careless or unobservant, will give the ratee' (Crichton and Doyle, 1975, quoted in Feldman, 1998).

Crichton and Doyle continue:

> 'this ignores the possibility that there may be a different 'true' value for each student, for example, because the instructor satisfied his [or her] needs or desires with respect to the function named to a differing degree'.

The assumption that there exists a 'true' score for a lecturer, is used to justify placing the focus in student feedback research on the class average rating, on the basis that the variations in responses will simply be randomly distributed around the mean. The most questionable aspect of this practice is that it precludes the possibility that the class may not all be evaluating the module and the lecturer on the same basis. If there are groups of students who are utilising different learning styles in the class (as discussed in section 2.5) the lecturer will be faced with students who desire different forms of teaching. The extent to which the lecturer's teaching style is congruent with the learning style utilised by each group of students is likely to be reflected in the evaluations that the students make of the lecturer. Therefore, the rating a lecturer receives from a student may reflect the extent to which the lecturer addresses the educational aims held by that student. In modules in which there are students utilising different learning styles the educational expectations of the students are not only likely to differ but are also potentially contradictory.

This therefore raises the issue of on what basis might the class not form an homogeneous entity. The likely answer to this is the same reason why the assumption that there is a direct relationship between teaching and learning discussed by Timpson and Andrew (1997) is almost certainly flawed.

The assumption that there is a direct relationship between teaching and learning, may seem commonsensical. A lecturer teaches and providing the lecturer is reasonably good and the student reasonably attentive and thoughtful, the student learns. Indeed the evidence from the student feedback literature suggests that there is a clear relationship between lecturers' student feedback ratings and student achievement in terms of examination performance on a particular module. For example, Cohen (1981) found that the overall correlation between these two indexes of teaching quality was 0.43 in a meta-analysis of 68 separate student feedback studies (see section 3.5.2). Therefore the quality of the module and the performance of the lecturer should be reflected in the ratings given by students. The students are in a reasonable position to judge whether their understanding has developed and if it has the lecturer warrants high ratings. Similarly, if the students do not feel that the lectures have aided their understanding the lecturer warrants lower ratings.

The problem with this assumption is that it takes a simplistic view of both teaching and learning and which once again raises questions over the appropriateness of treating the class as a single entity. Research from the educational psychological literature outlined in chapter 2 showed that there is significant diversity in the learning styles utilised by students and the teaching styles employed by lecturers. These differences call into question whether it is valid to portray a direct relationship between teaching and learning, they rather suggest that the relationship between teaching and learning may be more complex. The varying approaches students take to their learning (outlined in section 2.5) showed that students following different learning styles differ in their aims and the expectations they have of lecturers. Similarly, lecturers employing different teaching styles (outlined in section 2.6) are likely to differ in their aims when they enter a lecture theatre and in the expectations they have of students. These differences between a lecturer's teaching style and a student's learning style may play a significant part in how effective a student considers a lecturer to be.

The interaction between the lecturer's style and a student's style will also have important implications for lecturers' student feedback ratings. For example, a lecturer who seeks to challenge students to engage in a considerable amount of independent thought is unlikely to receive high ratings from a student who is following a 'surface' learning style, even if the lecturer concerned conducts interesting and insightful lectures. Similarly, a lecturer who fails to intellectually stimulate a student who is following a 'deep' learning style (for example, by simply presenting factual information) is also unlikely to receive high ratings from that particular student.

The potential flaw with the second assumption, that diversity in learning styles precludes there being a direct relationship between teaching and learning, explains why the assumption that there is a true score is also flawed. In other words the second assumption provides a reason to doubt that there is a single 'true' score for a lecturer for a particular module. Differences in learning styles imply that there could legitimately be more than one 'true' score for a lecturer on a module, i.e. a 'true' score for the lecturer from the students following a 'deep' learning style, a different 'true' score from those student following a 'surface' learning style and a different 'true' score for students following a 'strategic' learning style. The diversity in learning styles therefore becomes the principal reason to challenge the underlying assumption that links both the 'true' score and direct relationship between teaching and learning assumptions. As no single class is likely to consist of students who are all employing the same learning style, the appropriateness of treating a class as a coherent homogeneous entity for the purpose of student feedback is called into question.

3.9 Summary

This chapter has introduced the extensive literature on student feedback data. Student feedback data has been shown to be both valid, using multiple measures of teaching effectiveness and reliable, through being stable over time. One of the major themes in the literature is studies that attempt to measure the extent to which various variables influence student feedback data. Very few variables have been shown to consistently bias the ratings lecturers' receive. For the small number of variables that have been shown to influence the ratings lecturers' receive, the size of the impact has been generally shown to be small.

Much of this research has, however, proceeded on the basis of the questionable assumption that a class can be treated as a coherent homogeneous entity. Two explicit assumptions that follow as a consequence of this underlying assumption can be found in the literature, namely: there being a direct relationship between teaching and learning (Timpson and Andrew, 1997) and of there being a 'true' score for each lecturer on a module, with variation of responses merely reflecting random variation (Crichton and Doyle, 1975). The increasing diversity in U.K. higher education (discussed in chapter 2) and the research into different learning styles and different teaching styles published in the educational psychology literature (also discussed in chapter 2) suggests that it may not be appropriate to treat the class as a single homogeneous entity. The implications of the presence of discrete groups of students in Business School modules for both the appropriate statistical analysis of student feedback data and the presentation of student feedback results will be discussed in chapter 6. The methods commonly used to analyse student feedback data are based on the assumption that there is a single 'true' score for a lecturer on a module, which precludes the possibility that there could be more than one 'true' score for a lecturer resulting from the presence of discrete groups of students in the class who are employing different learning styles which leads them to differ in their appreciation of a lecturer's teaching style.

The organisation of teaching commitments in the Business School at Loughborough University provides an opportunity to use the Business School student feedback data to extend the analysis of potentially biasing variables, by introducing additional variables not previously featured in the published research. As the teaching load on many modules is split between different lecturers the data available for this thesis provides an opportunity to analyse the impact of variables that derive specifically from these shared teaching loads. The research presented in chapter 8 considers the influence on student feedback ratings of the proportion of lecturing hours undertaken by a lecturer on a module and the research presented in chapter 9 considers the effects for a lecturer's ratings of the ratings of the other lecturer(s) teaching on the module.

Chapter 4

DATA GATHERING

4.1 Introduction

This chapter describes the process by which the student feedback data is collected and processed in the Loughborough University Business School. Section 4.2 presents the student feedback form used in the Business School. Each question is considered in turn, with particular emphasis given to the wording of questions and how they may be interpreted (or misinterpreted) by the students. The questions that refer to teaching performance are compared to those in three well-known student feedback forms. The expectations and implications of the wording of questions for subsequent analysis are discussed. Section 4.3 outlines the process by which the student feedback data is collected, tracking the process step by step from the distribution of the forms in the lecture theatre through to the data being available to the Business School. Section 4.4 describes how the forms are read using an Optical Mark Reader (OMR). Various reasons for the OMR being unable to read some responses are discussed. Results from a statistical analysis of the extent to which the OMR fails to read apparently clear responses are also presented. Differences in the frequency of missing values between students at various levels of study are highlighted.

4.2 The structure of the student feedback form

4.2.1 General outline of the Business School student feedback form

The form used to collect student feedback data in the Business School has gradually developed over the last decade. Before 1990 various lecturers in the Business School designed their own feedback forms to collect the students evaluations of their modules and teaching. The collection of student feedback data was voluntary. In 1990 the student feedback form in the Business School was formalised with each lecturer required, for the first time, to collect student feedback data on each module in which they were involved. Each lecturer was evaluated on the same student feedback form.

Between 1990 and 1995 there was a gradual increase in interest within Loughborough University as a whole as to the importance of the collection of student feedback data (see section 1.2.4).

In 1995 Loughborough University introduced a university wide student feedback form that was then imposed on each department. The extent of the experience of designing the forms and collecting the data from them within the Business School, allowed the Business School to offer its model of student feedback to be used as the basis for the student feedback form controlled centrally by the University. In October 1999 the university introduced a new form that has very different questions to the version prepared in 1995.

This research is based on student feedback data collected in the period between October 1996 and June 1998. The student feedback form used in the Business School during this period is presented in Appendix 4.1. The form consisted of eighteen questions. These eighteen questions asked for responses based on a five point Likert scale ranging from strongly disagree to strongly agree. On the reverse side of the form students were encouraged to make written comments that elaborated on their responses to the eighteen set questions. Seven of these set questions were specified centrally by the University; the individual departments composed the remaining eleven. The student feedback form was split into two sections. The first of these consisted of up to twelve questions that were module-specific. The second section consisted of up to six questions that were lecturer-specific. The form had provision for up to three lecturers to be evaluated on each module.

4.2.2 The module-specific questions

Table 4.1 presents the first twelve questions on the student feedback form. These twelve questions varied in their purpose and their suitability for providing worthwhile data. Of these twelve questions the first six were specified centrally by the University, the Business School set the final six. Each of these were specific to the module. Each of the questions will now be examined in turn.

Table 4.1 Module-specific questions on the student feedback form.

	SD	D	N	A	SA
1. I was made aware of the module's aim and objectives	☐	☐	☐	☐	☐
2. The teaching methods for this module helped me understand the module	☐	☐	☐	☐	☐
3. The academic content of the module was stimulating	☐	☐	☐	☐	☐
4. I understood the assessment requirements of the module	☐	☐	☐	☐	☐
5. The teaching rooms for this module were fit for their purpose	☐	☐	☐	☐	☐
6. I found the library support for the module to be satisfactory	☐	☐	☐	☐	☐
7. The subject material in this module is challenging	☐	☐	☐	☐	☐
8. The subject material in this module was presented at about the right pace	☐	☐	☐	☐	☐
9. The subject material in this module was related to business situations	☐	☐	☐	☐	☐
10. The teaching on this module was well organised	☐	☐	☐	☐	☐
11. The coursework supported the module objectives	☐	☐	☐	☐	☐
12. The tutorials linked with this module were useful.	☐	☐	☐	☐	☐

1. I was made aware of the module's aim and objectives

As phrased this question should not provide much in the way of useable data. If the students answer the question literally then one would expect little variation in the ratings for an individual module. This is because it is Business School policy to provide students with a written statement of the module's aims and objectives. Normally this will be included with the reading list given to the students in the first lecture for the module. Therefore all students should answer either agree or strongly agree to this question. Given the wording of the question there is no scope for the student to reflect in their answer whether they *liked* the aims or objectives of the module; the question is restricted to a statement of whether *they were made aware*. If the responses on a module vary this may reflect students responding in terms of their feelings towards the module as a whole rather than the specific question asked.

2. The teaching methods for this module helped me understand the module

This is a standard type of question that one would expect to find on most student feedback forms. Students with different backgrounds may find different methods helpful. Therefore a greater degree of variation is expected within a class for Q2 compared to Q1.

3. The academic content of the module was stimulating

This question is similar to question 7, which reads:

7. The subject material in this module is challenging

The questions differ in their focus. Q3 is attempting to gauge the student's interest in the module whereas Q7 is measuring the perceived difficulty. Given the similarity between Q3 and Q7 it is expected that they will have a high positive correlation.

4. I understood the assessment requirements of the module

The problem with Q4 is similar to that of Q1. The students would have been informed about the assessments requirements as these are normally printed onto the reading lists given to the students at the first lecture for the module (as are the module's aims and objectives). It is possible that there is scope for a greater degree of legitimate variation in the ratings for Q4 compared to Q1; given the wording "I understood" compared to "I was made aware"; but this is unlikely. As with Q1 variations in responses are likely to reflect a student's attitude to the module as a whole rather than the specific question asked.

5. The teaching rooms for this module were fit for their purpose

This question has very little value on the student feedback form. The students are not really in a position to answer the question as phrased. What the students tend to answer is whether they *liked* the teaching room. For example the Loughborough University campus is large and lecture theatres are shared between departments and are spread throughout the campus. Therefore students can find their lectures being held in rooms a considerable distance from their own department, potentially involving a considerable walk. This situation may be reflected in a low score for this question. This situation is largely beyond the Business School's control and totally beyond the control of individual lecturers.

Not only is the question of little interest for the Business School, it is of very little practical value for the University's central administrators. This is because it is rarely possible to trace the room back from individual forms. This is because most modules will use a number of teaching rooms over the course of the module. The consequence of this is that from the numerical data based on the question it is usually impossible to tie particular rating scores to actual teaching rooms.

6. I found the library support for the module to be satisfactory

The problem with this question is that modules differ in their usage of the library. The required reading on some modules draws heavily on library resources, whereas others (especially some first year modules) stipulate required reading that is primarily based on a single textbook that the students are encouraged to purchase. There is therefore a problem in terms of between-module comparison of the responses to this question.

7. The subject material in this module is challenging

See comments under Q3.

8. The subject material in this module was presented at about the right pace

This question is ambiguous in nature. There is no sense of direction in the wording of the question. In particular, if the respondent disagrees with the statement this could imply that that the student perceived the pace as being *either* too fast *or* too slow. However there is value in this question in the context of its purpose. The purpose of the question is to determine whether there is a problem with the module. A response of strongly disagree or disagree highlights that there is a problem with the module in the mind of the student. Therefore the lower the class average for the module on this question the greater the problem. Although the responses to this question do not reveal whether the module was either too fast or too slow, an indication of this should be available to the lecturer from the written comments section on the reverse of the student feedback form.

9. The subject material in this module was related to business situations

This question reflects a key Business School objective that is set out in the Department's prospectus. The Business School chooses to use the student feedback mechanism as a means to gauge to what extent students perceive this to be realised.

10. The teaching on this module was well organised

There is an issue with this question as to the extent to which it reflects the student's judgement of the quality of teaching on the module. If this were to be the case there is a potential overlap with the questions in the second half of the student feedback form that deal specifically with the performance of the lecturer(s). However Q10 can be differentiated from the lecturer-specific questions, as there are often organisational matters central to the module that are independent of the lecturers. These could involve, for example, the integration into the module of outside speakers, the quality of fieldtrips or the appropriateness of the division of responsibilities between the lecturers involved with the module.

11. The coursework supported the module objectives

12. The tutorials linked with this module were useful

The final two questions in this section are only asked when relevant.

4.2.3 The lecturer-specific questions

There are six lecturer-specific questions on the student feedback form. These are presented in Table 4.2.

The questions are compared to those on three student feedback instruments that originate in the US. These three student feedback instruments were selected for two main reasons. Firstly, these three instruments have been extensively tested and shown to be valid in a variety of contexts (Clarkson, 1984; Watkins and Akande, 1992; Watkins and Gerong, 1992; Watkins and Regmi, 1992). Secondly, many of the current debates in the student feedback literature, for example the extent to which student feedback ratings are multidimensional (see chapter 3) base their analysis on available data from these instruments (Abrami and d'Apollonia, 1991; Abrami, d'Apollonia and Rosenfield, 1996; Cashin, Downey and Sixbury, 1994; Marsh, 1991a, 1991b, 1994, 1995).

The three selected instruments were:

a) The Endeavor Instructional Rating Form prepared by Peter Frey at Northwestern University (Frey, Leonard and Beatty, 1975);
b) The SEEQ Instrument prepared by Herbert Marsh at University of Southern California (Marsh, 1982, 1987);
c) The Instructional Development and Effectiveness Assessment (IDEA) System administered by the Center of Faculty Evaluation & Development, Kansas State University (Cashin and Downey, 1992; Cashin, Downey and Sixbury, 1994).

Table 4.2 Lecturer-specific questions on the student feedback form.

	SD	D	N	A	SA
13. The lecturer was well prepared	☐	☐	☐	☐	☐
14. The lecturer was enthusiastic about the subject	☐	☐	☐	☐	☐
15. The lecturer explained the subject clearly	☐	☐	☐	☐	☐
16. The lecturer encouraged participation and questions	☐	☐	☐	☐	☐
17. The lecturer overall, was effective	☐	☐	☐	☐	☐
18. The lecturer could be contacted for advice if needed	☐	☐	☐	☐	☐

The first four of these questions refer to different elements of teaching skill and reflect the Business School's concern to provide a high standard of teaching. The fifth (Q17) is an attempt at a general question that may be used to summarise the respondent's evaluation of the lecturer. As discussed in section 3.7 such a question is often termed 'global' in the student feedback literature (see for example Abrami and d'Apollonia, 1991; Abrami, d'Apollonia and Rosenfield, 1996). Such a question is a feature of many student feedback forms, for example Marsh's SEEQ form (Marsh, 1982). The final question on the form (Q18) was set centrally by the university. The Business School was compelled to have the question on the student feedback form. The results for each lecturer on this question were sent directly by Computing Services to the central university management. This is the only piece of lecturer-specific data to be received by the University management. Its presence on the form was controversial as many academics considered it an attempt by the University to gauge their movements outside of teaching commitments.

The lecturer-specific questions have considerable similarity with questions on other student feedback forms. Frey's Endeavor Instructional Rating Form has questions that closely resemble Q13, Q15 and Q16 on the Business School form. The three equivalent questions on the Endeavor Instrument are:

- 'Each class period was carefully planned in advance';
- 'The instructor planned the material carefully and summarised the major points';

- 'Class discussion was welcome in this course'.

Frey also has a question that is similar to Q18 on the Business School student feedback form. Frey's question reads:

- 'The student was able to get personal help in this course'.

Frey's wording is subtler than the wording on the Business School form as the latter refers specifically to lecturer's availability, whereas Frey's wording provides scope for an element of lecturer *attitude* to be included. Therefore Frey's question is not only attempting to measure the student's access to the lecturer but also attempting to measure the extent to which the student benefits from talking to the lecturer outside the confines of the lecture theatre. Therefore, in contrast to the Business School wording, Frey's question is able to more directly measure the usefulness of the student-lecturer interaction and is therefore likely to provide more meaningful data.

Marsh's SEEQ student feedback instrument has questions that closely resemble all six lecturer-specific questions on the Business School form. Unlike Frey, Marsh has an overall lecturer-specific question similar to Q17 on the Business School form. The equivalent SEEQ question reads:

- 'How does this instructor compare with other instructors that you have had at USC?'

Under the category of 'Individual Rapport', Marsh has four questions. Two of these questions are similar to the two variations of availability questions on the Business School and Frey's instrument. The two relevant questions on the SEEQ instrument read:

- 'Instructor made students feel welcome in seeking help and advice in or outside of class';
- 'Instructor was adequately accessible to students during office hours or after class'.

The IDEA system consists of two separate forms, which differ in their purposes. One form is a 47-item diagnostic form. Its aim is to help instructors improve their teaching by highlighting which aspects of teaching the students perceive the lecturer as being stronger or weaker at. The other form has 18 questions and is used exclusively for personnel decisions.

The 47-item form contains many of the questions on the Business School student feedback form. There are questions on encouraging participation and questions, lecturer's enthusiasm, organisation and clarity, clear statement of course objectives, the relation of the material to "real life situations" and the degree that the students are stimulated. There is also a question on the form that asks the students about the degree to which a lecturer was contactable outside the class, it reads:

- 'The instructor encouraged student-faculty interaction outside of class (office visits, phone calls, e-mails, etc.)'.

The high degree of similarity of the questions will ultimately facilitate a comparison of the statistical patterns from these instruments with the main models of the Business School student feedback data presented in chapter 10.

The eighteen question Loughborough University student feedback form covers most of the issues raised on the three U.S. student feedback forms discussed in this section. SEEQ and IDEA, which are both considerably longer than the Loughborough University form ask the students how they feel about the grading of their work on the module. This issue is not covered on the Loughborough University form. As a consequence of the time scales created by the modular system that divides the academic year into two fifteen week teaching periods, students are normally asked to complete student feedback forms for each module before they have received the results of their coursework for a module. The students therefore would be unable to meaningfully answer a question regarding grading. Another significant difference between the SEEQ and IDEA feedback forms and the Loughborough University form, is that the former ask questions that measure the student's motivation. For example, asking the student the number of hours they spend studying material pertinent to the module outside class and the level of interest they had in the subject prior to starting the module. The Loughborough University instrument has no questions of this type.

4.2.4 Expectations and implications for subsequent analysis

The expected pattern of responses for the module-specific questions (Q1-Q12) as outlined in section 4.2.2 will be explored in chapter 7, in which a sample of 65 modules are studied using cluster analysis. It is expected that Q1 and Q4 will be shown not to differentiate clusters within modules. Students on any particular module are expected to show very little differentiation in their ratings for either of these two questions. The analysis should also emphasise that students on a particular module give a similar rating for Q3 and Q7. The cluster analysis presented in chapter 7 will reveal the variables that show the most variation in each of the 65 modules sampled.

Regression analysis presented in chapters 8 and 9 use the most general of the lecturer-specific questions (Q17) on the Business School student feedback form as the dependent variable. This type of question has been shown to be a reliable way of representing a lecturer's overall performance (see section 3.7).

The main model of the Business School student feedback data presented in chapter 10 is run separately on each of the lecturer-specific questions (Q13-Q18). This allows an analysis of whether the lecturer-specific questions are affected by a set of predictor variables in similar ways. If one variable (e.g. Q18) is being treated differently by the students compared to the other lecturer-specific questions this will be apparent. Modelling student feedback data against various predictor variables simultaneously (as in chapter 10) provides an opportunity to measure which of the predictor variables affect each of the six lecturer-specific questions. It also allows comparisons to be made determining whether individual predictor variables are consistent in their impact on the different aspects of teaching effectiveness (i.e. the different lecturer-specific questions, Q13-Q18).

4.3 The mechanism for collecting student feedback

The academic year at Loughborough University is divided into two fifteen-week semesters. The first semester runs October-January and the second February-June. Loughborough University policy is to require student feedback data to be collected for a minimum of one third of a department's modules each semester. The Business School has chosen to collect student feedback data for all taught modules in each semester, therefore there are two sets of data for each academic year. Student feedback forms are printed and sent to departments by the University's Quality Assessment Unit. These forms have the seven mandatory questions set by the University already printed. The Business School has a further eleven questions that they overprint on each of the feedback forms. The feedback forms for semester one modules are distributed in week ten of the semester (early December). In semester two the forms are also distributed to the students in week ten of the semester which falls shortly after the Easter vacation, in early May.

The procedure for collecting student feedback can be outlined as follows:

- Lecturers distribute the feedback forms to students in lectures following a prescribed set of guidelines which aims to foster consistency. Students are asked to fill in the forms whilst in the lecture theatre. The policy of asking students to respond to the questionnaire whilst remaining in the lecturer theatre is designed to maximise response rates. However it is recognised that this may lead to forms being completed quickly and with only limited thought being given to each of the responses.
- The forms are collected and returned to the Undergraduate Office. The forms from all the semester's modules are collated, packed up and sent to Computing Services.
- In Computing Services, the forms are read electronically via the use of an Optical Mark Reader. The data currently is sent to the Business School by e-mail, but in the period of this research 1996-1998, the data was put onto a disc and sent to the Business School with the student feedback forms.

- Summary sheets reporting the means for each question for each module are prepared and sent, with the original student feedback forms, to the leading lecturer on the module.
- The lecturers on the module consider the quantitative data and the written comments on the reverse of the forms.
- The summary data is also forwarded to the Subject Area Co-ordinator (SAC) who is the person responsible for the relevant section of the department.

The leading lecturer on the module also has the opportunity to read the written comments on the reverse side of the feedback forms prior to the forms being sent to Computing Services. This allows the lecturers to incorporate any constructive criticism or to build on the well-received aspects of the module. Although the summary data is not available to the lecturer until after the module is completed, the opportunity to read comments made by the students makes some swift responses possible. One of the potential problems with this system is that it allows for the possibility of unscrupulous lecturers removing from the batch any forms that depict them in a negative light. To guard against this, response rates for each module are calculated and circulated to every member of academic staff in the department. The response rate is the percentage of registered students on the module who completed a student feedback form for that module. It can be formally written as:

$$\text{Response rate} = \frac{\text{Number of responses}}{\text{Number of registered students}} \times 100$$

In the past response rates have ranged between 14% and 100%. Nearly two thirds of the modules (62.7%) had response rates of 70% or better. The average response rate was 72.7%. This contrasts favourably with response rates found in many questionnaire-based studies. A high response rate creates confidence that the available data is representative of the undergraduate student population in the Business School.

Response rates are the only piece of information that is widely circulated in this manner. The data on individual modules is only seen by the Undergraduate Programmes Director (UPD), the lecturer(s) on the individual module and the subject area cordinator (SAC) who is responsible for the relevant section of the department. The department considers that this policy of confidentiality should put lecturers at their ease and make the exercise less threatening and more productive.

4.4 Data preparation

Completed student feedback forms are packed up and sent to the University's Computing Services section. They are fed through an Optical Mark Reader (OMR) with the data for each module being saved in a separate text file. These files are put onto a disc that is sent, with the original feedback forms, back to the Business School. The Business School converts each text file into an Excel worksheet and each worksheet in turn is fed into an Excel template to produce the summary printout.

The use of the OMR allows data to be read quickly and cost effectively. Unfortunately some responses are lost if the forms themselves are damaged or if the respondent has not closely followed the instructions for filling in the forms. Responses will not be read if:

- There are multiple marks for a single question, i.e. if more than one response category has been marked for a single question. This is often the result of the respondent having changed their mind, i.e. two boxes have been filled in and one crossed out. The OMR has no way of understanding this, but the respondent's intention is clear when one looks at the actual form. Another example of this is occasionally found in response to question 1, occasionally respondents mistake the direction of the scale and respond (for example) 1 instead of 5. In all of these cases a ">" symbol will appear in the text file.
- The marks are non-centralised, i.e. if the mark only covers one side of the box.

- Alternative marks such as a tick or cross have been used. Only a solid mark through the entire box will be read by the OMR.
- A red or very light blue pen has been used. Responses made in green pen are usually read. In situations in which the OMR is unable to detect a response the relevant cell in the text file is blank.

The following analysis measures the extent to which the data that the OMR registered as missing was actually missing rather than being misread or marked in an unclear manner. If the data had been misread or marked in an unclear manner then it would be available upon examination of the original student feedback forms. To measure the extent to which 'missing' cases were not actually missing but unread or unreadable, a sample of 65 modules were analysed. In total these 65 modules consisted of 5012 student feedback forms. The 65 modules were selected on the basis of being modules with the greatest number of responses. Consideration was also given to response rates. Some of the largest modules, in terms of registered students, were not selected as they had very low response rates. The problem with low response rates is the question this raises over the generalizability of the sample. For each module the data set was checked and the total number of missing values was noted. A note was also made of the number of cases and variables within the data set that contained missing values. Once this had been done the original feedback forms were compared with their entry in the data set. This was easy in those cases where the forms had been read but not shuffled by the relevant lecturer, so that the order of the forms in the pack and in the data set was the same. For modules where the original forms had been shuffled it was usually possible to match forms with their entry in the data set by observing the pattern of responses. The average response rate for the 65 selected modules was 72.7%. This mirrors exactly the average response rate for all 305 modules. Allied to the basis of selecting the modules, this gives a degree of reassurance that the sample is representative of the whole data set.

The 65 modules were taken from all three of the undergraduate levels (i.e. first year undergraduate, second year undergraduate and finalists). None of the 65 modules were drawn from the taught course postgraduate modules as only one postgraduate module had greater than 50 responses (54). Other that this exception, the highest number of returned forms for any of the taught course postgraduate modules was 35.

Table 4.3 Pattern of missing values in a sample of 65 modules.

Average number of:	Data set read by OMR, pre checking of original forms	Data set after the original forms were checked
Missing values	24.5	8.2
Variables with missing values	12.4	5.0
Cases with missing values	9.1	3.6

Table 4.3 shows the pattern of missing values for the sample of 65 modules. The second column in Table 4.3 reports the average number of missing values in the data prior to the data sets being checked. This average was 24.5; i.e. each Excel worksheet contained an average of 24.5 data cells in which no data point was present. The number of missing values was significantly different over the three module levels. The average number of missing values for forms returned by first year undergraduates was 24.9; this closely mirrors the overall average. The average for second year undergraduates was 28, significantly above the average. In contrast final year undergraduates tended to leave fewer questions on the form unanswered or incorrectly marked. The average number of missing values for final year undergraduates was 16.2. The differences between the average number of missing values for each module level category and the overall number of missing values was statistically significant at the 1% level.

The final column in Table 4.3 reports the number of missing values in the data once the original student feedback forms had been checked. The average number of genuinely missing values in the 65 selected data sets reduced significantly once the original student feedback forms had been checked. The average number of genuinely missing values (post-checking) was 8.2. This implies that only a third (33.6%) of the values entered as missing on the Excel spreadsheets were genuinely missing.

Interestingly this pattern was not consistent over the three module levels. One could hypothesise that the proportion of missing values that are due to incorrectly marking the student feedback forms, i.e. not actually 'missing' upon examination of the original student feedback forms, would be lower for second and final year undergraduates compared to first year undergraduates. The reasoning would that as the students develop an increasing degree of experience in completing the feedback form they become increasingly competent in their ability to complete the form. The data supports this hypothesis. For the first year undergraduates 71.5% of the data points registered as missing by the OMR were distinguishable as valid responses once the original student feedback forms had been checked. This is in contrast to figures of 62.1% and 64.8% for second year and final year students respectively. The differences between both the first and second year undergraduates to the overall sample mean were statistically significant at the 1% level. The difference between the final year undergraduates and the overall sample mean was statistically significant to the 5% level ($p = 0.015$). This suggests that those in the second and final year of their degrees are making fewer errors in the completion of their forms, for example not adequately filling the whole area of the relevant box, or marking more than one box for a particular question.

Although this analysis illustrates that by checking the original forms a large percentage of values that the OMR registered as missing were apparent upon reference to the student feedback forms, the problem of missing values overall is very small. The average number of missing values for each module, 24.5 (see Table 4.3), represents only 1.5% of the total number of possible responses from the students who filled out student feedback forms. For example in a two-lecturer module in which all twelve module-specific questions have been used there are 24 variables (twelve module-specific questions and two sets of six lecturer-specific questions). If 50 students have filled out student feedback forms for the module there are 1200 possible data points for the module (i.e. 24 x 50). Therefore an overall average of missing value rate of 1.5% would imply that there were 18 missing values on this module (i.e. 1200 x 0.015). This evidence illustrates that there is not a significant problem with missing values in the data set used in this research.

Whilst the number of missing values is not a problem when the purpose is the general averaging of ratings, the number of values registered as missing by the OMR that are apparent upon inspection of the original student feedback forms becomes more important if the purpose is to conduct more sophisticated statistical analysis. Techniques such as factor analysis and cluster analysis include only complete cases in the calculations; cases with missing values are discarded. The final row of Table 4.3 presents the figures for the average number of cases with missing values in the data sets read by the OMR and compares this to the average number of cases with missing values after the original forms had been inspected. Prior to the inspection of the actual forms the average number of cases with missing values was 9.1. Given that the average number of cases in the 65 modules sampled was 77.1, this represents 11.8%, i.e. 11.8% cases were not available for factor analysis and cluster analysis. After the inspection of the original forms the average number of cases with missing values falls to 3.6, representing 4.7% of cases. This decrease is appreciable and stresses the value of checking the original forms where practicable, especially if the data is to be subjected to multivariate techniques such as factor analysis and cluster analysis.

The means and standard deviations for each of the questions on the student feedback form are shown in Table 4.4.

Table 4.4 Means and standard deviations for each of the questions on the student feedback form

Question	Mean	Standard Deviation	Number of responses
Q1	3.90	.69	13619
Q2	3.70	.87	13653
Q3	3.51	.95	13662
Q4	3.91	.73	13603
Q5	3.86	.83	13629
Q6	3.27	.83	13553
Q7	3.70	.86	13597
Q8	3.62	.90	13585
Q9	4.03	.78	13608
Q10	3.92	.81	13611
Q11	3.92	.77	12891
Q12	3.71	1.00	6311
Q13 Lecturer 1	4.19	.68	13195
Q13 Lecturer 2	3.95	.79	5946
Q13 Lecturer 3	3.79	.72	1839
Q14 Lecturer 1	4.26	.73	13181
Q14 Lecturer 2	3.99	.84	5953
Q14 Lecturer 3	3.65	.82	1835
Q15 Lecturer 1	3.85	.90	13188
Q15 Lecturer 2	3.62	1.01	5952
Q15 Lecturer 3	3.38	.98	1835
Q16 Lecturer 1	3.89	.89	13179
Q16 Lecturer 2	3.74	.91	5957
Q16 Lecturer 3	3.51	.89	1833
Q17 Lecturer 1	3.93	.86	12910
Q17 Lecturer 2	3.68	.96	5796
Q17 Lecturer 3	3.38	.92	1750
Q18 Lecturer 1	3.76	.83	12919
Q18 Lecturer 2	3.51	.83	5775
Q18 Lecturer 3	3.36	.77	1766

It is clear from looking at Table 4.4 that the standard deviations for each of the questions on the student feedback form are very similar. The standard deviations range from 0.69 to 1.01. To the extent to which there is a little variation in the standard deviations, the variation is in line with the expectations outlined in section 4.2.2. The two module-specific questions that were expected to have the smallest standard deviations (i.e. Q1 and Q4 - which essentially ask students questions of fact rather than opinion) are the two module-specific questions with the smallest standard deviations.

Another noticeable feature of Table 4.4 is that there are clear differences in the means for each of the questions between the three positions on the student feedback form. The possible reasons for these differences will be considered in chapter 8.

Chapter 5

METHODS OF DATA ANALYSIS

5.1 Introduction

In this chapter the methods by which the data sets used in the research were constructed and analysed are explained. Section 5.2 provides a detailed description of how the master data sets used in the research were constructed from the original text files prepared by Computing Services at Loughborough University. The composition of these data sets are outlined in section 5.3 which presents statistical summaries of descriptive variables that may influence the ratings received by lecturers. These descriptive variables were entered into the data sets alongside the rating scores given by the students. These descriptive variables are used in the analyses presented in chapters 8-10. The following section (5.4) describes the mechanism whereby the module level data set was created based on the data sets described in sections 5.2 and 5.3. The three main multivariate statistical techniques used to analyse the data: cluster analysis, factor analysis and multiple regression are discussed in section 5.5. Each of these techniques is taken in turn with the emphasis placed on discussing the issues raised by the implementation of these techniques in the analyses presented in chapters 6-10.

5.2 Construction of master individual form level data sets

The data used in this research is drawn from two consecutive academic years, 1996-1997 and 1997-1998. The total data set consists of student feedback data for all 305 modules taught during the four semesters October 1996-June 1998. The number of modules and the number of individual feedback forms for each of the four semesters is presented in Table 5.1.

Table 5.1 Number of modules and feedback forms by semester

		NUMBER OF MODULES	NUMBER OF FEEDBACK FORMS
Academic Year 1996-1997:	Semester 1	78	3226
	Semester 2	69	3173
Academic Year 1997-1998:	Semester 1	78	3652
	Semester 2	80	3763
	TOTAL	**305**	**13814**

In chapter 4 the process whereby the student feedback data is read and stored on disc by Computing Services was outlined. The disk sent to the Business School by Computing Services consists of a text file for each module and these text files are routinely converted into Microsoft Excel worksheets. To construct the master data sets used in this research, the data for each module needed to be combined into a single data file. To do this the first text file was opened and converted into an Excel worksheet. It was then copied and pasted into a new Excel worksheet. Subsequently, each text file was opened in sequence, converted into a worksheet, copied and pasted into the new spreadsheet with the first entry of the latest module directly following the final entry of the previous module.

One problem with this was that Computing Services change the layout of their student feedback text files at regular intervals, so that the column ordering was not the same for the text files over the four semesters used in this research. To overcome this complication the process outlined above was performed separately for the four semesters. Adjustments using the standard features of the Excel package were made to standardise the format of each semester's set of data. At this point the data for all four semesters were joined into a single data file.

5.3 Composition of master individual form level data sets

The data set was arranged into two different versions to aid the analysis.

5.3.1 Version A

The first version consisted of all the student feedback data and was the result of the process described in section 5.2. The Excel spreadsheet consisted of 40 columns and 13814 rows. This represents the total number of student feedback forms returned from the 305 modules. Integral to the analysis presented in chapters 8, 9 and 10 is the measurement of the extent to which various predictor variables can influence the student feedback ratings that students give lecturers. Therefore in addition to the thirty columns of student responses (i.e. columns for the twelve module-specific questions and six lecturer-specific questions for up to three lecturers), there are ten columns containing descriptive data of the individual modules. These are:

a) *Year.* The academic year in which the module was taught.

b) *Semester.* The semester in which the module was taught.

c) *Module Level.* This ranges between 1-4 where 1 indicates first year undergraduate and 4 indicates Masters level taught module. Table 5.2 presents the number of modules at each level, with breakdowns for each semester.

Table 5.2 Frequency of modules at each level by semester

	MODULE LEVEL				
	1	2	3	4	TOTAL
1996-1997: Semester 1	18	20	25	15	78
Semester 2	18	21	21	9	69
1997-1998: Semester 1	17	19	23	19	78
Semester 2	18	20	23	19	80
TOTAL	**71**	**80**	**92**	**62**	**305**

d) *Module code.* This is the individual code used to identify each module. The module code consists of a six digit number where the first two digits represent the year of the data (either 96 or 97), the third the module level (ranging from 1-4) and the remaining three digits repeated those of the original module code as it appeared in the files prepared by Computing Services. These original codes consisted of both letters and numbers, to allow the data to be more easily handled in SPSS it was necessary to make these codes wholly numerical.

e) Subject area. The modules were divided into 10 categories these are listed in Table 5.3. In cases of ambiguity, the module was coded as 10 which was termed 'other'. Table 5.3 shows that students at the Business School study a wide range of academic modules. These range from modules that could be placed into the category humanities (e.g. Human Resource Management), through modules that could be considered social science (e.g. Marketing Analysis and Decision Making) to those that could be termed science (e.g. Business Forecasting). Research evidence indicates that different academic disciplines have differing average ratings (see section 3.6). As a result it is potentially important to consider the effects of subject area as the different subject groups within the Business School could be considered surrogates for academic discipline. Table 5.4 presents the frequency of modules in each subject area. This table gives both overall totals for each subject area and subtotals of each semester.

Table 5.3 Listing of the ten subject areas

Code	**Subject Area**	**Abbreviation**
1	Accounting and Financial Management	AFM
2	Banking	BK
3	Economics	ECON
4	Human Resource Management	HRM
5	Information Systems	IS
6	Marketing	MAR
7	Strategic Management	SM
8	Retailing	RET
9	Quantitative Methods	QU
10	Other	OTHER

Table 5.4 Frequency of modules in each subject area

	SUBJECT AREA										
	1	2	3	4	5	6	7	8	9	10	TOTAL
1996-1997: Semester 1	14	5	8	15	6	10	3	4	7	6	78
Semester 2	16	5	4	13	4	7	2	4	9	5	69
1997-1998: Semester 1	13	4	6	17	6	11	4	4	6	7	78
Semester 2	16	6	4	11	6	8	3	4	10	12	80
TOTAL	**59**	**20**	**22**	**56**	**22**	**36**	**12**	**16**	**32**	**30**	**305**

f) Class size. For each module, data was available for both the number of registered students on the module as well as the number of actual responses. As discussed in section 3.6 there is clear evidence in the literature that class size affects ratings. As the number of registered students will be a closer representation of the number of students generally attending the teaching sessions compared to the number of feedback forms completed for the module, the number of registered students is taken as the measure of class size. For the modules used in this research, the number of registered students ranges between 3 and 248, the average number of registered students is 65.5.

g) Number of lecturers on the module.

h) Lecturer 1 code number.

i) Lecturer 2 code number.

j) Lecturer 3 code number.

In total 45 lecturers are included in the data set. All of these lecturers are full time members of the department; they span the full range of the academic hierarchy. Data for 16 additional lecturers was available for a small number of modules. This data was not included in the research as the lecturers concerned tended to be highly inexperienced and were not permanent members of the department (often only visiting the Business School to give specific lectures). Each lecturer in the Business School was assigned a code number. For each lecturer on each module this code number was entered into the data set and could therefore be used to identify each of the lecturers. As discussed in chapter 4, the structure of the student form allows for up to three lecturers to be evaluated on a module, therefore three columns of lecturer code numbers were added to the data set. Again numerical code numbers were used, as these are more convenient in terms of sorting the data and converting it between Microsoft Excel and SPSS.

5.3.2 Version B

The aim in creating a second version of the data set was to separate lecturer student feedback data so that the data set could conveniently be sorted by lecturer. This version (version B) contains only the data referring to the lecturer (i.e. questions 13-18 on the student feedback questionnaire) plus the descriptive data. The student feedback form contains space for feedback on up to three lecturers on an individual module. The text files prepared by Computing Services place the lecturer data for questions 13-18 directly after the data for questions 1-12. The data for each of the questions is placed in sequence, implying that the data for the lecturers is interleaved. This spreadsheet design does not allow the data to be easily sorted by lecturer. This is because an individual lecturer is likely to vary in the position that they appear on the student feedback questionnaire and hence their feedback will be recorded in different positions. In some modules a particular lecturer will be lecturer 1, in others they will occur as lecturer 2 or lecturer 3.

The spreadsheet in version B was arranged so that data for the first named lecturer on the module is placed directly above that the second named lecturer on the module. Similarly the data for the third named lecturer on the module is placed directly below that of the second named lecturer. For example a module taught by two lecturers with 50 completed feedback forms would cover 50 lines in version A of the master data set but would cover 100 lines in version B of the master data set. The 50 cases for lecturer 1, each with descriptive information about the module would be followed by the 50 cases for lecturer 2 again with the descriptive information. For analysis purposes it is important that the descriptive information relating to the module is duplicated.

Just over half of the modules (155) were taught by a single lecturer (51%), 34% of modules (103) had two named lecturers and 13% of modules (41) had three named lecturers. Six modules (2%) had no named lecturers, these modules were student centred and therefore involved little lecturing.

In total the spreadsheet for version B of the master data set consists of fourteen columns and 20213 rows. Six of the fourteen columns give data for questions 13-18. The other eight columns duplicate those in version A, namely: year, semester, module level, module code, subject area, class size, number of lecturers on the module and lecturer code. There are not ten additional columns as in version A, as in version B there is only one lecturer code per row as each lecturer on each module is placed as a separate row in the data set.

The spreadsheets of both versions of the data were then converted into SPSS files.

5.4 Construction of master module level data set

The master module level data set was created directly from the individual form level data set described in section 5.3.2, by using the subtotal command in Microsoft Excel.

The master module level data set consists of 458 rows, with a separate row for each lecturer on each module in the data set. Alongside the class averages for each variable there are the eight descriptive variables previously discussed: year, semester, module level, module code, subject area, class size, number of lecturers on the module and lecturer code. In addition there were six extra descriptor variables not included in the individual level data sets. These are linked to the regression analysis in chapters 8, 9 and 10 and further details are given in those chapters.

The module level data set was created in order to run the regression analysis presented in chapters 8, 9, and 10. The sequencing of the following chapters reflects the statistical techniques used and the level of data that is most appropriate to use with the various techniques. Chapters 6 and 7 analyse the degree of structure in the data at the individual form level. One of the central aims of chapters 6 and 7 is to explore whether there is a common view in the class of a module's quality and the lecturer's performance, with the students responses being distributed around a single mean, or whether the students divide into discrete groups who judge the quality of the module and the performance of the lecturer(s) differently. For this reason the individual form level data was used. Chapters 8, 9 and 10 use regression as the principal data analysis tool. These chapters aim to model whether and to what extent a range of predictor variables affect the student feedback ratings for a lecturer. For this analysis interest in the degree to which these variables influence ratings is at the module level. It should be noted that if the regression analyses were run at the individual form level then the effects of the predictor variables would be swamped by the differences between respondents, i.e. by the extent to which individual students differed in their judgement of the quality of the module and the effectiveness of the lecturer(s).

5.5 Statistical methods

5.5.1 Statistical techniques used

The three main multivariate statistical techniques used in this research were cluster analysis, factor analysis and multiple regression analysis.

Cluster analysis is used in chapters 6 and 7 to explore the appropriateness of treating the class as a coherent homogeneous entity (as discussed in chapter 3). In chapter 6 emphasis is given on the appropriateness of factor analysis to adequately extract the underlying dimensions of student feedback data when the assumption of the class being an homogenous entity is violated. Cluster analysis is used in chapter 7 to explore the extent to which there are distinct groups of students in a sample of Business School modules who are systematically evaluating the quality of the module and the performance of the lecturer differently.

Multiple regression is used in chapters 8, 9 and 10. In chapter 8 it is used to determine whether the position that a lecturer appears on the student feedback form influences the ratings that a lecturer receives. In chapter 9 multiple regression is used to determine whether there is a contrast effect in lecturers' student feedback ratings, through examining whether the ratings that a lecturer receives on a module are influenced by the performance of the lecturer(s) with whom they teach on a particular module. In chapter 10 multiple regression is used to model lecturers' student ratings against a range of predictor variables that have been suggested in the literature or in this research as possibly influencing lecturers' student feedback ratings.

In this section the key features of these multivariate statistical techniques and the key decisions involved in carrying out these techniques are discussed. In particular the reasons why the methods employed were selected and alternative methods that could have been used instead will be discussed.

5.5.2 Cluster Analysis

Cluster analysis is a classification technique that aims to place observation into groups based on their similarity to each other, 'in a broad sense clusters are thought of as collections of points which are relatively close, but which are separated by empty regions of space from other clusters' (Sneath, 1969, p.260).

Clusters of objects should 'exhibit high internal (within-cluster) homogeneity and high external (between-cluster) heterogeneity' (Hair, et al., 1995, p.423). In other words, if the classification is successful, the objects within each cluster will be close together and objects within different clusters will be far apart.

In contrast to factor analysis in which links between variables are studied, cluster analysis studies links between individual cases, i.e. between students. Cluster analysis is an important technique in student feedback research as it allows the researcher to examine the extent to which there are discrete groups of students in the class who are evaluating the quality of the module and/or the performance of the lecturer differently.

Milligan and Cooper (1987) suggest that there are seven steps in performing a cluster analysis:

1. The cases to be clustered need to be selected. These cases should be representative of the population.
2. The variables to be used in the analysis need to be selected. These variables must contain sufficient information to permit the clustering of the cases.
3. The analyst must decide whether to standardise the data.
4. The similarity or dissimilarity measure needs to be selected. These measures assess the degree of closeness or separation between the cases.
5. The clustering algorithm needs to be selected. Different clustering algorithms are designed to find different types of cluster structure. For example the single link method is designed to detect outliers.
6. The number of clusters needs to be determined. There are a large number of methods for deciding on the number of clusters. None of the methods seem to be universally successful.
7. The cluster solution needs to be validated. For example, the researcher may wish to test the generalizability of the cluster structure by attempting to replicate it on another data set.

The first step of the cluster analysis process is to determine the cases to be clustered. The cluster analysis presented in chapters 6 and 7 used all the returned student feedback forms for each of the selected modules baring those that contained missing values (see discussion in section 4.4). Sustained attempts were made to reduce the number of forms that had to be discarded due to the presence of missing values by carefully checking the original student feedback forms for each of the modules used in the cluster analysis. The next step (step 2) of the cluster analysis process is to select the variables to be used in the analysis. The cluster analysis was carried out using each of the variables on the student feedback form. Hence in the present case little judgement was needed in steps 1 and 2 as it was decided to be comprehensive and simply use all available data while taking care to maximise the availability of data by systematically checking the original student feedback forms to reduce the number of missing values (as discussed in section 4.4).

The next step (step 3) in conducting cluster analysis is to decide whether to standardise the data. The most common methods of standardisation is the conversion of each variable used in the analysis by subtracting the mean and dividing by the standard deviation for that variable. This process converts each variable into a standardised variable with a zero mean and a unit standard deviation (Hair, et al., 1995, p.435). There are two main situations in which standardising the data is appropriate. The first is when the variables used in the analysis are measured on different scales, for example when some of the variables are measured on a Likert scale and other variables are measured as percentages. The second is when the variables used have significantly different dispersions. Variables that have a larger dispersion than others will have a greater weight in the cluster analysis (Hair, et al., 1995). The data used in this research was not standardised as all the variables are measured on the same five point Likert scale and the standard deviations of the variables used in the analysis were similar (see Table 4.4, section 4.4).

Step 4 of the cluster analysis procedure (i.e. selecting the similarity measure) is often closely linked to step 5 (i.e. selecting the clustering algorithm). The decision making in step 5 will be considered first. In deciding which cluster algorithm to use, the researcher can choose from several hierarchical cluster algorithms; examples include single linkage, complete linkage, average linkage, Ward's minimum variance method and the centroid method. Hierarchical methods require some measure of distance so that the concept of closeness can be measured. In the single linkage method, for example, the algorithm calculates the two closest points and joins them together. It then calculates the case that is closest to one member of the cluster and adds that case to the cluster. The cluster is thus extended to all possible cases that are continuously linked together via at least one member (Lorr, 1983). When the aim of the analysis is to produce clusters that are as homogeneous as possible the long extended chains of cases produced by the single linkage method is an weakness. This would be a particular problem in a situation where two broad clusters each has a case that is close to the boundary where the distance of these two cases was small. The single linkage method would have joined these two cases together at an early stage therefore distorting the cluster structure in the data.

The average linkage method, which is computed on the basis of average distance from all cases in one cluster to all cases in another. This method avoids the problems of outlying cases, but tends to be biased toward producing clusters with approximately the same variance (Hair et al., 1995).

The cluster analyses presented in this research were carried out using Ward's minimum variance method (Ward, 1963). At each stage in the clustering procedure, clusters are combined so that the within-cluster sum of squares is minimised over all partitions obtainable by combining two clusters from the previous stage (Hair, et al., 1995).

There are two reasons why Ward's minimum variance method was selected. Firstly, there is considerable evidence that it is the best hierarchical clustering method (Mojena, 1977; Blashfield, 1976). Secondly, Ward's minimum variance method is relatively robust compared to other hierarchical methods such as average linkage method in that it will continue to operate effectively when clusters are not well separated (Mojena, 1977). This is important in the analysis of Business School data as evidence presented in chapter 7 illustrates that the clusters are often not well separated. Ward's method also has the benefit of minimising the within-cluster differences and avoids problems with the 'chaining' of cases which is a problem for some of the linkage methods (Hair, et al., 1995).

An alternative to using a hierarchical clustering algorithm is to use non-hierarchical methods, one example of which is k-means cluster analysis. K-means cluster analysis is an iterative partitioning method which divides observations into some predetermined number of clusters. In contrast to hierarchical methods which use a tree like construction process, non-hierarchical methods use cluster seeds by which to group cases which are closest to those seeds.

The centroids of these clusters are calculated and observations are then reassigned to clusters (if they are closer to another cluster than the one they were originally assigned) until some decision rule terminates the process (Punj and Stewart, 1983).

Non-hierarchical methods include the 'sequential threshold' method, the 'parallel threshold' method and the 'optimization' method. They differ in terms of the method of selecting the starting seeds, the cluster reassignment process, the decision rule used for terminating the clustering process and the frequency with which cluster centriods are updated during the reassignment process.

Although k-means cluster analysis has the advantage of being relatively unaffected by outlying cases and the inclusion of irrelevant or inappropriate variables in the analysis (Punj and Stewart, 1983), k-means has the major problem of the determination of the starting seed used in the analysis.

The selection of the cluster seed is a crucial part of the process. Many statistical packages (including SPSS) use essentially random starting seeds, often based on the ordering of cases in the data set. The implication of this is that were the ordering of cases in the data set to be changed (i.e. the data set were to be shuffled) the result of the cluster analysis would also change.

Hair, et al., (1995) comments:

> 'the use of non-hierarchical techniques with random seed points is markedly inferior to the hierarchical techniques. Even a non-random starting solution does not guarantee an optimal clustering of observations' (Hair, et al., 1995, 442).

For this reason a hierarchical method (specifically Wards minimum variance) was selected for the cluster analysis for the various sets of data used in this research.

One possibility is to use k-means analysis as a follow up to a hierarchical cluster analysis (e.g. conducted with Ward's minimum variance method) using the cluster centriods obtained from Ward's minimum variance method as the starting seeds to be used in the k-means analysis. This approach would allow the advantages of k-means analysis to be potentially realised without the weakness of random cluster seeds.

In effect such an approach would be an act of 'fine tuning', which would permit the switching of cluster membership of some of the cases in the data set. Since there is considerable evidence that Ward's method is the best hierarchical clustering method, one would expect little difference in the k-means cluster formations compared to the results of Ward's minimum variance method. The application of a k-means cluster analysis following a hierarchical cluster analysis would possibly lead to some cases (i.e. students) 'jumping' clusters (i.e. students that were on the boundary of one cluster may switch to another) however such fine tuning was considered unnecessary as the aim was to study the broad picture for a large number of modules. In particular the emphasis was on the existence of clusters not their precise location.

The two-phase methodology (i.e. hierarchical cluster analysis followed by k-means cluster analysis) would be appropriate if the primary aim of the analysis was to understand which students fit into a particular cluster (i.e. where the actual placement of individual cases, in this case students, was important). This research is not concerned with the cluster membership affiliation of individual students, indeed given the anonymity of the student feedback form, the researcher has no useful information with which to determine which individual students belong to which cluster. Rather, the aim of the research is to understand the nature of the clusters and to explain the variation of perspective in the class in a meaningful manner. A small number of individual students 'jumping' clusters would not change the overall nature of the cluster. It was decided therefore that the two-phase methodology was not necessary for the data used in this research.

Returning to the issue of similarity measure (step 4), squared Euclidean distance is the recommended measure for Ward's method of clustering (Hair, et al., 1995) and so this similarly measure was used. This reflects the fact that Ward's minimum variance method is based on the minimum sums of squares so squared Euclidean distance which is itself based on sums of squares is ideal. There are, in general, other distance measures that could have been used, for example the city-block approach. In contrast to squared Euclidean distance, which is based on the sum of the squared distances between cases, the city-block approach is based on the sum of the absolute differences of the co-ordinates for the cases. The city-block method has the methodological weakness of assuming that the variables are not correlated with each other. In data sets in which the variables are correlated with each other the clusters computed using this method are not valid (Hair, et al., 1995).

The next step (step 6) in the cluster analysis process is for the researcher to determine the number of clusters present in a data set. One of the most difficult and most important decisions that the researcher has to make in the cluster analysis process is to decide whether there are genuine clusters in the data (i.e. whether the data set should be partitioned at all). The second and equally difficult and important decision is to decide on the number of clusters that represents the most suitable partitioning.

The first issue involves differentiating a genuine cluster structure from the division of a continuum. This is difficult in cluster analysis, since hierarchical cluster analysis necessarily produces apparent clusters even when no meaningful groups are embedded in the data (Barney and Hoskisson, 1990). One method of determining whether there is a genuine cluster structure in the data is to compute the cluster centroids for each variable in each cluster separately and then examine these centroids to determine whether there is a meaningful pattern suggestive of a genuine cluster structure. Evidence of a cluster structure would be indicated by several variables clearly differentiating between the clusters, by virtue of having substantially greater differences between their centriods compared to the other variables in the analysis. In contrast, evidence of a continuum would be found if the between-cluster differences in the centriods of each variable were similar. This would suggest that the cases in a data set with two clusters had simply been divided into the highest scoring half and the lowest scoring half.

There are a number of approaches to the second issue. One way of determining the most suitable partitioning of a data set is to use one of the available stopping rules. There are a large number of stopping rules that differ both in their complexity and the extent to which their methodology is inherently subjective or objective. The performance and general value of the Calanski and Harabasz stopping rule in analysing student feedback data is discussed in section 7.6.

The final step (step 7) in the cluster analysis process is to validate the cluster structure. Several ways to validate a cluster structure have been suggested. The most common way is to divide the sample into two halves and carry out clustering on each half. The researcher can then check the similarity between the cluster solutions from each half and from this information can assess the validity of the cluster structure (Punj and Stewart, 1983). This strategy was not appropriate in this research for two reasons. Firstly, the primary purpose in conducting the cluster analysis was to determine whether clusters existed, rather than describing their precise location. Secondly, many of the modules on which cluster analysis was carried out were not sufficiently large for cluster analysis to be carried out on separate halves on the total cases in each module. Another way of validating a cluster structure is to use an alternative method. Chapter 7 presents two alternative methods of validating a cluster structure. Firstly, the examination of the distribution of the data and secondly examination of the bivariate correlations of variables in the entire data set and within each cluster separately. Each of these approaches were conducted in a thorough manner. These methods provided quite convincing evidence supporting the presence of cluster structure in some Business School modules.

5.5.3 Factor Analysis

Factor analysis is a commonly used tool in data analysis. Its use is particularly prevalent in the analysis of responses to questionnaires. Nunnally (1978) reported that *Psychometrika* has devoted more space to articles based on factor analysis than any other multivariate technique.

The reason for this prevalence of use of factor analysis is that the general aim of factor analysis is to simply a data set by reducing an observed set of variables into a smaller set of underlying dimensions (factors) with the minimum loss of information (Hair, et al., 1995). Factor analysis offers the opportunity to reduce the overall data set to a more manageable size by analysing the structure of the interrelationships (i.e. the correlations) among the variables to determine common underlying dimensions, potentially easing the difficulties and reducing the time required in analysing the data set. However these benefits may be outweighed by methodological weaknesses in the technique in certain circumstances. Research presented in chapter 6 illustrates one such situation.

As with cluster analysis a key problem with factor analysis is that the computer programs available (e.g. SPSS) will always offer a factor structure. Thus, factor analysis is always a potential candidate for the "garbage in, garbage out" phenomenon (Hair, et al., 1995). One of the problems with the questionnaire used in this research is that it is not obvious that the questions form logically coherent groupings defining different dimensions of teaching effectiveness (i.e. factors). For this reason alone factor analysis results based on the student feedback data need to be treated with a degree of scepticism. A further reason for this scepticism is discussed in chapter 6 where the influence of clusters in the data set on factor analysis is examined.

In selecting the modules to be cluster analysed in chapter 7, emphasis was placed on selecting modules that had the greatest number of returned student feedback forms. This was important as adequate sample size is an important consideration in cluster analysis. It is recommended that factor analysis should not be performed on data sets containing fewer than 50 cases, data sets consisting of 100 or more cases are more preferable (Hair, et al., 1995). For any data set there should be at least five times as many cases as there are variables used in the analysis (preferably the ratio should be closer to 10:1), i.e. a factor analysis of 18 variables should ideally have a minimum of 90 cases (i.e. 90 returned student feedback forms). It is desirable to have the highest possible cases:variables ratio as this will reduce the chance of "overfitting" the data, i.e. producing factors that are sample specific with little generalizability.

As with cluster analysis, the approach to conducting a factor analysis can be broken down into seven steps. These seven steps (adapted from Hair, et al., 1995), reflect the major decisions that have to be made during factor analysis and the main checks that need to be performed. These checks are needed to ensure that the results obtained are meaningful.

1. The variables to be used in the analysis need to be selected.
2. Initial checks for structure in the data set need to be undertaken to ensure that the data are appropriate for factor analysis.
3. A factor method needs to be selected. The two most widely used methods are Common Factor Analysis (termed Principal Axis Factoring in SPSS) and Principal Components Analysis.
4. The number of factors needs to be determined. There are a large number of methods for deciding on the number of factors.
5. The method of rotation needs to be selected.
6. The rotated factor matrix needs to be interpreted to gauge which variables load on which factor and whether these loadings are meaningful, i.e. describe relationships that seem logical given the researcher's previous knowledge of the research and the literature in the area.
7. The factor solution needs to be validated. For example, the researcher could run factor analysis independently on split samples from a single data set and compare the factor structures.

Having selected the variables to be used in the analysis (in this research the questions on the Business School student feedback form) the next step (step 2) in conducting factor analysis is to check that there is sufficient structure in the data to make factor analysis appropriate. In chapter 6 the presence of sufficient structure in a data set was checked in three ways. Firstly, the KMO measure of sampling adequacy was examined. The criteria for interpreting the degree of structure in the data set depicted by the KMO value is presented in Table 6.4. Secondly, the overall correlation matrix was examined to check that a sufficient proportion of bivariate correlations were significant. Finally, the partial correlation matrix (i.e. the anti-image correlation matrix in SPSS) was checked to ensure that few of the bivariate partial correlations were significant. Another possible method of checking for sufficient structure in the data is Bartlett's test of sphericity. This method tests the null hypothesis that the variables in the population are uncorrelated. A significant result therefore suggests that there are non-zero correlations among (at least) some of the variables. One of the problems with Bartlett's test of sphericity is that it is very sensitive especially as the sample size increases. For this reason it was not used for checking for adequate structure in the data sets used in this research.

The next step (step 3) is to choose between the several factor analysis methods available. The two most common methods are Common Factor Analysis (termed Principal Axis Factoring in SPSS) and Principal Components Analysis. The main difference between these two methods is in the type of variance accounted for in each method and hence the values on the diagonal of the correlation matrix. The factors resulting from Principal Axis Factoring are based only on common variance (i.e. variance shared by all the variables in the data set). Hence a measure of common variance (i.e. the communality) is inserted on the diagonal of the correlation matrix for each variable. In Principal Components Analysis three types of variance are accounted for: common variance, unique variance (i.e. variance associated only with a specific variable) and error variance (i.e. variance due to measurement error, data gathering problems, or the effect of random components). Hence the value of unity is placed on the diagonal of the correlation matrix for each variable. There is considerable debate over which factor model is more appropriate (e.g. Snook and Gorsuch, 1989; Gorsuch, 1990; Mulaik, 1990).

Principal Axis Factoring is often recommended when the main objective is to identify the underlying dimensions represented in the original variables and the researcher has little knowledge about the amount of unique variance and error. Principal Components Analysis is appropriate when the analyst is primarily concerned about prediction or the minimum number of factors needed to account for the maximum proportion of the variance represented in the original set of variables (Hair, et al., 1995). Principal Axis Factoring does however, suffer from the problem of factor indeterminacy. The problem of factor indeterminacy is due to the fact that several different factor scores can be calculated for each respondent for any particular estimated factor model (Mulaik and McDonald, 1978).

Monte Carlo studies have, however, demonstrated that there is little difference in the results of Principal Components Analysis and Principal Axis Factoring (Velicer and Jackson, 1990). In most analyses the results will be nearly identical provided that the number of cases exceeds 30 (Gorsuch, 1983) and the communalities of most variables exceeds 0.6 (Hair, et al., 1995). Given the problem of factor indeterminacy with Principal Axis Factoring and the empirical evidence that there is little difference between the two methods, Principal Components Analysis was selected for use in this research.

The next step (step 4) in the factor analysis procedure is to decide the criterion for determining the number of factors to extract. There are several methods of factor extraction that a researcher can choose from. Two of the most commonly used methods are Kaiser's eigenvalue greater than 1 rule and the scree test.

Kaiser's eigenvalue greater than 1 rule (Kaiser, 1970, 1974, 1981) is sometimes referred to as the latent root criterion. Its premise is that any factor must explain at least as much of the variance as one of the variables in the analysis. As each variable contributes a value of 1 to the total of the eigenvalues, a factor which has an eigenvalue lower than one is discarded from the analysis.

The scree test is for use with Principal Components Analysis. Since the scree test aims to identify the number of factors that should be extracted before the amount of unique variance begins to dominate the common variance structure (Cattell, 1966). The researcher considers as factors all the components in the sharp descent of the scree plot *before* the first one on the line where the gradient is comparatively small. The logic of the scree plot is that the common variance is large in the first few factors and hence the eigenvalues are large. In the latter factors the common variance is essentially zero and hence the eigenvalues are small (i.e. essentially reflecting unique variance).

Stevens (1996) concluded that when the number of variables used in the analysis is fewer than 30 and the communalities are greater than 0.7, or when sample size exceeds 250 and the mean communality is greater than 0.59 the eigenvalue value greater than one rule is suitable. In situations where these criteria are not met the scree test would be more desirable *providing* that the sample size exceeds 200. Very few Business School modules have a sample size exceeding 200. Note that sample size in this context refers to the number of returned student feedback forms rather than the number of students enrolled on the module. For this reason the eigenvalue greater than one rule was normally used as the criterion for determining the number of factors to extract for this research.

Fava and Velicer (1996) conducted a simulation study on the consequences for a factor structure of either underextraction (i.e. extracting too few factors) and overextraction (i.e. extracting too many factors). Fava and Velicer found that the consequences of underextraction were particularly damaging to the results of factor analysis. This is a particular problem for the eigenvalue greater than one rule as the eigenvalue greater than one rule has a tendency to underextract, particularly in data sets that contain fewer than 20 variables (Hair, et al., 1995). The data sets factor analysed in this research consisted of between 17 and 36 variables. Particular care was therefore taken with the data sets consisting of either 17 or 18 variables, which represent the modules taught by a single lecturer. The scree test in contrast tends to result in *at least* one more factor being considered significant compared to the latent root criterion (Cattell, 1966).

It can be the case that there are eigenvalues *just* under 1 (i.e. between 0.95 and 0.99). The use of an arbitrary cut off rule (i.e. at the eigenvalue of one) may trigger a problem of underextraction. Care should be taken not to simply ignore factors with an eigenvalue just below one (nor to always extract factors with an eigenvalue just exceeding one). In marginal cases the Principal Components Analysis should be re-run using a different number of factors from that suggested by the eigenvalue greater than one rule and the results compared to see which factor solution produced the most coherent set of factors, i.e. the factor solution in which the variables loaded on to factors in a manner which allowed the factors to be logically named.

At this point the initial (unrotated) factor solution was examined to determine whether a single factor solution best fitted the data set (i.e. where all the variables loaded onto a single factor). Data sets that showed evidence of a multi-factor solution were then rotated.

In deciding which method of rotation to use (step 5) the first decision the researcher faces is between orthogonal rotation and oblique rotation. The basis of orthogonal rotation is that the axes of the factors are maintained at 90° . In other words, the factors are extracted in such a way that they are not correlated with each other. The second factor will be derived from the proportion of variance remaining after the first factor has been extracted. In contrast to orthogonal rotation which presumes that the factors are uncorrelated, oblique rotation allows the factors to be correlated and provides the actual correlations as output.

It is certainly the case that orthogonal rotation is more commonly used in published research than oblique rotation. Ford, MacCallum and Tait (1986), analysed 152 studies published in three particular psychological journals (over a ten year period). They found that of those that clearly expressed which rotation method was used 87% used the orthogonal method, with only 13% employing the oblique method.

One of the main problems with oblique rotation is that there is a danger with small samples or data sets that have a low cases:variables ratio that the results of an oblique rotation will be sample specific. Hair et al. (1995) suggest that the procedures for performing oblique rotations are not as well developed as those for orthogonal rotation and are still subject to considerable controversy.

There are several methods of orthogonal rotation. Two of the most common methods are QUARTIMAX and VARIMAX. The QUARTIMAX technique focuses on rotating the initial factor solution so that all the variables load highly onto one factor (i.e. a general factor) and in addition each variable loads highly onto one (and only one) other factor. In contrast, the VARIMAX technique focuses on rotating the initial factor solution so that each variable loads highly on one and only one factor. The VARIMAX technique tends to produce a mixture of some high loadings and some very low (i.e. close to zero) loadings in *each* column of the factor matrix. Therefore the VARIMAX method tends to produce factors that are more clearly separated.

Ford, MacCallum and Tait (1986) found that of the orthogonally rotated studies they analysed, 93% used the VARIMAX method of rotation. Kaiser (1970, 1974) used simulated data to demonstrate that the factor pattern obtained using VARIMAX rotation tended to more consistent than the alternative QUARTIMAX technique. In the light of the widespread use of the VARIMAX method and its methodological advantages, the VARIMAX method of rotation was used in this research.

Having rotated the principal components, the next step (step 6) is to interpret them. This was carried out by studying the factor loadings (i.e. the correlation between the original variable and a particular factor) on the rotated matrix. Hair, et al. (1995) offers guidelines on identifying significant factor loadings, in which the significance of a factor loading varies with the size of the data set. For example, for a data set consisting of 164 cases (as does the example presented in section 6.4.3), Hair et al. recommends that a factor loading should be considered significant if it is greater than 0.4 to 0.45. All variables with a factor loading greater than 0.4 were, therefore, considered when giving a name to the factor. As with the decision on eignvalues (discussed in step 4) care was taken however, to not arbitrary ignore variables that had factor loadings just under 0.4, or to automatically include variables that had factor loadings just above 0.4. The criteria that were used to decide whether variables with factor loadings close to 0.4 should be used to describe the factor were:

a) whether the variable enhanced the description the factor (i.e. fitted logically with the other variables that loaded on that factor);
b) whether the variable had a higher loading on another factor.

The final step (step 7) is to validate the factor structure. One of the key aims of the validation process is to check the extent to which a factor structure found in one data set is generalizable to the wider population. One method of validation is to analyse a new sample and compare how similar the two factor structures are. A second method is to run factor analysis independently on split samples from a single data set. In both these methods a similar factor structure in corresponding samples would be evidence supporting the validity of the factor structure. The split sample method was used to validate the degree of structure in the data set used in section 6.4. A set of ten random samples (each with one third of the number of cases in the entire data set) were computed. Factor analysis was run on each of these and the resulting KMO statistics were compared for consistency.

5.5.4 Multiple Regression

Multiple regression is a dependence technique that measures the relationship between a single metric dependent variable and several independent variables. The regression model takes the form:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \ldots\ldots\ldots\ldots b_pX_p + \varepsilon$$

Where: Y is the dependent variable

b_0 is the constant (which is also the Y axis intercept)

$X_1 - X_p$ are the predictor variables

$b_1 - b_p$ are the weights associated with each predictor variable. These weights denote the relative contribution of the predictor variable in explaining the variation in the dependent variable

ε is the residual, or prediction error

The objective of multiple regression is to model the changes in the dependent variable in terms of changes in the predictor variables (Hair, et al., 1995).

Carrying out multiple regression can be depicted as a six-step process:

1. The dependent and predictor variables need to be selected. These variables must arise out of the research problem.

2. If necessary, the data needs to be transformed to meet the assumptions of multiple regression. In this research several non-metric (categorical) variables needed to be transformed into dummy variables to allow them to be used in the regression analysis.
3. The method of selecting predictor variables to be included in the final model needs to be chosen.
4. The statistical significance of the regression model needs to be assessed.
5. The effect of the influential points need to be measured.
6. The residuals need to be studied to ensure that there are no obvious problems with the model. Influential cases need to be considered for removal from the model.

The first stage of the multiple regression process is to select the variables to be used in the analysis. The multiple regression analysis aimed to model lecturers' student feedback ratings against a range of predictor variables that have often been shown to influence lecturers' ratings.

The next step (step 2) of the multiple regression process is to decide whether the data needs to be transformed to meet the methodological assumptions of multiple regression. Such transformation was necessary for several of the variables used in this research. Many of the variables used in student feedback research are non-metric, for example the module level or the academic department or subject group to which the module belongs. In cases where the predictor variable is categorical it needs to be transformed into a dummy variable in order to be used in a regression analysis. The following example illustrates the point. There are four levels at which modules can be taken in the Business School, ranging from first year undergraduate to taught course postgraduate. To be able to use this data in a regression analysis the variable *level* has to be transformed into indicator variables. Four separate indicator variables were created, one for each level. For each of these new variables either a 0 or 1 was entered for each case depending on whether the module was taught at that level or not. All but one of the indicator variables for a non-metric variable are entered as predictor variables in the regression model, the omitted indicator variable is used as a reference level.

The next step (step 3) is to chose the methods with which to select the predictor variables to be included in the final model. Backward elimination was used as the method of selecting predictor variables to be included in the final model. In this method all the predictor variables are initially entered into the model and then predictor variables are removed one at a time. At each step, the predictor variable with the smallest partial correlation with the dependent variable is considered for removal. This variable is removed from the model if it is not significant at the 5% level. The new model is then fitted to the data and the new variable with the smallest partial correlation with the dependent variable is then considered for removal. The process continues until all of the variables in the model are significant at the 5% level.

As an alternative to backward elimination, forward selection could have been used as the method of selecting predictor variables to be included in the final model. In contrast to the backward elimination method which deletes variables in sequence, forward selection starts with no variables in the model and adds variables in sequence. Forward selection starts with a simple regression model with only one predictor variable, this being the predictor variable most highly correlated with the dependent variable. The partial correlation coefficients are then examined to determine which additional predictor variable explains the largest portion of the variation in the dependent variable remaining from the first regression equation. The regression model is then recomputed with both predictor variables and the new predictor variable tested for significance. This process is continued until none of remaining predictor variables makes a significant contribution to the model.

A third alternative method of selecting variables to be included in the model is stepwise estimation. This is very similar to forward selection except that at each step all predictor variables are tested for significance and non-significant variables are deleted form the model.

The main problem with the forward selection and the stepwise estimation procedures is that they only consider adding one variable at a time. In situations in which two predictor variables are not significant in themselves, but interact so that together they explain a significant proportion of the variance in the dependent variable, forward selection and stepwise estimation may not result in either variable appearing in the final model. For this reason backward elimination was used as the initial method of variable selection.

In any case, to check the results of the backward elimination method, the regression models were re-run using the forward selection method for a sample of the regression models. In each case the regression equations using the different methods were either the same, (i.e. with the same set of variables in the final model) or very similar with one or at most two variables that were present in one of the equations being absent in the other. These were always the variable(s) with the lowest t-value(s). In these cases there was little difference in the coefficients of the corresponding variables in the final model or in the R^2 values.

The next step (step 4) involves assessing the statistical significance of the regression model. This was achieved through an examination of the coefficient of determination (R^2), the F-value, the regression coefficients and their t-values. The researcher using one of the available statistical computer packages to run multiple regression (such as SPSS) has the opportunity to select the significance level at which an independent variable will be included in the model. For most of this research the significance level was set at 5%.

Steps 5 and 6 of the process involve the examination of the residuals and the effect that the leverage points have on the model. Residuals measure the difference between the observed and predicted values. They are therefore a measure of error (Hair, et al., 1995). A case was considered for removal if its standardised residual was large, i.e. greater than 3. This is equivalent to a value being more than three standard deviations from the mean. It is important to check for outliers as the aim is to model the vast bulk of the data without the results being substantially effected by one or two abnormal (i.e. outlying) cases.

There are several statistical measures of influence that can be used to detect cases that have a large effect on the model. The measure used in this research was the DFFIT scores for each case. This measures the impact that an individual case has on the overall model fit by assessing the degree to which the fitted values change when the case is deleted (Hair, et al., 1995). Two methods were used to assess whether any of the cases had a large influence on the model. The first was to use a rule of thumb proposed by Myers (1990); this rule considers a case to have a large influence on the model when its DFFIT value exceeds:

$$2p / n$$

where:
p is the number of predictor variables in the model.
n is the sample size.

The second method was to generate a box plot of the DFFIT scores, the cases identified as outliers by the box plot were also considered influential. In practice the Myers method defined more cases as influential than the box plot method. One of the problems with multiple regression is that the researcher may 'overfit' the model by deleting too many cases. This is a problem because the model can become too sample specific and therefore lack generalizability to the wider population. For this reason the box plot method was the predominant method used in deciding whether a case was influential and therefore should be removed from the analysis.

Multiple regression is an appropriate tool for the analysis of student feedback data because it provides a way of disentangling interrelationships between predictor variables. Multiple regression allows the effects of several predictor variables on the dependent variable to be displayed in a single model and the sizes of these effects to be computed. There are, however, alternative methods that could have been used to model the influences on lecturers' student feedback ratings. These alternative methods include:

a) (Multiple) analysis of variance;
b) Structural equation modelling;
c) Multilevel (hierarchical) modelling.

Each of these methods will be considered in turn.

a) Analysis of variance

Analysis of variance (ANOVA) is a dependence technique that measures the differences in a metric dependent variable based on a set of nonmetric predictor variables (Anderson, Sweeney and Williams, 1996). Multivariate analysis of variance (MANOVA) extends the ANOVA method to allow more than one dependent variable.

ANOVA and MANOVA are particularly suitable when used within an experimental design, i.e. a research design in which the researcher directly controls a predictor variable to determine its effect on the dependent variable(s). This type of analysis is used in student feedback studies that attempt to experimentally manipulate the students' experiences of lecturing in a laboratory, in an attempt to isolate the influence of a particular predictor variable on student feedback ratings (see for example, Abrami, Dickens, Perry and Leventhal, 1980; Marsh, 1984b).

In many situations ANOVA is an equivalent technique to multiple regression as standard ANAOVA models can be recast as regression models. ANOVA was not used in this research because there are a number of continuous predictor variables. MANOVA was not used in this research because there was interest in each of the dependent variables separately. It was expected that the predictor variables would influence the different aspects of lecturing (i.e. the different lecturer-specific questions) on the student feedback form in different ways. Therefore a different model was expected for each of the dependent variables.

b) Structural equation modelling

In contrast to multiple regression which aims to explain how an observable dependent variable depends on a number of manifest indicators (predictor variables), the aim of structural equation modelling is to explain the structure or pattern among a set of latent variables, each measured by one or more indicators. Latent variables are theoretical concepts and can only be measured indirectly by observable or measured variables (indicators).

Structural equation modelling combines two techniques, multiple regression and factor analysis (Hair et al., 1995). There are two features of structural equation modelling that distinguish it from multiple regression:

- the model can incorporate more than one dependent variable;
- the model can measure latent relationships (i.e. unobserved concepts) and account for measurement error in the estimation process. The ability of structural equation modelling to build on latent variables has the advantage of significantly increasing the sophistication of models that can be developed.

To conduct structural equation modelling the researcher needs to develop a theoretically based model, ensuring that all key predictor variables are included. The researcher then creates a path diagram of causal relationships. These are either exogenous (i.e. the predictor variables are not 'caused' or predicted by any other variables in the model) or endogenous (i.e. the predictor variables are predicted by one or more of the constructs). The next step is to convert the path diagram into a set of structural equations and specify the measurement model and the structural model. The quality of the resulting model can then be evaluated using a range of goodness-of-fit measures.

Structural equation modelling is not used in this research because each predictor variable is treated as being linked directly to the dependent variables and there are no latent constructs being considered. Although several dependent variables are considered in subsequent chapters the relationships between them are straightforward and there are no theoretical reasons for identifying any latent constructs. Since of the lecturer-specific questions only one question (Q17) might be conceived as depending on the others, the path diagram would be relatively trivial and could easily be investigated using multiple regression, there was considered no advantage in introducing the extra complexity of structural equation modelling.

c) Multilevel (hierarchical) modelling

The third alternative to multiple regression for modelling student feedback ratings is to use a hierarchical modelling technique. Hierarchical modelling is of particular use when there are multiple levels in the data set, with each level being nested in the level below. The hierarchical structure of student feedback data is shown in Figure 5.1.

Figure 5.1 A form of hierarchical structure of student feedback data

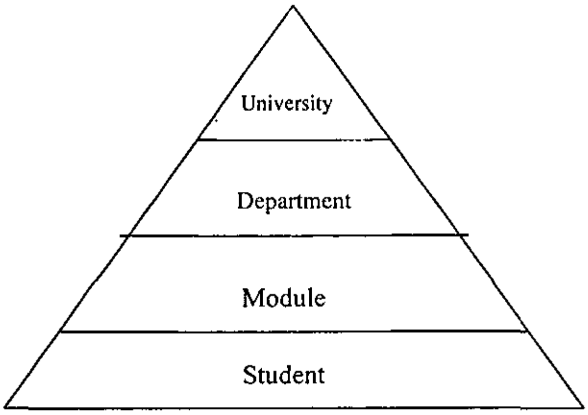


Figure 5.1 depicts the various layers of student feedback data. Student feedback forms are filled in by students for a particular module, which forms part of the teaching of a particular academic department at Loughborough University. In the terminology of multilevel modelling, Figure 5.1 has a hierarchical structure with the

student nested within a module which is nested within an academic department, which is nested within a particular university.

Hierarchical modelling is often used in education research (see for example, Goldstein and Spiegelhalter, 1996; Raudenbush, 1993). For example, using hierarchical modelling students can be considered to be nested in classes, which are nested within schools, which are themselves nested within a particular geographical area.

An example of a two level model is presented as equation 5.1. In this example (taken from Plewis, 1997) the two levels represent students and schools and models students reading attainment.

$$y_{ij} = b_0 + u_j + e_{ij} \qquad [5.1]$$

Where:

y_{ij} = reading attainment for student i, in school j
b_0 = overall mean
u_j = departure of school j from the overall mean
e_{ij} = residual term for students (random variation)

The aim being for the model to estimate:

- the overall mean (b_0)
- the between-school variance in students reading attainment (σ^2_u)
- the between-student within-school variance (σ^2_e)

The results of using this model will indicate the ratio of between school variance to total variance (often termed the intra-unit correlation).

Plewis (1997) argues that multilevel linear modelling has two particular strengths:

- by taking into account all the variability in the data (both between pupil and between schools), the standard errors of the regression coefficients will be correctly estimated. This corrects the tendency of the coefficients in single level regression equations to be overstated, in other words, for the standard errors to be underestimated;

- multilevel level modelling offers researches the possibility of being more imaginative theoretically and as data from each available levels can be analysed, the method should avoid the problems of aggregation and the ecological fallacy.

The ecological fallacy is the presumption that the relationships in the data at one level of the hierarchy is necessarily representative of the potential relationships present in the data at other levels in the hierarchy. At each level in the hierarchy there is likely to be variability in student responses.

One of the core ideas of hierarchical modelling is about students, classes and schools being random samples of all possible students, classes and schools. In standard multiple regression the effect of a particular school would be accounted for by using indicator variables and treating each school as a predictor variable in the equation, with all but one school separately entered into the regression model as an indicator variable. Therefore, if the data was drawn from five schools, four schools would be entered as indicator variables in the model. The effect of the school is then interpreted in terms of its similarity to one of the other schools used in the analysis.

In contrast, hierarchical modelling treats the school as a 'random effect'. This reflects the fact that the schools from which the data has been drawn are considered a random sample of all the schools from which data could have been drawn. In statistical terms, the implication of this is that the differences between schools are described in terms of the variance of the responses between schools.

Clearly hierarchical modelling would be very applicable for a broader study covering a number of departments at each of a number of universities. In this case the departments and universities could be considered as random samples from all departments and all universities. However, hierarchical modelling is not applicable for the data used in this research because, as discussed in section 5.2, the data used in this research consisted of the student feedback data collected in every Business School taught course module over a two academic year period. The modules used, therefore, are not a sample from a wider population, they constitute the entire population of modules over the time period in question. Therefore, part of the primary value of hierarchical modelling, i.e. the ability to account for 'random effects', would not be applicable in this research.

This research is principally interested in modelling the data at the module level. This reflects the fact that the differences between the particular modules used in the analysis are of interest. For example, there is a rational basis for expecting differences between students' evaluations of a first year economics module and a first year human resource management module. The research seeks to measure the influence of a range of predictor variables on the student feedback ratings lecturers receive. Another central aspect of the research involves the modelling of student feedback data to measure the effects on lecturers' ratings of the ratings of those with whom they teach on a particular module, this also requires the data to be analysed at the module level.

One way in which hierarchical modelling might have been relevant for the data used in this research would have been if there had been additional levels in the student feedback data, for example if tutorial groups formed a distinct and identifiable level in the hierarchy. Adding an additional level of tutorial group would acknowledge the fact that in many modules the students' educational experience is spilt between lectures and tutorials. The students' educational experience on a module will, therefore, be affected by the tutorial group to which they are assigned as well as the common lecture. In the data structure depicted in Figure 5.1, the between-tutorial group variance is subsumed in the between-student variance because data on tutorial group was not collected. Also, for many modules (particularly final year modules) there are no tutorials to support the lectures. Hence, the nature of the Business School student feedback data precludes this level of analysis, as there is no provision on the student feedback form for the student to indicate the tutorial group to which they belonged.

The same lack of data collection also precludes the analysis of variations in students' evaluations of teaching between students on the six different undergraduate degree programmes taught in the Business School. Most of the modules will be taken by students drawn from more than one Business School degree programme and the student feedback forms collect no information about which degree programme any individual student belongs to. This potential layer could, therefore, not be untangled from the available data.

Given the impossibility of untangling the potential levels of tutorial group and degree programme from the available data, the data used in this research lacks the multilevel structure that would make hierarchical modelling a useful method with which to analyse the data.

Chapter 6

SYSTEMATIC INTERRATER VARIANCE

6.1 Introduction

The increasing diversity within contemporary higher education both in terms of the students entering higher education and the variation in the learning styles utilised by students and the teaching styles employed by lecturers calls into question the practice of considering any single class a coherent homogeneous entity. In section 3.8 it was pointed out that the practice of considering a class as a homogeneous entity is a pervasive underlying assumption in the student feedback literature. This assumption is most clearly manifest in the practice of computing class average ratings and presenting these as a measure of the students' satisfaction with the module. This chapter considers whether there is a problem with using class averages and in particular whether the statistical technique of factor analysis, itself based on averaging ratings, is an appropriate statistical tool to use on data drawn from an environment as diverse as contemporary higher education.

Factor analysis is regularly utilised in student feedback research. The traditional approach to statistical analysis in student feedback research assumes that there is a true objective rating for a lecturer for each of the questions on a student feedback form. This implies that there is no systematic interrater variance, the variability that does exist between responses is merely random variation. In particular this chapter illustrates the effects on factor analysis that result from the existence of groups or clusters within the class, i.e. when the variance in responses has a systematic component.

The use of factor analysis in the student feedback literature is outlined in section 6.2 and in section 6.3 the rationale for treating the class as a homogeneous entity is analysed. The presence of systematic interrater variance in a Business School module is considered in section 6.4 and three discrete groups of respondents, each with different response patterns, are highlighted. It is shown that the cluster patterns are suggestive of different learning styles being utilised by students in the class. The section also illustrates the effect that the clusters have on the factor structure by showing how the strength of the factor structure, as indicated by the KMO statistic, significantly weakens when each of the three clusters are analysed in turn. Section 6.5 examines the methodological relationships between cluster and factor analysis and illustrates the interplay between the two. Simulation data is used to construct four scenarios, each with a different cluster pattern, to examine how varying cluster structures impact on factor structure; these findings are presented in section 6.6. The conclusion, section 6.7, calls for researchers to be more wary about relying on class average ratings as a measure of student satisfaction with the module. More emphasis should be placed on the investigation of the variation in responses within a class, particularly in how these variations may reflect different learning styles being employed by the students in the class.

6.2 Use of factor analysis in student feedback research

Factor analysis is one of the dominant statistical techniques used in student feedback research. Its use dates back to the origins of modern research in the field. H. H. Remmers (Smalzried and Remmers, 1943) published one of the earliest uses of factor analysis to analyse the results of student feedback. Remmers is described as the 'Father of Student Evaluation Research' by Marsh (Marsh, 1987). Remmers used factor analysis to extract two 'higher order' factors, which he termed empathy and professional maturity, from the original ten questions posed to students (Purdue Rating Scale, 1927). More recently influential authors such as Frey (Frey et al., 1975; Frey, 1978) and Marsh (1982), have used factor analysis in the construction and the analysis of their respective rating instruments. Both Marsh's SEEQ and Frey's

ENDEAVOR instrument have been replicated in various parts of the world (see Marsh and Roche, 1992; Watkins and Akande, 1992; Watkins and Gerong, 1992; Watkins and Regmi, 1992).

The use of factor analysis is continuing with the introduction into student feedback research of newer and more advanced covariance methods based on the ideas of factor analysis. In particular structural equation modelling (SEM) and similar techniques such as confirmatory factor analysis have been shown to have distinct applications to the discipline (Marsh 1991b; Stringer and Irwing, 1998). These techniques are an extension of multivariate techniques such as multiple regression and exploratory factor analysis. The main advantage of SEM compared to the other multivariate statistical techniques is its ability to examine a series of dependence relationships simultaneously (Hair, et al., 1995, p.617).

Factor analysis can be constructively used in the construction phase of the questionnaire, aiding in the selection of the instrument's questions from an item pool. The technique is also routinely used to analyse student feedback data collected in the classroom. This is not surprising as factor analysis acts as a data reduction technique providing an indication of which variables (or questions) are being answered in similar ways. Long questionnaires containing many questions can then be reduced into a small set of "composite" variables that are the amalgam of individual variables that have been shown to measure the same construct.

In this chapter the appropriateness of using factor analysis as a statistical tool for student feedback data is examined. The use of factor analysis is questioned in circumstances in which there are discrete groups (or clusters) of students in the class which each view the teaching differently, as a result of utilising different learning styles that differ in the type of teaching the student most desires from the lecturer.

6.3 The implications of averaging class responses

In chapter 3 it was pointed out that much of the published student feedback literature is based on the premise that it is appropriate to treat a class as a coherent homogeneous entity. This underlying assumption in the student feedback literature is most clearly manifested in the practice of computing class average ratings for a module and using these averages as the primary measure of student satisfaction with the module. Thus a single 'composite' rating is considered adequate to represent the view of the class. The degree of diversity, continually increasing in higher education, in terms of the characteristics of students entering higher education and the differences in both students' learning styles and lecturers' teaching styles (outlined in chapter 2) call into question the appropriateness of using class average ratings as the predominate measure of students satisfaction with the module.

As discussed in section 3.8 the assumption behind taking the class average for each variable is that 'there exists a true value on a given trait for the ratee which every rater, if he [or she] is not biased or unmotivated or careless or unobservant, will give the ratee' (Crichton and Doyle, 1975, quoted in Feldman, 1998). Crichton and Doyle continue 'this ignores the possibility that there may be a different "true" value for each student, for example, because the instructor satisfied his [or her] needs or desires with respect to the function named to a differing degree'.

In other words, variation in the individual ratings that the class gives the lecturer on each individual question on the evaluation form is assumed to be merely random variation. If however there are distinctive groups within the class (clusters of respondents) who *systematically* evaluate a lecturer differently on particular aspects of their performance, then the variation cannot be considered random and the validity of averaging responses is called into question. In section 6.4 an example of a module in which systematic interrater variance exists is presented and the groups of questions (or dimensions) that separate the clusters of respondents are pinpointed.

Crichton and Doyle (1975) are far from alone in recognising the possibility of clusters in a class (or 'true variance'). Feldman (1977) identifies the need for work on clusters and Marsh comments that:

> 'various subgroups of students within the same class may view teaching differently, and may be differently affected by the instruction which they receive, but there has been surprisingly little systematic research to examine this possibility' (Marsh, 1987, p. 277).

Despite these acknowledgements that clusters may well be present in data sets, the use of class averages is the norm. One of the main claims used to justify taking class averages of ratings is that individual idiosyncrasies and limitations tend to cancel out (Feldman, 1998). The evidence presented in this chapter and in chapter 2 regarding different learning styles being utilised by students, supports Feldman's belief that this claim is questionable. It is demonstrated that far from eliminating idiosyncrasies and limitations in a data set, the practice of taking class averages leads to the loss of potentially significant information. This loss could be acute if there are distinctive clusters of respondents (systematic interrater variance) in a class.

6.4 Systematic interrater variance in a Loughborough University Business School module

6.4.1 <u>The aim of the investigation</u>

To illustrate the existence of clusters of respondents, the ratings from a first year undergraduate module in the Business School has been selected as an example. This data was collected using the standard form described in chapter 4. This module was selected because the enrolment was large and not self-selecting in the sense that it was a compulsory module for the students. In total 164 students completed student feedback forms for this module, a response rate of 65.6%.

The aim was to use cluster analysis to examine whether there are distinctive groups of students who systematically judge the module and the lecturer's performance differently. If there are distinct groups then this would indicate that the differences in the class responses for each variable was not just random error, but included systematic interrater variance.

6.4.2 Results and interpretation of the cluster analysis

Most statistical computer packages use an agglomerative schedule to formulate clusters. As discussed in section 5.5.1, the principle behind agglomerative methods is that each observation begins in its own cluster. In each subsequent step, the two closest clusters are joined together to form a new aggregate cluster, thus reducing the number of clusters by one at each step. This process continues until all the observations are grouped together into a single all encompassing cluster. The consequence of this hierarchical procedure is that the results at an earlier stage of the agglomerative procedure are nested (or contained) within the results at a latter stage of the process (Hair, et al., 1995).

The gradual build-up of clusters generated by the agglomeration schedule is represented diagrammatically as a dendrogram. The dendrogram for the data set under consideration is presented as Figure 6.1. As mentioned in section 5.5.1, this was computed using the software package SPSS for Windows and used Ward's method with squared Euclidean distance, the recommended distance measure for this particular clustering algorithm (Hair, et al., 1995).

As a result of the sequential nature of cluster formation it is possible to argue that there are different numbers of clusters depicted in the dendrogram. The analyst has to make a judgement of how to partition the dendrogram. Partitions occur when the analyst selects one of the solutions in the nested sequence of clusterings that comprise the overall hierarchy (Everitt, 1993). A key determinant in the decision of how to partition the dendrogram is the degree of within cluster similarity deemed necessary.

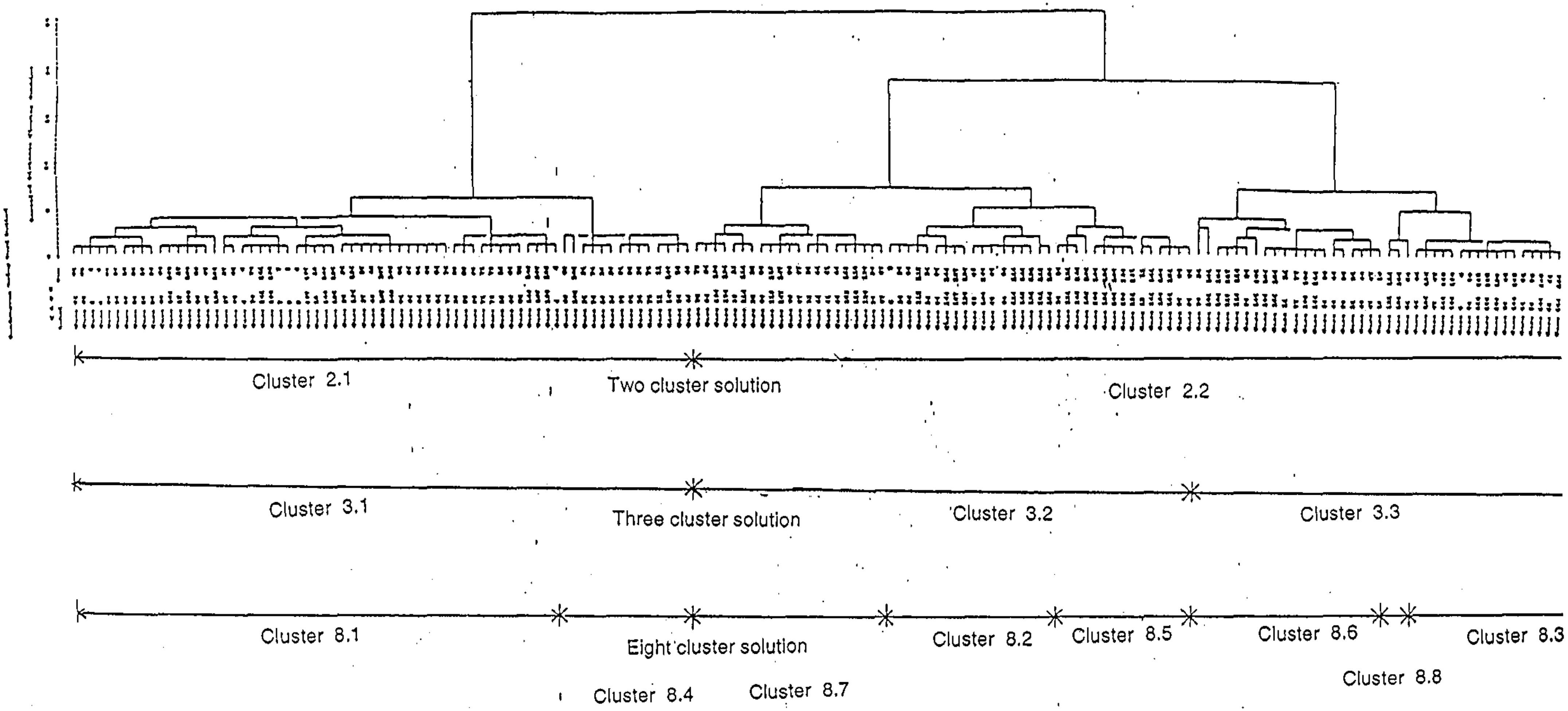
If the analyst wishes there to be *very high* degrees of inter-object similarities in each of the clusters it is possible to conclude that there are eight clusters in Figure 6.1. This highlights a fundamental trade-off in hierarchical cluster analysis. The consequence of maximising within cluster homogeneity, by partitioning the dendrogram into a large number of clusters, is that the between cluster heterogeneity will be small. If one examines the eight-cluster solution depicted in Figure 6.1, one notices that several of the clusters have very little between cluster heterogeneity. Two good examples of this are clusters 8.1 and 8.4 as well as 8.2 and 8.5. In both cases a small increase in the agglomeration coefficient, equivalent to a short line on the dendrogram, leads to the clusters being joined together.

Similarly if one selects the two cluster solution one has maximised the between cluster heterogeneity but has severely reduced the within cluster homogeneity. The within cluster variation in cluster 2.2 is very high, in that it consists of both clusters 3.2 and 3.3 in the three cluster solution, that have considerable between cluster heterogeneity.

Systematic interrater variance in the data set can be demonstrated with any meaningful cluster structure. As a result, the objection to using class averages remains unaffected by any ambiguity in the analyst's judgement of the number of clusters in the data set. A commonly used approach to determining the optimal number of clusters contained in the data set is to employ one of the available stopping rules. These are algorithms that attempt to calculate the most suitable partition of the data set. Monte Carlo analysis by Milligan and Cooper (1985) indicated that the best stopping rule was that devised by Calinski and Harabasz (1974). Applying this method to the data set suggests that either a two or a three-cluster solution represent an appropriate division of the data set. There was only a small difference between the Calinski and Harabasz statistic for the two and three cluster solution. For the analysis in this chapter the three-cluster solution was selected. The three-cluster solution was selected was for two reasons. Firstly, underestimating the number of clusters is a more serious problem than overestimating the number of clusters, as significant information will be lost if distinct clusters are merged (Milligan and Cooper, 1985).

Secondly, the dendrogram (Figure 6.1) suggests that the three cluster solution maintains a high degree of homogeneity within each cluster, whilst allowing the greatest degree of heterogeneity between the three clusters as indicated by the similarity scale that runs across the top of the dendrogram.

Figure 6.1 Dendrogram depicting three alternative cluster partitions within the full data set of 164 cases



Having detected a possible cluster structure, it is important to check that the groupings are meaningful. There is a danger when using cluster analysis that the method will produce an arbitrary division of the data set. Cluster analysis will, by its nature, always produce an outcome that indicates groupings even when the points form a continuum. In other words, clusters might simply be the highest scoring half and the lowest scoring half (or equally thirds) of a continuous range of scores. To explore the cluster structure, the means for each question in each cluster were calculated. These are displayed in Table 6.1.

Table 6.1 Cluster means

Question	Cluster 1	Cluster 2	Cluster 3	All cases
1	4.04	4.00	3.71	3.95
2	4.26	4.04	**3.38**	3.96
3	4.09	3.76	**2.88**	3.67
4	4.06	**3.37**	3.74	3.75
5	4.29	4.00	4.00	4.12
6	3.19	3.04	3.02	3.10
7	4.09	3.78	**2.86**	3.67
8	3.99	3.85	**3.05**	3.70
9	4.32	3.89	3.57	3.99
10	4.44	4.17	3.88	4.21
11	4.06	**2.39**	3.81	3.45
12	4.22	**3.24**	3.76	3.78
13	4.66	4.35	4.12	4.42
14	4.93	4.78	4.60	4.79
15	4.66	4.46	4.05	4.45
16	4.75	4.52	4.50	4.61
17	4.65	4.43	4.05	4.42
18	4.01	3.69	3.64	3.81
Overall	*4.26*	*3.88*	*3.70*	*3.99*
	n = 68	n = 54	n = 42	n = 164

These cluster means reveal that groups of students vary noticeably in their views of the module and the lecturer. The values in bold in Table 6.1 highlight several key variables that define cluster membership. In cluster 2 for example, the values for questions 4, 11 and 12 (especially question 11) provide evidence that there is not a simple continuum with values descending predictably from cluster 1 to cluster 3. The values for all these questions are *substantially* lower than the corresponding values for cluster 3. Cluster 3 is principally defined by four variables (questions 2, 3, 7 and 8) whose values, relative to the other two clusters, are significantly lower than the norm. This evidence shows that there is a meaningful pattern of interrater variance that precludes the possibility that variance in this data set is simply random variation, or that the clusters have been arbitrarily constructed.

The clusters can be characterised in the following way. Those in cluster 1 were distinctively very happy with the module and the instruction that they received. Those in cluster 2 were not happy with the assessment mechanisms, including the tutorials and coursework (questions 4, 11 and 12). However, overall they remain relatively happy, though clearly not as satisfied as those in cluster 1. Students in cluster 3 did not find the module stimulating or challenging and were critical of the pace of teaching (questions 2, 3, 7 and 8). However, these students were relatively happy about the assessment mechanisms (questions 4, 11 and 12).

The cluster partition in this module is highly congruent with the presence of different learning styles in the class. The students in cluster 2 are the most critical of the assessment mechanisms of the module, yet their ratings differ far less from the students in cluster 1 in terms of their evaluation of the intellectual aspects of the module (e.g. whether the module was challenging). This suggests that the students in cluster 2 are following a 'strategic' learning style, in which the student is primarily concerned with maximising their examination performance by pragmatically approaching the course, focusing their time and effort directly around the aspects of the module on which they will ultimately be evaluated. In contrast, students in cluster 3 are primarily critical of the academic elements of the module criticising the teaching for failing to challenge or stimulate them. Students in cluster 3 also criticise the teaching methods and the pace of teaching for not helping them to develop their understanding. It is possible that many of the students in cluster 3 were following a 'deep' learning style and found that the teaching failed to provide them with sufficient impetus to develop their appreciation of the material.

Criticism of the teaching methods by students in cluster 3 may suggest that the lecturer was following an 'instruction paradigm' in which the emphasis of the teaching was to impart knowledge to the student, rather than a 'learning paradigm' in which the emphasis is on the lecturer acting as a facilitator aiding the individual student develop their own understanding of the subject. Students utilising a 'deep' learning style are unlikely to appreciate lecturers who teach within an instruction paradigm as the emphasis of the lecturer will not match the aims and objectives of these students.

The module under consideration was taught to first year undergraduate students in the first semester of their university studies. Following the research outlined in section 2.5 it is not surprising that students following a 'deep' learning style are in the minority. It is expected that the majority of first year students will not follow a 'deep' learning style, but will rather utilise an approach more akin to the 'surface' learning style that is likely to resemble the style that they used during their 'A' level studies.

6.4.3 Results and interpretation of the factor analysis

Having demonstrated the existence of systematic interrater variance for a particular module, the investigation into the consequences of averaging class responses can now be broadened. In particular, the value of using factor analysis as a data reduction technique will be analysed. The first step was to run an exploratory factor analysis, to see whether the overall data set displayed any meaningful factor structure. As mentioned in section 5.5.2, the factor analysis was performed using SPSS and used Principal Components Analysis and Varimax rotation. Principal Components Analysis was selected to be consistent with Chang (1983) since Chang's analysis of the relationship between cluster and factor structure is presented in section 6.5. The resulting statistics suggest that a clear factor structure exists and the eigenvalues are presented as Table 6.2.

Table 6.2 Factor analysis of all 164 cases using Principal Components Analysis (with Varimax rotation).

Factor	Eigenvalue	Percent of variance	Cumulative variance
1	4.79	26.6	26.6
2	1.84	10.2	36.9
3	1.51	8.4	45.3
4	1.22	6.8	52.0
5	1.12	6.2	58.3
6	1.06	5.9	64.2
7	0.88	4.9	69.1
8	0.81	4.5	73.6
9	0.71	3.9	77.5
10	0.69	3.8	81.3
11	0.61	3.4	84.7
12	0.57	3.2	87.8
13	0.49	2.7	90.6
14	0.46	2.6	93.1
15	0.43	2.4	95.5
16	0.39	2.1	97.6
17	0.27	1.5	99.2
18	0.15	0.8	100

Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) value = 0.78.

Factor analysis was also carried out to examine the factor structure within each cluster independently. If the factor structure that was identified when analysing all 164 responses is a genuine representation of the data set, then a similar structure should be found in each of the clusters separately. Table 6.3 summarises the results of the factor analysis for each of the three clusters separately.

Table 6.3 Factor analysis within each cluster

Cluster	KMO	Maximum eigenvalue	Eigenvalues greater than 1
Cluster 1	0.54	3.61	7
Cluster 2	0.46	3.41	7
Cluster 3	0.51	3.58	7

The standard convention (see for example Kaiser and Rice, 1974 and Stewart, 1981) for interpreting the appropriateness of factor analysis as indicated by the KMO measure of sampling adequacy, is presented in Table 6.4. From the table it is clear that if a factor structure is present in the data set (i.e. if factor analysis is appropriate) then the KMO value should be *at least* 0.6.

Table 6.4 Calibration for Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) values

Value	Description
0.9 and above	Marvellous
0.8-0.89	Meritorious
0.7-0.79	Middling
0.6-0.69	Mediocre
0.5-0.59	Miserable
0.49 and below	Unacceptable

Reproduced from Kaiser and Rice, 1974; Stewart, 1981.

Using the entire sample, there is clear evidence to suggest that there is a factor structure within the data. In particular, the KMO measure of sampling adequacy is 0.78 and the maximum eigenvalue is 4.79. However using the data for each of the clusters the KMO statistics are 0.54, 0.45 and 0.51 strongly suggesting that there is no factor structure within any of the three clusters. This places a question mark over the use of factor analysis as a technique and the practice of averaging class responses, both traditional approaches in the student feedback literature. The results raise the distinct possibility that the existence of the three clusters (systematic interrater variance) generated an illusory factor structure in the data that took no account of systematic interrater variance present in the class ratings.

It is an established fact that the KMO measure of sampling adequacy has a degree of sensitivity to sample size (Kaiser, 1970, 1974, 1981). The possibility that this effect alone could account for the large fall in the KMO value in the three clusters compared to the overall data set needed to be examined. To examine this possibility SPSS was programmed to extract a set of ten random samples from the data set. The number of cases in each random sample was set as 55 which represents approximately one third of the total number of cases in the data set. The KMO value of each of these random samples was then calculated, these are reported in Table 6.5.

Table 6.5 KMO values of a set of ten random samples

	KMO value
Whole data set	**0.78**
Random Sample 1	0.67
Random Sample 2	0.61
Random Sample 3	0.68
Random Sample 4	0.69
Random Sample 5	0.64
Random Sample 6	0.57
Random Sample 7	0.69
Random Sample 8	0.62
Random Sample 9	0.64
Random Sample 10	0.68

Table 6.5 illustrates that each of the ten random samples has a KMO value lower than the KMO of the whole data set. In all but one of the samples the KMO value has remained above 0.6, a value considered to represent a degree of factor structure in the data (see Table 6.4). A comparison of the within cluster KMO values presented in Table 6.3 and the KMO values of the random samples presented in Table 6.5, reveals that the KMO values in the random samples are much higher than those in the clusters. This is despite the fact that all the random samples and clusters have been drawn from the same data set and that the sample size of the random samples and the clusters are approximately the same. From this it can be concluded that although the KMO value falls when sample size falls, sample size alone cannot account for the fall in the KMO values to 0.54, 0.45 and 0.51 in the three clusters respectively compared to 0.78, the KMO value for the whole data set.

Having presented an empirical illustration demonstrating the problems that systematic interrater variance has for factor analysis, a more theoretical consideration of the relationship between the presence of clusters and the (apparent) factor structure within a data set will be explored.

6.5 Methodological links between cluster structures and factor structures

For any given set of data there will be interplay between the cluster and factor structures. In essence the interplay between cluster and factor structures is explained by the distribution of points in multidimensional space. This is illustrated graphically in two-dimensional space in Figure 6.2.

Figure 6.2a A continuous model: An illustration of a continuous underlying factor structure

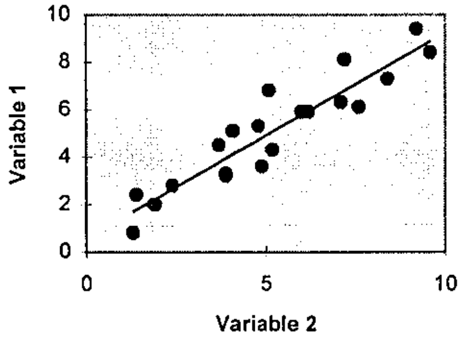
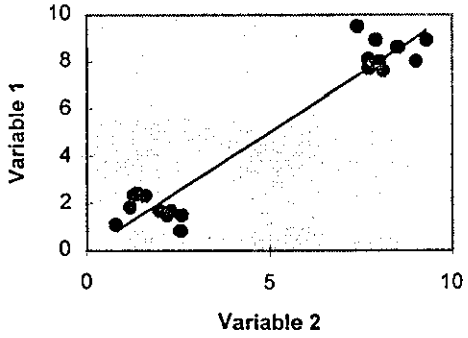


Figure 6.2b A discrete model: An illustration of a discrete underlying cluster structure



In Figure 6.2a the points are distributed throughout the two-dimensional space. In this situation one would expect to find a strong factor structure. In Figure 6.2b the points are distributed in a discrete fashion within the two-dimensional space, i.e. there are regions of high density separated by regions of low density.

Figure 6.2b depicts two clusters. One of the clusters can be characterised as having "high" scores and the other as having "low" scores. For a particular case, if the value for one particular variable (say variable 1) is high then the value for another variable (say variable 2) will also tend to be high because of the cluster structure. Similarly for a different case, if the value for one particular variable (say variable 1) is low then the value for another variable (say variable 2) will also tend to be low, again because of the cluster structure. As a result, the observed correlations between the variables will not be small and this will lead to a large value for the KMO measure of sampling adequacy. In other words, a superficial analysis of the data distributed in a fashion similar to that represented in Figure 6.2b would conclude that there was strong evidence of a factor structure. Scenarios 2, 3 and 4 in the simulation study described below consider this situation.

For the situation in which there are two clusters with several variables having the same variance-covariance matrix for each cluster, the effect on the correlations of the combined sample can be analysed theoretically as follows (Chang, 1983):

Let y be a k dimensional random variable, that is y consists of the different variables that are used to measure the various cases. Let a proportion p of the cases be from cluster 1 and let the means of the variables for cluster 1 be summarised as $\mu 1$. Let the remainder (proportion 1-p) of the cases be from cluster 2 and let the means of the variables for cluster 2 be summarised as $\mu 2$. Let the variances and covariances of the various variables within each cluster (i.e. for cluster 1 and for cluster 2 separately) be summarised as the matrix Σ. (Formally, y is a k dimensional random variable with a mixture of two normal distributions with means $\mu 1$ and $\mu 2$, mixing proportions of p and (1-p) respectively and a common covariance matrix Σ.)

Then the matrix summarising the variances and covariances of the mixture of variables from cluster 1 and cluster 2 together is given by (Chang, 1983):

$$V = p(1-p)dd' + \Sigma$$

where $d = \mu 1 - \mu 2$ and d' is the transpose of d.

In other words, the variances and covariances (or correlations) of the mixture of variables from the two clusters depends not only on the variances and covariances (or correlations) for each of the clusters separately but also on the proportion from each cluster (p) and the distance between the clusters (d).

In section 6.6 the effect of the clusters will be studied more generally using simulated data. In particular the effect of clusters on the KMO measure of sampling adequacy will be analysed.

6.6 Effects of clusters on factor analysis illustrated with simulated data

6.6.1 Construction of the simulated data

Presented below is a set of illustrations using simulation data. Four scenarios were created which are described in detail below. For each of the four scenarios, ten sets of data were created, each set of data consisting of 100 values for each of eighteen variables. Eighteen variables were used to be consistent with the number of variables on the questionnaire.

The random numbers were simulated using the NAG routine G05DDF, which creates normal random variates, based on a specified mean and standard deviation. The random variates were truncated so that the data used in the simulation study were simply the integers 1, 2, 3, 4 and 5 the same as the responses from the questionnaire. Each of the simulations used a different random number seed to give different sets of random numbers and hence the results from the simulations are independent.

Although the means vary according to the different scenarios, for example to simulate clusters, the (marginal) standard deviations are always the same. In brief, there is a cluster structure only in scenarios 2 to 4. In scenario 1 there is no cluster structure and there is no factor structure, hence scenario 1 provides a baseline. The factor analysis was carried out using SPSS for Windows. The cluster analysis was carried out in SPSS for Windows using Ward's method with Euclidean distance.

6.6.2 Scenario 1

The data used in scenario 1 had the same (marginal) mean and standard deviation as the responses from the student feedback forms, which are presented in Table 6.6. There should be no evidence of clusters or factors for this scenario. Table 6.7 presents the values for KMO, maximum eigenvalue and number of eigenvalues greater than one. Using the conventional rules for factor analysis (KMO under 0.5 being unacceptable, under 0.6 being miserable, Kaiser and Rice, 1974 and Stewart, 1981) there is no evidence of any factors which is consistent with the way in which the data were simulated.

Table 6.6 Means and standard deviations for responses to the eighteen questions, based on a five point scale

Variable	Mean	Std. deviation
1	3.95	0.68
2	3.96	0.64
3	3.67	0.81
4	3.75	0.73
5	4.12	0.62
6	3.10	0.67
7	3.67	0.78
8	3.70	0.84
9	3.99	0.75
10	4.21	0.59
11	3.45	0.96
12	3.78	0.84
13	4.42	0.58
14	4.79	0.44
15	4.45	0.62
16	4.61	0.56
17	4.42	0.64
18	3.81	0.71

Table 6.7 Summary of scenario 1

Simulation	KMO	Maximum eigenvalue	Eigenvalues greater than 1
1	0.48	1.76	8
2	0.47	1.78	8
3	0.44	1.63	9
4	0.45	1.76	8
5	0.40	1.76	9
6	0.47	1.84	10
7	0.46	2.01	8
8	0.45	1.87	8
9	0.48	1.84	8
10	0.45	1.82	9

6.6.3 Scenario 2

In scenario 2 there are two clusters in the data. The means for the first six variables differed by one standard deviation between the clusters, the remaining twelve variables had the same means. Many authors including Punj and Stewart (1983), comment that including even one or two spurious variables, that is variables which do not contribute to distinguishing between clusters, is likely to distort a cluster solution. Hence it is expected that there will be no evidence of clusters for this scenario. However, there will be some correlation between the variables and it is expected that the value of KMO will be increased.

The values for KMO, maximum eigenvalue and number of eigenvalues greater than one are given in Table 6.8. Using the conventional rules for factor analysis (see Table 6.4) there is no evidence of any factors which is consistent with the way in which the data were simulated. It should be noted though, that although there is only a weak cluster structure, this has had a small but statistically significant effect on the KMO measure of sampling adequacy. A two sample t-test shows that there is evidence, significant at the 5% level, that the mean KMO value for scenario 2 is greater than the mean KMO value for scenario 1 (t = 2.76, p = 0.013 with 18 degrees of freedom).

Table 6.8 Summary of scenario 2

Simulation	KMO	Maximum eigenvalue	Eigenvalues greater than 1
1	0.55	2.51	8
2	0.48	2.31	9
3	0.50	2.12	7
4	0.45	1.98	9
5	0.52	2.29	7
6	0.50	2.26	9
7	0.48	2.01	8
8	0.42	1.90	7
9	0.54	2.31	8
10	0.50	2.10	8

6.6.4 Scenario 3

In scenario 3 there are two clusters in the data. The means for all eighteen variables differed by one standard deviation between the clusters. Since all of the variables contribute to distinguishing between clusters it is expected that there will be clear evidence of clusters for this scenario. There will also be some correlation between the variables and it is expected that the value of KMO will be increased. The values for KMO, maximum eigenvalue and number of eigenvalues greater than one are given in Table 6.9. The KMO scores over 0.7 are indicative of a clear factor structure. The eigenvalue data suggests that this factor structure consists of six or seven distinct factors.

Table 6.9 Summary of scenario 3

Simulation	KMO	Maximum eigenvalue	Eigenvalues greater than 1
1	0.77	4.20	6
2	0.71	3.80	6
3	0.74	4.04	6
4	0.62	3.33	7
5	0.74	3.88	7
6	0.68	3.78	7
7	0.77	4.74	6
8	0.74	4.43	6
9	0.75	4.24	5
10	0.71	3.96	6

The way in which the data was simulated illustrates the interplay between the cluster and factor structures. Given the way the data was simulated the configuration of the points should resemble Figure 6.2b with the points distributed discretely in multidimensional space. The high value for the KMO statistic and the existence of six or seven factors in the apparently clear factor structure are simply the product of the correlations between cases because of the cluster structure.

6.6.5 Scenario 4

To illustrate this point further, scenario 4 presents a similar situation to scenario 3, but this time there are three clusters in the data. The means for all eighteen variables differed by 3/4 of a standard deviation between the first and second clusters and similarly between the second and third clusters. Hence, the means for all eighteen variables differed by one and a half standard deviations between the first and third clusters. Again because there will be some correlation between the variables, it is expected that the value of KMO will be increased. Taken at face value, the data presented in Table 6.10 indicates that there is an extremely clear factor structure in the data.

Table 6.10 Summary of scenario 4

Simulation	KMO	Maximum eigenvalue	Eigenvalues greater than 1
1	0.83	5.35	5
2	0.83	5.25	5
3	0.85	5.32	5
4	0.85	5.22	5
5	0.83	5.27	5
6	0.86	5.55	5
7	0.86	5.61	5
8	0.84	5.43	5
9	0.84	5.51	5
10	0.81	5.08	6

The presence of the three clusters will create a significant degree of correlation between the variables. It is this correlation that results in the high KMO statistics presented in Table 6.10. There is a clear risk that an analyst looking at these KMO values would conclude that there was a factor structure in the whole data set. If the possibility of clusters (i.e. that the points are discretely positioned in multidimensional space) is not considered then incorrect conclusions about the true structure in the data may be drawn.

6.7 Conclusions

One of the clearest manifestations of the assumption that a class can be treated as a homogeneous entity is the practice of computing the class average ratings and using these as the primary measure of students' satisfaction with the module. This practice is based on the premise that there exists a 'true' value for a lecturer on each of the questions on a rating form for a particular module. Differences between students are assumed to reflect misjudgements or individual bias on behalf of a student and therefore are considered to be random variation. By averaging class responses the impact of these inaccuracies is reduced. However averaging class responses also precludes the possibility that there is systematic interrater variance in the data, or to phrase it differently, that there are groups of students who are systematically viewing the course and the teaching differently (Crichton and Doyle, 1975).

The possibility of clusters of responses occurring in a class (systematic interrater variance) has previously been recognised (Feldman, 1977; Marsh, 1987). However, most of the empirical work in the field has proceeded on the basis that taking class averages of responses is valid. Evidence presented in section 6.4 illustrates that clusters exist in the Business School student feedback data. It was argued that the cluster patterns reveal differences in students' learning styles in that particular module.

Given the diversity of the students entering higher education (discussed in chapter 2) and the range of institutions offering undergraduate education, that differ in their missions, it is not surprising that there are distinct groups of students in Business School modules that take different approaches to their studies. Given that different learning styles are such a clear feature of contemporary higher education such clusters should be expected. In other words, it is unrealistic not to expect a significant degree of systematic interrater variance in students' evaluations of modules and lecturers.

Doubts as to the appropriateness of treating a class as a homogeneous entity (illustrated through the presence of systematic interrater variance) have significant implications for the use of factor analysis on student feedback data. Although factor analysis is a useful and important tool, it is recommended that researchers should always check first for clusters within their data set. If systematic interrater variance exists, then conclusions based on averaging class responses (e.g. factor analysis) are likely to be misleading. Conclusions based on factor analysis should only be considered reliable if the same factor structure is found in each of the clusters. The evidence in this chapter calls for researchers to place far more emphasis on the evaluation of systematic interrater variance. Researchers need to consider whether there are groups of students who come to significantly different judgements about the module and the performance of their lecturer as a result of entering the lecture theatre with different aims and objectives that derive from the employment of alternative learning styles. This chapter has shown the problems that exist when systematic interrater variance exists in Business School modules. The next chapter will explore to what extent these problems are manifested in the Business School student feedback data by examining the extent to which there are clusters of students with different learning styles present in Business School modules. The following chapter will also consider the implications of the presence of different learning styles in a class for lecturers and departmental managers.

Chapter 7

CLUSTER STRUCTURES IN THE DATA

7.1 Introduction

The previous chapter illustrated the potential problems with the practice of reporting the results of student feedback in the form of class averages, the standard practice in the Business School. Chapter 6 demonstrated that when there are groups of students in the class who systematically view the quality of the module and the skills of the lecturer differently, the reported class averages could be misleading. For example, a lecturer could receive satisfactory ratings in the reported class averages, indicating that there was no need for the lecturer to substantially change the module or their lecturing habits. However, their 'satisfactory' class averages could be masking a situation in which the teaching style of the lecturer was suited to the learning style of a proportion of the students (i.e. those who rated the skills of the lecturer clearly *above* satisfactory), but not suited to the learning style of other students (i.e. those who rated the skills of the lecturer clearly *below* satisfactory). In other words, contrary to the class average scores indicating that the module and lecturing were suited to the aims and objectives of the students, the teaching may not in reality have been particularly suited to a significant proportion of the students on the module. These differences between the usefulness of a particular lecturer's teaching method for different groups in the class should ideally be addressed by the lecturer(s).

The differences in students' perceptions could reflect the use of different learning styles being present in the class. There was some evidence presented in section 6.3.2 that suggested that this was the case in a particular Business School module. In this chapter the extent to which different learning styles are present in Business School modules is examined by investigating the extent to which clusters appear in the data collected on modules and lecturers in the Business School at Loughborough University. In section 7.2 the reasons for expecting some Business School modules to possess a cluster structure and some reasons for expecting some clusters explicit to the Business School are discussed. In section 7.3 an example of a cluster structure in a module taught by a single lecturer is presented. The cluster means are analysed and are shown to depict a cluster structure. The presence of a cluster structure is strengthened through analysis that contrasts the differences of the between and within-cluster correlations of the variables that are considered key in determining cluster structure. In section 7.4 the cluster patterns of a sample of 41 modules taught by a single lecturer are discussed. A clear pattern emerges which supports the presence of different learning styles as discussed in section 7.3. In a similar vein to the analysis of modules taught by a single lecturer, section 7.5 presents an example of a cluster structure in a module taught by a more than one lecturer and section 7.6 presents the cluster patterns of a sample of 24 modules taught by more than one lecturer. A clear pattern emerges which supports the presence of different teaching styles and a potential interaction between students' learning styles and lecturers' teaching styles. In section 7.7 the analysis of all the 65 modules is extended by using one of the most successful objective tests for determining whether clusters are present in the data and if there are what the most suitable partition of the data set is. The results of this 'stopping rule' support the presence of clusters in the Business School student feedback data, but suggest that the clusters are not highly differentiated. The general conclusions are presented in section 7.8.

7.2 Expected cluster patterns in the Business School data

The increasing number of students entering higher education over the last 40 years was discussed in chapter 2. Students in higher education are becoming increasingly diverse with an increasing proportion of female students, students from working class backgrounds and students from the ethnic minorities. The types of institutions offering undergraduate education are now more diverse than ever before, differing in their missions. Within this diversity there are alternative learning styles that students can utilise in the development of their understanding. As discussed in section 2.5 each of these learning styles involves the student approaching their studies with different aims and objectives. Students following different learning styles differ not only in their own study behaviour, but also in the type of teaching they desire from their lecturers. Similarly, as discussed in section 2.6, lecturers have been shown to vary in their teaching styles. Lecturers employing different teaching styles will enter the lecture theatre with different aims and objectives.

There was evidence in chapter 6 of different learning styles being employed by the students in a particular Business School module. Students in one of the clusters were suggested to be following a 'deep' learning style. These students were critical of the teaching methods and the pace of the teaching. In this chapter the frequency of which clusters of students who approach their studies with different learning styles is considered.

The different learning styles may be shaped, in part, by reasons specific to the context of the Business School modules. Some of the modules involve considerable statistical and mathematical context. The undergraduates entering the Business School are likely to vary in the subjects that they studied at ‘A’ level. Some students will have studied statistics or mathematics at ‘A’ level whilst others will have not. This difference in previous academic background may impact on how comfortable students feel with the material and this, in turn, may influence how they approach their learning for these modules. Similarly between 1/4 to 1/3 of the undergraduates enter the department having studied economics. On a more general level, the compulsory modules that all undergraduate students have to attend covers the full spectrum of sub-disciplines in the Business School. No one individual student is likely to feel equally comfortable with disciplines of such varying nature. Neither is any one particular learning style likely to be equally appropriate.

If there are clearly defined differences in students' learning styles in a class then these will be reflected in the cluster patterns for the module. The reason for this is that the teaching style of the lecturer will be more preferable to students following one learning style to those following another. For example, a student following a 'surface' learning style is likely to appreciate lecturers who lecture in a highly organised, structured manner, guiding the student through the material and providing the student with handouts of material and lecture summaries (i.e. a lecturer broadly following an 'instruction paradigm'). Lecturers employing the alternative 'learning paradigm' will place the focus on stimulating the student's interest in the subject material and in organising the module in a manner that challenges the student. A student following a 'surface' learning style is unlikely to appreciate (and rate highly) such a lecturer. Therefore, the root cause of clusters in a class is a mismatch between the lecturer's teaching style and the learning styles of some of the students on the module.

A cluster structure is also expected in some of the modules taught by more than one lecturer. The psychological concept of 'contrast effects' was discussed in section 2.6.4. A contrast effect in student feedback ratings implies that the rating a lecturer receives on a module may be influenced by the performance of the other lecturer(s) teaching on that particular module. The comparative judgement of lecturers' performance is likely to reflect different teaching styles being employed by lecturers and the extent to which these different styles interact with the learning styles being utilised by the students in the class.

The student feedback form used in the Business School has no provision for collecting information relating to student characteristics. Therefore it is not possible to explain clusters in terms of student characteristics such as previous academic experience. However, this analysis remains important given the implications for the practice of reporting student feedback ratings in the form of class averages, if different groups of students that are approaching their studies with different aims and objectives are present in the class.

7.3 An empirical example of cluster structure in a module taught by a single lecturer

7.3.1 Design of the cluster analysis

To explore the presence and nature of clusters in Business School modules, a particular module, A060~97 was selected for analysis. This module was selected for analysis as it has a large number of registered students and is compulsory, thereby eliminating any bias caused by student self selection. An example of student self selection bias would be students taking a module because, based on prior experience, they had a liking for the lecturer.

The actual student feedback forms for module A060~97 were examined and an attempt was made to fill in as many of the 'missing values' as possible. As discussed in chapter 4 many data values that are recorded as being missing are not actually missing they are simply unread. Getting the data set as complete as possible is important, as when computing cluster analysis SPSS can only process cases that are complete (i.e. cases that have no missing values).

Hierarchical cluster analysis was run in SPSS using Wards Method with squared Euclidean distance. The resulting dendrogram was analysed and a judgement of the number of clusters it depicted was taken. The analysis was then re-run, with SPSS set to save the cluster membership of each case using the number of clusters deemed most appropriate based on analysis of the dendrogram. The means for each variable in each cluster were then calculated and these were used in conjunction with the dendrogram to determine whether a cluster 'structure' was present and if so, which questions were key to determining cluster membership.

7.3.2 Examination of the dendrogram and cluster means for A060~97

Table 7.1 presents the cluster means for module A060~97.

Table 7.1 Cluster means for module A060~97

Question	Cluster 1	Cluster 2	Difference	All cases
1	4.32	3.84	0.48	4.11
2	**3.94**	**2.76**	**1.18**	**3.43**
3	3.48	3.00	0.48	3.28
4	4.03	3.70	0.33	3.89
5	4.03	4.04	-0.01	4.03
6	3.23	2.86	0.37	3.07
7	4.12	4.28	-0.16	4.19
8	**3.89**	**2.14**	**1.75**	**3.14**
9	3.94	3.48	0.46	3.74
10	4.18	3.76	0.42	4.00
11	3.89	3.66	0.23	3.79
12	2.91	3.34	-0.43	3.09
13	4.53	4.12	0.41	4.35
14	4.17	3.74	0.43	3.98
15	**4.11**	**2.60**	**1.51**	**3.46**
16	**3.65**	**2.54**	**1.11**	**3.17**
17	**4.15**	**3.00**	**1.15**	**3.66**
18	3.61	3.04	0.57	3.36
Overall	*3.90*	*3.33*	*0.57*	*3.65*
	n = 66	n = 50		n = 116

Table 7.1 reveals that five variables have appreciably greater differences in their means between the two clusters than the other variables. These are shown in bold in Table 7.1. For each of these five variables the difference in the mean scores between the clusters exceeds 1. Amongst the thirteen remaining variables, the next highest difference is 0.57.

The five variables that predominantly separate the clusters are questions:

2. The teaching methods for this module helped me understand the subject;
8. The subject material in this module was presented at about the right pace;
15. The lecturer explained the subject clearly;
16. The lecturer encouraged participation and questions;
17. The lecturer overall was effective.

For the cluster structure to be meaningful the variables that define cluster membership need to logically relate to each other.

The students in cluster 2 were particularly critical of the lecturer for each of the five questions. These students were critical of the teaching methods and the pace of teaching. They were also critical of the lecturer, rating the lecturer low on clarity, encouragement of participation and overall effectiveness. There is an interesting link between the cluster structure of this module and the cluster structure of the module presented in the previous chapter (section 6.4.2). In that module the most critical students returned low ratings for the teaching methods and the pace of teaching, but were *not* critical of the lecturer (i.e. students in both clusters rated the lecturer highly on each of the lecturer-specific questions). The inference from this is that the students in cluster 2 in module A060~97 are following a 'surface' learning style.

It is noticeable that 'teaching methods' (Q2) and 'pace of teaching' (Q8) are closely related to the lecturer's overall effectiveness (Q17) for these students. In other words, for these students there is not merely a problem with the teaching methods and the pace of teaching, but in the students' minds this is the fault of the lecturer. Students following a 'surface' learning style are more likely to be critical of a lecturer who employs a teaching style incompatible with their learning priorities compared to students following a 'deep' learning style. This is because students utilising a 'surface' learning style are more reliant on the lecturer compared to students following the more independent thinking 'deep' learning style. Therefore the style of a lecturer following a more challenging 'learning paradigm' form of teaching becomes an *actual problem* for 'surface' learners, this is likely to be reflected in low student feedback ratings.

The close link between lecturers' clarity (Q15) and the other five questions which distinguish the students in cluster 2 is also interesting. Clarity for these students may have less to do with the direct performance of the lecturer, but may in contrast reflect more about the interaction between the lecturer's teaching style and the student's learning style. Clarity therefore reflects how compatible the lecturer's style was with the students' learning style, reflected in the students aims and objectives as they entered the lecture theatre.

7.3.3 The application of between and within-cluster correlations

One of the major problems with cluster analysis is that the sorting ability of the cluster analysis algorithms is powerful enough to produce clusters even when there are no meaningful groupings embedded in the data set (Barney and Hoskisson, 1990).

One piece of evidence that is suggestive that the cluster solution is not merely an arbitrary division of a continuum is the evidence of the cluster means. The cluster means illustrate the mean for each variable in each of the clusters separately. This method is illustrated in Table 7.1. Table 7.1 reported that five variables had significantly greater differences in their means between the two clusters compared to the other 11 variables. These five variables were therefore considered to define cluster membership. The fact that the defining variables formed a logical explanation of the clusters is evidence that a meaningful cluster structure exists in this module.

A second piece of evidence that can be used to support the presence of cluster structure in the data set is the contrast between the bivariate correlations of the key variables in the data set as a whole and the bivariate correlations of the key variables within each cluster. The logic behind this approach can be illustrated though four idealised scatterplots.

Take for example Figures 7.1, 7.2, 7.3 and 7.4. Each of these figures presents a two dimensional image that illustrates different combinations of bivariate correlation and cluster membership. Figure 7.1 illustrates a situation in which there is both cluster structure and high overall correlation. The variable 1 - variable 2 (X-Y) correlation is strong and one would expect a high value for R^2, indicating that there was a strong positive correlation between the two variables. With this high X-Y correlation there are two distinct clusters, one of which is scoring low and the other high on both dimensions. What is most interesting in this scenario is that although there is high correlation between the two variables in the data set as a whole, the correlation between the two variables within each of the clusters is very low. The value of R^2 for the nine points in the top right quadrant of Figure 7.1 would be close to 0.

Figures 7.2 and 7.3 both display a high correlation between the two variables. They differ in the presence of clusters. Figure 7.2 displays a strong two cluster structure, as in Figure 7.1 one cluster consists of high scores for both of the variables and the other cluster represents the converse - low scores for both variables. The cluster pattern in Figure 7.2 can be contrasted with that in Figure 7.1 in terms of the degree of within cluster correlation. In Figure 7.2 there is a high within cluster correlation between variable 1 and variable 2. The within cluster correlation of the two variables matches that of the entire data set.

In the three models discussed so far a cluster structure is present in two. In both cases there is a high degree of correlation within the data set as a whole. It would be erroneous to assume that such a correlation between the variables in the data set was a prerequisite for there to be a cluster structure. Figure 7.4 illustrates a situation in which there is a clear cluster structure but a zero correlation between the two variables in the data set as a whole. The within cluster correlation is high.

Figure 7.1
Scatterplot illustrating high overall correlation and two clusters that have low within-cluster correlation

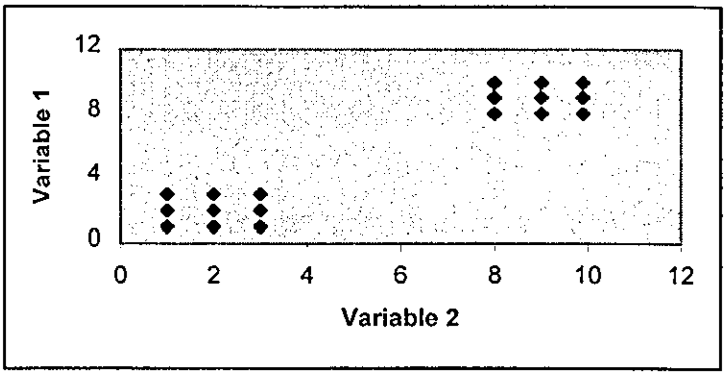


Figure 7.2
Scatterplot illustrating high overall correlation and two clusters that have high within-cluster correlation

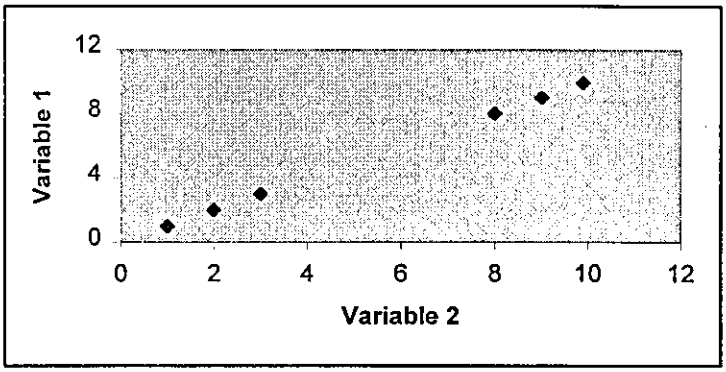


Figure 7.3
Scatterplot illustrating high overall correlation without clusters

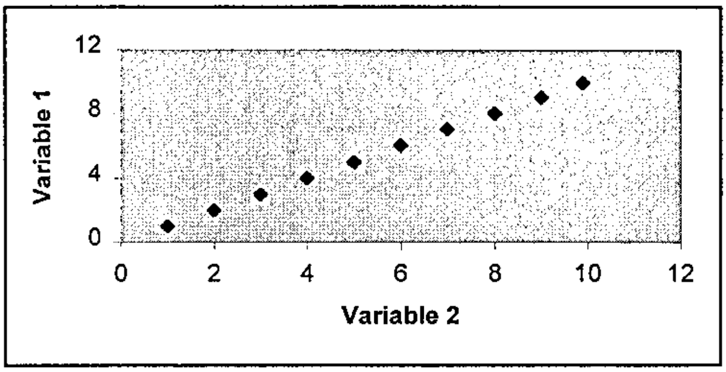
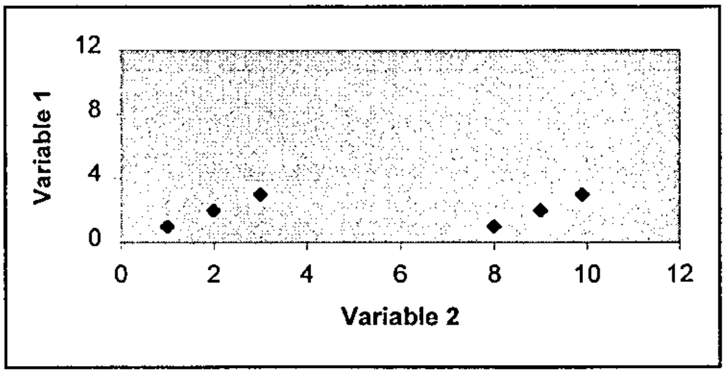


Figure 7.4
Scatterplot illustrating low overall correlation but with two clusters that have high within-cluster correlation



With these idealised models in mind we can examine the overall correlations and the within cluster correlations of the variables that compose the key variables in module A060~97. Table 7.2 presents the bivariate (Pearson Product Moment) correlation coefficients for the five variables over the data set as a whole. Correlations shown in bold are significant at the 1% level. The most cursory glance at Table 7.2 tells the reader that there is a strong degree of correlation in the whole data set. The correlations range between 0.438 and 0.784.

Table 7.2 Correlation matrix for key variables in module A060~97

	Q2	Q8	Q15	Q16	Q17
Q2	1	**0.622**	**0.68**	**0.438**	**0.659**
Q8	**0.622**	1	**0.601**	**0.439**	**0.567**
Q15	**0.68**	**0.601**	1	**0.54**	**0.784**
Q16	**0.438**	**0.439**	**0.54**	1	**0.571**
Q17	**0.659**	**0.567**	**0.784**	**0.571**	1

These empirical correlations would suggest that of the four models (Figures 7.1 to 7.4) either Figure 7.1, Figure 7.2 (which both depict a strong overall correlation and a strong cluster structure), or Figure 7.3 (which depicts a strong overall correlation but no cluster structure) could depict this data set. In other words the correlational information so far does not provide guidance as to whether a cluster structure is present in the data set. To establish whether a cluster structure is present it is necessary to examine the correlations of the same variables within each of the two clusters separately. These are presented in Tables 7.3 and 7.4.

Table 7.3 Correlation matrix for key variables in module A060~97 within cluster 1

	Q2	Q8	Q15	Q16	Q17
Q2	1	**0.336**	0.276	0.149	**0.436**
Q8	**0.336**	1	0.088	-0.148	0.129
Q15	0.276	0.088	1	0.303	**0.532**
Q16	0.149	-0.148	0.303	1	**0.38**
Q17	**0.436**	0.129	**0.532**	**0.38**	1

Table 7.4 Correlation matrix for key variables in module A060~97 within cluster 2

	Q2	Q8	Q15	Q16	Q17
Q2	1	0.091	0.357	0.131	0.328
Q8	0.091	1	-0.076	0.216	0.032
Q15	0.357	-0.076	1	0.238	**0.568**
Q16	0.131	0.216	0.238	1	0.341
Q17	0.328	0.032	**0.568**	0.341	1

A quick glance clearly indicates that the degree of correlation in Tables 7.3 and 7.4 is far lower than those in Table 7.2. In Table 7.3 only four bivariate combinations (Q2, Q8; Q2, Q17; Q15, Q17; Q16, Q17) and in Table 7.4 only one bivariate combination (Q15, Q17) are significant to the 1% level. Take for example Q8. In the overall data set it correlated very strongly with each of the other four variables: 0.622, 0.601, 0.439 and 0.567 respectively. These strong correlations disappear when Q8 is correlated with the other four variables in the two clusters separately. In cluster 1 they correspond to: 0.336, 0.088, -0.148 and 0.129 while in cluster 2: 0.091, -0.076, 0.216 and 0.032.

Having examined the within cluster correlations we can return to the theoretical correlational models previously discussed. Given that the evidence presented in Tables 7.2, 7.3 and 7.4 illustrate a high overall correlation between the five key variables determining cluster membership in module A060~97, but a low within correlation between these same variables, Figure 7.2 and Figure 7.3 can be dismissed as representing the structure of this data set. In contrast Figure 7.1 does meet the core characteristics of this data set depicting a pair of discrete clusters. This does not necessarily imply that the distribution of points of the key questions in A060~97 actually resembles the distribution illustrated in Figure 7.1. Evidence presented in section 7.7 illustrates that the two clusters are likely to be less clearly differentiated than Figure 7.1 with a degree of overlap at the margins. The correlations do though strengthen the judgement initially derived from the visual inspection of the dendrogram and the examination of the cluster means that a cluster structure exists in module A060~97.

7.4 General pattern of cluster structure for modules taught by a single lecturer

Having detected evidence of different learning styles amongst the students in the class leading to a cluster structure in one Business School module it is now important to examine the extent to which evidence of different learning styles is found in other modules taught by a single lecturer.

In total 64 further modules were cluster analysed. These were the same 65 modules used to analyse the extent of missing values in the data set presented in section 4.4. Each of the modules have between 17 and 30 variables (depending primarily on the number of lecturers who taught on the module). The modules were selected on the basis of being the larger modules in terms of number of registered students. Several of the largest modules were not selected as they had very low response rates. The problem with low response rates is the question mark that this raises over the generalizability of the results based on the sample.

As with module A060~97, the actual student feedback forms were examined and an attempt was made to fill in as many of the 'missing values' as possible. Given that many of the modules had numbers of responses that were close to the realistic minimum number of cases suitable in a cluster analysis, it was important to maximise the number of valid (i.e. complete) cases in each module. The actual numbers of responses for a particular module were important as conventional guidelines on data set size for conducting cluster analysis is that the technique should not be performed on data sets containing fewer than 50 cases. A minimum of 100 cases is recommended by some researchers (Hair, et al., 1995). The cluster analysis method followed that outlined in section 7.3.1.

Given the expectations outlined in section 7.2 regarding the potential of contrast effects between lecturers in modules in which the lecturing responsibilities are shared, modules taught by a single lecturer are reported separately from modules taught by either two or three lecturers. In this section modules taught by a single lecturer are discussed. Modules taught by either two or three lecturers are discussed in sections 6.6 and 6.7.

In the overall sample of 65 modules 41 were taught by a single lecturer. There was no evidence of a cluster structure in the responses for 24 (59%) of these 41 modules. Table 7.1 reports on the remaining 17 modules (41%) where the responses were considered to possess a cluster structure, by virtue of several variables having appreciably greater difference in their means between the clusters compared to the other variables.

Table 7.5 Modules taught by a single lecturer possessing a cluster structure

Code	Reg.Studs	#cases	Any Clusters	Key Variables	No Difference	Cluster sizes
A052~96	147	119	TWO	2 3 15 17	4 6 11	45/72
B070~96	188	133	TWO	2 3 8 15 17		110/23
A030~97	100	69	THREE	2 3 13 15 17		10/46/13
A035~97	61	50	TWO	2 3 17 18		12/38
A060~97	207	116	TWO	2 8 15 16 17	5	66/50
A100~97	122	62	TWO	3 7 8		16/33/13
A530~97	85	54	TWO	10 13	9 18	32/21
B005~97	82	61	TWO	2 8 15 17	4 6 9 18	29/32
B028~97	65	56	THREE	2 3 5 8 10 15 17		14/23/19
B105~97	158	75	TWO	2 3 7 10 12-17		42/33
B510~97	95	73	THREE	10 13 15		44/22/7
B555~97	59	43	TWO	2 3 17 18		23/20
B560~97	78	53	TWO	2 3 15 16 17		11/42
C022~97	66	55	THREE	5 6	10 13 14 16 17	22/13/20
C110~97	110	78	THREE	2 3 7 8 16 17		26/13/39
C190~97	73	63	TWO	2 3 7 8 10 13-17		48/15
C565~97	67	52	TWO	2 3 13 15 17	4 18	24/28

The second and third columns in Table 7.5 present the number of registered students and the number of forms (cases) that had no missing values and therefore could be included in the cluster analysis. The fourth column reports on the number of clusters that seemed appropriate from a visual inspection of the dendrogram. The variables that were the most important in differentiating between the clusters are listed in the fifth column. The variables listed in the sixth column are those questions whose means were very similar between the clusters and hence played no part in differentiating between the clusters. The number of respondents in each of the clusters is reported in the final column. A table showing the frequency with which each question played a key role in determining between cluster membership is presented in Table 7.6.

Table 7.6 Frequency with which each question is key in determining cluster membership in the sample of modules taught by a single lecturer

Question	Frequency
1	0
2	13
3	12
4	0
5	2
6	1
7	4
8	7
9	0
10	5
11	0
12	1
13	6
14	2
15	11
16	5
17	13
18	2

Table 7.6 illustrates the frequency with which each of the eighteen variables were 'key variables' in determining cluster membership. Key variables are defined as those variables that have a particularly large difference in their means between clusters. From Table 7.6 it can be seen that questions 2, 3, 15 and 17 are commonly key in determining cluster membership. These questions relate to the understanding gained by the student on the module and the lecturer's ability at conveying the material. Specifically the questions cover the extent to which the teaching helped the student develop their understanding of the subject, how stimulating the student found the academic content, how clearly the lecturer explained the material and the lecturer's overall effectiveness. As has been illustrated in sections 6.4.2 and 7.3.2 these questions on the student feedback form relate closely to alternative learning styles. It is not surprising that the overall measure of lecturing effectiveness (Q17) is related to Q2 and Q3 as these questions are so closely reflect differences in students learning styles. Students' judgements about clarity (Q15) is likely to directly relate to how well the lecturer satisfies the students aims and objectives, which are derived from their learning style.

Questions 7, 8, 10, 13 and 16 are moderately common in determining cluster membership. These questions relate mainly to the content of the module (was it challenging and was it related to business situations) and the degree to which the lecturer was well prepared and the module well organised. Students following a 'surface' learning style are likely to favour a lecturer who presents the material in a highly structured manner and who provides the student with learning aids such as, printed handouts and lecture summaries.

Questions 5, 6, 12, 14 and 18 are relatively uncommon in determining cluster membership. As discussed in section 4.2.2, Q5 (quality of teaching rooms) and Q6 (library support) are not well prepared questions in which the students responses are not expected to necessarily yield useful information. The presence of Q12 (tutorials) in this category is a little misleading. This question can only be asked in modules in which tutorials are part of the module. In only 7 of the 17 single lecturer modules sampled are the students able to respond to Q12. Of the six lecturer-specific questions on the student feedback form, the two lecturer-specific questions that are relatively uncommon in determining cluster membership (Q14 and Q18) are the two questions that have the least impact on students' learning styles. Although an enthusiastic lecturer is considered more desirable than a lecturer who lacks enthusiasm (as it is considered a motivational benefit for the student) the lecturer's enthusiasm is far less related to the students aims and objectives when they enter the lecture theatre, rather than, for example, the extent of lecturers' preparation.

Questions 1, 4, 9 and 11 are never key in determining cluster membership. The fact that Q1 and Q4 have little impact in determining cluster membership is to be expected given the wording of the questions, as discussed in chapter 4. Both these questions broadly ask the students about events such as being made aware of the module's aims and objectives. As it is Business School policy to distribute such information to every student on each module, all students on a particular module should have the same experience and therefore rate the lecturer similarly. One would also expect Q9 that asks the students to judge the extent to which any module is related to business situations to have little differential ability within a class. Some modules will by their nature be more applied (e.g. some marketing modules) whereas other modules will by their nature by more theoretical (e.g. some quantitative modules). For this reason one might postulate that there would be a high degree of between-module variation for Q9, but little within-module variation. Q11 asks students whether the coursework supported the modules aims and objectives. Once again, one would expect the students in the class to respond to this question in a similar way, either the coursework did or did not support the modules aims and objectives. Furthermore, questions 1, 4 and 9 do not reflect differences in learning styles utilised by students. Therefore there is no reason to expect students following different learning styles, to differ in the ratings they give on these questions.

The wider range of questions that regularly appear in the fifth column of Table 7.5 are particularly relevant to differences in learning styles being utilised by students. Specifically they all refer to how well the lecturer conveys knowledge to the students, how successful they are in developing the students interest in their subject and the extent to which the lecturer's teaching helps the student in developing their understanding of the subject. The fact that this pattern regularly repeats itself is encouraging as one way of validating a cluster structure is to see if a corresponding pattern is apparent in other data sets. As all the modules are evaluated on the same rating instrument the feedback results on each module can be considered separate samples. Separate samples producing similar structure suggest generalizability and validity of the cluster structure (Hair, et al., 1995).

Table 7.5 shows that in almost all the modules taught by a single lecturer in which Q15 and Q17 were key variables in determining cluster membership, Q2 and Q3 were also key. This suggests that Q2 (teaching methods helped me understand) and Q3 (academic content was stimulating) are linked in the student's minds to the lecturer. This is an issue with the structure of the student feedback form as Q1-Q12 are intended to be module-specific, whereas Q13-Q18 are intended to be lecturer-specific. Evidence in Table 7.5 suggests that this distinction is not reflected in the answering patterns of the students, who are seemingly considering the skills of the lecturer when answering some of the module-specific questions.

There is another interesting point that can be drawn from Table 7.5. Two of the modules A530~97 and B510~97 had a slightly different cluster structure to most of the other modules. In both of these modules the clusters were differentiated by differences in the responses to teaching organisation (Q10), lecturers' preparation (Q13) and lecturers' clarity (Q15). The same lecturer taught both of these modules, which is interesting as it suggests some between-module consistency in the student's judgements of the lecturing strengths and weaknesses of an individual lecturer.

7.5 An example of cluster structure in a module taught by more than one lecturer

As mentioned in section 7.2 some of the modules taught by more than one lecturer are expected to have a cluster structure as a result of contrast effects between the lecturers teaching on the module. Such a contrast effect may reflect different teaching styles being employed by the lecturers on the module. The extent to which lecturers' teaching styles are congruent with the students' aims and objectives as they enter the lecture theatre (themselves a reflection on the students' learning style) is likely to be reflected in lecturers' student feedback ratings. Modules taught by more than one lecturer are less likely to show a clear relationship between the module-specific questions (particularly Q2 and Q3) and the lecturer-specific questions (particularly Q15 and Q17). This reflects the fact that the students' responses for the module-specific questions will be based on their experience of more than one lecturer. Therefore it will be more difficult to distinguish the impact that a lecturer has on the student feedback ratings for the module-specific questions in the modules taught by more than one lecturer.

To examine this possibility a sample module was chosen for analysis. The cluster centroids for the selected module (B530~97) are presented in Table 7.7.

Table 7.7 Cluster means for module B530~97

Question	Cluster 1	Cluster 2	Difference	All cases
1	4.15	3.58	0.57	3.82
2	4.04	3.26	0.78	3.58
3	3.56	2.74	0.82	3.08
4	4.11	3.79	0.32	3.92
5	4.19	3.79	0.40	3.95
6	3.56	3.16	0.40	3.32
7	4.07	3.71	0.36	3.86
8	4.04	3.18	0.86	3.54
9	4.07	3.82	0.25	3.92
10	4.15	3.39	0.76	3.71
11	4.26	3.71	0.55	3.94
13 L1	4.48	4.34	0.14	4.40
14 L1	4.37	4.11	0.26	4.22
15 L1	4.26	3.95	0.31	4.08
16 L1	4.22	3.87	0.35	4.02
17 L1	4.15	3.82	0.33	3.95
18 L1	3.78	3.39	0.39	3.55
13 L2	**4.22**	**3.26**	**0.96**	**3.66**
14 L2	**4.30**	**3.39**	**0.91**	**3.77**
15 L2	**4.04**	**2.45**	**1.59**	**3.11**
16 L2	**4.30**	**3.11**	**1.19**	**3.60**
17 L2	**3.93**	**2.29**	**1.64**	**2.97**
18 L2	**4.00**	**3.16**	**0.84**	**3.51**
Overall	*4.10*	*3.45*	*0.65*	*3.72*
	n = 27	n = 38		N = 65

Examination of Table 7.7 shows that the clusters are principally defined by the ratings the students gave to the two lecturers. In cluster 1 there is very little difference between the ratings that the students gave the two lecturers. The biggest variation is for Q13, this is only 0.26. Cluster 2, in contrast, differentiates strongly between the ratings they gave the lecturers. Three of the variables Q13, Q15 and Q17 have differences exceeding 1, with Q14 and Q16 having very large variations of 0.72 and 0.71 respectively. This indicates that the students are clearly differentiated by the ratings they gave lecturer 2. The students in cluster 2 are less happy generally with the module and the lecturers. They clearly were not impressed with lecturer 2 who scores relatively well with the respondents in cluster 1. One explanation for this cluster structure is that lecturer 2 may have had a teaching style that more clearly polarised the students; a teaching style that particularly did not fit with the learning style of students in cluster 2. It is noticeable that other than the lecturer-specific questions the questions with the next four highest differences between the two clusters all related to lecturing style (i.e. questions 2, 3, 8 and 10).

The two clusters are of similar size, so the effect could not be explained by several students having an irrational dislike of lecturer 2. If cluster 2 consisted of a small number of respondents these individuals could be classed as outliers and the analyst would have to be cautious in the interpretation of the cluster structure. A relatively even split in the size of cluster membership encourages confidence that the cluster structure is a meaningful division of the classes perception of the quality of the lecturing on the module. The cluster structure in this module cannot be explained by lecturer 2 being a probationary or guest lecturer lacking classroom experience. Both of the lecturers teaching on this module are experienced, regular members of faculty.

Following the methodology used in section 7.3.3 the between and within-cluster correlations for module B530~97 were examined. Table 7.8 presents the bivariate (Pearson Product Moment) correlation coefficients for the six lecturer-specific questions for lecturer 2 over the data set as a whole. Correlations shown in bold are significant at the 1% level.

Table 7.8 Correlation matrix for the six lecturer-specific questions for lecturer 2 in module B530~97

	L2 Q13	L2 Q14	L2 Q15	L2 Q16	L2 Q17	L2 Q18
L2 Q13	1	**.530**	**.694**	**.475**	**.689**	.309
L2 Q14	**.530**	1	**.630**	**.627**	**.664**	.279
L2 Q15	**.694**	**.630**	1	**.663**	**.841**	**.393**
L2 Q16	**.475**	**.627**	**.663**	1	**.622**	**.363**
L2 Q17	**.689**	**.664**	**.841**	**.622**	1	**.433**
L2 Q18	.309	.279	**.393**	**.363**	**.433**	1

The correlations presented in Table 7.8 suggest that there is a strong structure in the data set as a whole for these variables. The correlations range between 0.279 and 0.841. Following the methodology in section 7.3.3, if there are two clusters in the data set for the six lecturer-specific questions for lecturer 2, the within-cluster correlations should be substantially lower than the between-cluster correlations. Tables 7.9 and 7.10 illustrate the bivariate correlations for each of the six lecturer-specific questions for lecturer 2 for each of the clusters separately.

Table 7.9 Correlation matrix for the six lecturer-specific questions for lecturer 2 in module B530~97 within cluster 1

	L2 Q13	L2 Q14	L2 Q15	L2 Q16	L2 Q17	L2 Q18
L2 Q13	1	.312	.208	.312	**.500**	.290
L2 Q14	.312	1	.296	.345	.483	-.091
L2 Q15	.208	.296	1	.296	**.533**	.000
L2 Q16	.312	.345	.296	1	.378	.362
L2 Q17	**.500**	.483	**.533**	.378	1	.218
L2 Q18	.290	-.091	.000	.362	.218	1

Table 7.10 Correlation matrix for the six lecturer-specific questions for lecturer 2 in module B530~97 within cluster 2

	L2 Q13	L2 Q14	L2 Q15	L2 Q16	L2 Q17	L2 Q18
L2 Q13	1	.358	**.647**	.182	**.541**	.015
L2 Q14	.358	1	**.466**	**.470**	**.484**	.015
L2 Q15	**.647**	**.466**	1	.392	**.736**	.137
L2 Q16	.182	**.470**	.392	1	.286	-.026
L2 Q17	**.541**	**.484**	**.736**	.286	1	.151
L2 Q18	.015	.105	.137	-.026	.151	1

A quick glance clearly indicates that the degree of correlation in Tables 7.9 and 7.10 is far lower than those in Table 7.8. In Table 7.9 only two bivariate combinations (Q13, Q15; Q15, Q17) and in Table 6.10 only six bivariate combinations (Q13, Q15; Q13, Q17; Q14, Q15; Q14, Q16; Q14, Q17; Q15, Q17) are significant to the 1% level. Each of these bivariate correlations significant to the 1% level in Table 7.10 are lower than the corresponding correlations in Table 7.8. The difference between the between and within-cluster correlation coefficients are not quite as clear in this example compared to the example of a single lecturer module presented in section

7.3.3. However, the correlation coefficients presented in Tables 7.8 to 7.10 still provide evidence of cluster structure in module B530~97.

The validity of this methodology at depicting whether clusters are present in the data set can be tested by examining the differences between the between and within-cluster correlations for lecturer 1 in module B530~97. As there is little difference in the ratings given to lecturer 1 by students in either of the clusters, it is expected that the six lecturer-specific questions for lecturer 1 will show high between-cluster correlations *and* high within-cluster correlations. Table 7.11 presents the bivariate correlation coefficients for the six lecturer-specific questions for lecturer 2 over the data set as a whole. Correlations shown in bold are significant at the 1% level.

Table 7.11 Correlation matrix for the six lecturer-specific questions for lecturer 1 in module B530~97

	L1 Q13	L1 Q14	L1 Q15	L1 Q16	L1 Q17	L1 Q18
L1 Q13	1	**.733**	**.511**	**.582**	**.481**	**.433**
L1 Q14	**.733**	1	**.630**	**.527**	**.548**	**.403**
L1 Q15	**.511**	**.630**	1	**.658**	**.824**	**.505**
L1 Q16	**.582**	**.527**	**.658**	1	**.660**	**.537**
L1 Q17	**.481**	**.548**	**.824**	**.660**	1	**.410**
L1 Q18	**.433**	**.403**	**.505**	**.537**	**.410**	1

As expected, Table 7.11 clearly shows a strong degree of correlation in the lecturer-specific questions for lecturer 1 in the data set as a whole. The correlations range between 0.403 and 0.824.

Tables 7.12 and 7.13 illustrate the bivariate correlations for each of the six lecturer-specific questions for lecturer 1 for each of the clusters separately.

Table 7.12 Correlation matrix for the six lecturer-specific questions for lecturer 1 in module B530~97 within cluster 1

	L1 Q13	L1 Q14	L1 Q15	L1 Q16	L1 Q17	L1 Q18
L1 Q13	1	**.796**	**.648**	.485	**.579**	.391
L1 Q14	**.796**	1	**.764**	**.583**	**.651**	.439
L1 Q15	**.648**	**.764**	1	**.682**	**.882**	**.512**
L1 Q16	.485	**.583**	**.682**	1	.735	**.586**
L1 Q17	**.579**	**.651**	**.882**	.735	1	**.533**
L1 Q18	.391	.439	**.512**	**.586**	**.533**	1

Table 7.13 Correlation matrix for the six lecturer-specific questions for lecturer 1 in module B530~97 within cluster 2

	L1 Q13	L1 Q14	L1 Q15	L1 Q16	L1 Q17	L1 Q18
L1 Q13	1	**.690**	**.433**	**.626**	**.418**	**.437**
L1 Q14	**.690**	1	**.549**	**.448**	**.471**	.314
L1 Q15	**.433**	**.549**	1	.629	.793	**.477**
L1 Q16	**.626**	**.448**	.629	1	.603	**.451**
L1 Q17	**.418**	**.471**	.793	.603	1	.300
L1 Q18	**.437**	.314	**.477**	**.451**	.300	1

Tables 7.12 and 7.13 show very high correlations between the six lecturer-specific questions for lecturer 1 in module B530~97. The correlations in cluster 1 (Table 7.12) are generally *higher* than the corresponding correlations in the data set as a whole (Table 7.11). Whilst the correlations in cluster 2 (Table 7.13) tend to be lower than the corresponding correlations in the data set as a whole. However, the differences are slight, all but two bivariate combinations (Q14, Q18; Q17, Q18) are statistically significant at the 1% level. These results strengthen the conclusion that module B530~97 has a cluster structure defined by students' judgements of one of the lecturers (lecturer 2).

7.6 General pattern of cluster structure for modules taught by two or three lecturers

The evidence in section 7.5 depicts a cluster structure relating to differences between the lecturers in one Business School module taught by more than one lecturer. It is now important to examine the extent to which this cluster structure is representative of other modules taught by either two or three lecturers.

In the overall sample of 65 modules, 24 were taught by either two or three lecturers. There was no evidence of a cluster structure in the responses for 11 (46%) of these 24 modules. Table 7.7 reports on the remaining 13 modules (54%) where the responses were considered to possess a cluster structure.

Table 7.14 Modules taught by more than one lecturer possessing a cluster structure

Code	Reg.Studs	#cases	Any Clusters	Key Variables	No Difference	Cluster sizes
A015~97	100	78	TWO	L1 13-18	7 11 L2 13&18	60/18
A057~97	201	113	THREE	2 3 L1 16&17		54/27/32
A085~97	159	106	TWO	L3 13-18		43/63
A110~97	77	45	TWO	2 3 8 9 12 L2 17	L3 15	16/29
A525~97	162	105	THREE	3 L2 13-17	5 6	34/51/20
B012~97	79	55	TWO	2 3 7 8 9 10 L1 15-17		17/35
B040~97	165	109	TWO	2 3 7 10 L1 15&17	13 L316&18	55/54
B120~97	85	67	TWO	2 3 8-11 L1 17 L2 17	L3 13-18	45/22
B530~97	94	65	TWO	3 8 L2 13-18		27/38
C050~97	159	99	TWO	9 10 L2 13, 15&17 L1 15-17		33/66
C055~97	85	43	TWO	2 3 8 10 L1 13-17		28/15
C072~97	59	52	THREE	3 L2 13 15-17		20/20/12
C132~97	78	70	THREE	5 L2 15-17		17/22/31

The main pattern reflected in Table 7.14 is that in many of these modules the clusters are differentiated on the basis of students' perceptions of one of the lecturers.

Table 7.15 Frequency with which each question is key in determining cluster membership in the sample of modules taught by more than one lecturer

Question	Frequency
1	0
2	6
3	8
4	0
5	1
6	0
7	2
8	5
9	4
10	5
11	1
12	1
13	7
14	5
15	10
16	9
17	13
18	3

Table 7.15 shows the frequency with which each of the eighteen variables were key variables in determining cluster membership. Table 7.15 can be directly compared to Table 7.6 which reported the corresponding frequencies with which each of the eighteen variables were key variables in determining cluster membership in the sample of modules taught by a single lecturer. The most salient point to note when comparing Table 7.6 and Table 7.15 is the change in the frequency with which Q2 and Q3 are key variables compared to Q15 and Q17. Whereas in Table 7.5 when Q15 and Q17 were key variables in differentiating clusters, they were almost always accompanied by Q2 and Q3 which were also key variables, an examination of Table 7.14 reveals that this relationship is not as strong in modules taught by more than one lecturer. This is in line with the expectation mentioned in section 7.5. It reflects the fact that in modules taught by more than one lecturer the data on the module-specific questions are the amalgam of students experiences of different lecturers, whose teaching methods and ability to stimulate the interest of the students may differ.

7.7 The application of a stopping rule

7.7.1 The need for stopping rules

One of the most problematic decisions that the analyst has to make during the process of cluster analysis is to determine whether there are any clusters in the data set and if so, which cluster partition is the most appropriate division of the data.

The initial method of determining the most suitable partition of the data set is the examination of the dendrogram. There are two inherent problems with relying on a visual inspection of the dendrogram. Firstly, the clustering power of the cluster algorithms is so strong that clusters will be depicted even when none are present in the data set (Barney and Hoskisson, 1990). Secondly, the decision over the most appropriate partitioning of the data set is a subjective decision if the analyst relies simply on a visual inspection of the dendrogram (see Ketchen and Shook, 1996). Virtually all the clustering algorithms provide little, if any, information as to the number of clusters present in the data (Milligan and Cooper, 1985).

However, the analyst is not restricted to a visual inspection of the dendrogram as a way of deciding on the most appropriate partition of the data set. To reduce the element of subjectivity in the decision of how to partition the data set, various objective procedures to determine the most suitable partition of the data set have been proposed. When applied to the results of hierarchical cluster analysis these techniques are termed 'stopping rules' (Milligan and Cooper, 1985). Milligan and Cooper (1985) conducted a simulation study that contrasted the performance of 30 stopping rules. The stopping rules that were found to be generally accurate and reliable include the Ratio-criterion method (Duda and Hart, 1973); the C-Index method (Hubert and Levin, 1976); the Gamma method (Baker and Hubert, 1975); the F-ratio test (Beale, 1969) and the Mojena rule (Mojena, 1977).

7.7.2 The Calinski and Harabasz stopping rule

The stopping rule found to give the best cluster recovery in the Milligan and Cooper simulation study was devised by Calinski and Harabasz (1974). The Calinski and Harabasz index was found to perform consistently well across varying numbers of clusters in the simulated data and across varying levels of noise (Milligan and Cooper, 1985).

The Calinski and Harabasz index computes the ratio of the mean between-cluster variance to the mean within-cluster variance. The formula can be stated as:

$$\mathrm{VRC} = \frac{\mathrm{BGSS} / (k - 1)}{\mathrm{WGSS} / (n - k)}$$

where:

VRC = variance ratio criterion
BGSS = between-group sums of squares
WGSS = within-group sums of squares
k = number of clusters
n = number of cases

Therefore each possible cluster partition returns a VRC value. In interpreting the VRC values Calinski and Harabasz instruct that the most suitable partition of the data set is:

> 'that number k for which the VRC has an absolute or local maximum, or at least has a comparatively rapid increase' (Calinski and Harabasz, 1974, p.12).

The Calinski and Harabasz stopping rule was run on each of the 65 modules. The results for all 65 modules are presented in Appendix 7.2. From the results it is clear that in all but one of the 65 modules, the highest variance ratio criterion (VRC) value occurs when the number of clusters is two. In many of these modules the VRC value increases at a steady rate as the number of clusters in the data falls. This indicates that there is no meaningful cluster structure in these modules:

> 'a uniform distribution of points in space will be usually reflected by a smooth run of values of the VRC' (Calinski and Harabasz, 1974, p.11).

However for some of the modules there was a comparatively rapid increase in the VRC value between two possible cluster partitions:

> 'if the points are grouped into...natural clusters, with small within-cluster variation, the change [between cluster partitions] will cause a...rapid rise of the VRC, possibly forming a hump' (Calinski and Harabasz, 1974, pp.11-12).

When this strategy was implemented the Calinski and Harabasz stopping rule provided evidence that supported the presence of a cluster structure in many of the Business School modules.

For example, there are eight modules in which three clusters were deemed to be present from the visual inspection of the dendrogram and the examination of the cluster means (see Appendix 7.1). The results of the application of the Calinski and Harabasz stopping rule for each of these eight modules are presented in Table 7.16.

Table 7.16 Stopping rule evidence of modules that have three clusters that agree with previous assessment

Code	**K = 6**	**K = 5**	**K = 4**	**K = 3**	**K = 2**	**%dif.6-5**	**%dif.5-4**	**%dif.4-3**	**%dif.3-2**	**Any Clusters**
A030~97	17.45	19.44	21.01	24.15	26.77	11.40	8.08	**14.95**	10.85	THREE
A057~97	14.97	16.32	18.29	22.6	27.49	9.02	12.07	**23.56**	21.64	THREE
B028~97	13.88	15.47	17.9	22.64	25.99	11.46	15.71	**26.48**	14.80	THREE
B510~97	11.6	12.99	15.43	19.06	21.35	11.98	18.78	**23.53**	12.01	THREE
C022~97	10.67	11.18	11.57	13.03	14.08	4.78	3.49	**12.62**	8.06	THREE
C072~97	8.95	10.01	11.89	14.98	18.34	11.84	18.78	**25.99**	22.43	THREE
C110~97	11.25	12.59	14.89	17.98	22.53	11.91	18.27	**20.75**	25.31	THREE
C132~97	8.84	9.49	10.49	12.31	13.65	7.35	10.54	**17.35**	10.89	THREE

In each of the eight modules the difference from a four-cluster to a three-cluster solution is the biggest single percentage increase in the VRC value between the partitions presented in Table 7.16. Therefore the Calinski and Harabasz stopping rule supports the previous judgement of the presence of clusters in the Business School student feedback data for these modules.

However, using the same criteria the Calinski and Harabasz rule would suggest that there are three clusters in each of the nine modules reported in Table 7.17. In each of these modules the previous assessment as reported in Table 7.5 and Table 7.14 (and Appendix 7.1) did not concur with the three-cluster assessment.

Table 7.17 Stopping rule evidence of modules that have three clusters that is counter to previous assessment

Code	**K = 6**	**K = 5**	**K = 4**	**K = 3**	**K = 2**	**%dif.6-5**	**%dif.5-4**	**%dif.4-3**	**%dif.3-2**	**Any Clusters**
A052~96	18.56	20.27	22.13	25.27	26.33	9.21	9.18	**14.19**	4.19	TWO
A530~96	10.75	11.86	13.6	16.96	18.5	10.33	14.67	**24.71**	9.08	Continuum
B510~96	10.37	10.94	11.86	14.08	16.61	5.50	8.41	**18.72**	17.97	Continuum
A070~97	15.09	16.92	18.93	22.74	26.99	12.13	11.88	**20.13**	18.69	
A100~97	10.49	11.62	13.04	15.19	16.29	10.77	12.22	**16.49**	7.24	TWO
A530~97	8.64	9.1	9.32	9.95	10.46	5.32	2.42	**6.76**	5.13	TWO
B020~97	9.34	9.83	10.5	11.04	10.84	5.25	6.82	**5.14**	-1.81	None
B022~97	12.65	13.91	15.07	17.16	18.48	9.96	8.34	**13.87**	7.69	Continuum
C005~97	5.93	6.16	6.49	7.43	8.12	3.88	5.36	**14.48**	9.29	Unclear

From the Calinski and Harabasz stopping rule data the following two observations can be made. Firstly, the results of the Calinski and Harabasz stopping rule and the results from the inspection of the dendrogram and the cluster means do not always agree. Secondly, there is a tendency for the VRC value to increase at a steady rate with the two-cluster solution almost always having the highest VRC value.

These results do not invalidate the previous assessment that many Business School modules have a cluster structure. The prevalence of the two-cluster solution having the highest VRC value of the alternative cluster partitions is likely to indicate that there is a degree of overlap between the clusters. The Milligan and Cooper simulation study (1985) which concluded that the Calinski and Harabasz method was the best stopping rule of the stopping rules tested, used 'truncated multivariate normal mixtures *to ensure that the clusters did not overlap*' (Milligan and Cooper, 1987, p.330, italics added). This method was taken to conform to the concept of 'natural clusters' whereby clusters should form distinct groups with internal cohesion and external isolation (Cormack, 1971). In other words, groupings in a data set should be considered as clusters if they are mutually exclusive (i.e. they do not overlap).

Atlas and Overall (1994) argue that:

> 'the problem with much of this work [stopping rules] has been an unrealistic degree of separation between clusters or between the latent populations from which mixture samples are drawn' (Atlas and Overall, 1994, p. 581).

Atlas and Overall illustrated that the Calinski and Harabasz stopping rule will not perform particularly well when, to follow the terminology of Sneath (1969), the clusters are not separated by 'empty regions of space from other clusters'.

The results of the Calinski and Harabasz stopping rule offer support for the judgement in sections 7.3 to 7.6 that a cluster structure is present in some of the modules taught in the Business School. In many of the modules for which a suitable partition had been previously suggested the Calinski and Harabasz stopping rule supports this judgement. The VRC values do though suggest that the degree of distinctiveness between the clusters is unlikely to be considerable and that the clusters have a degree of overlap.

7.8 General conclusions

The research presented in chapter 6 illustrated that there was a potential problem with the reporting of class average ratings in modules in which there are groups of students who systematically view the quality of the module and the skills of the lecturer(s) differently. Given that it is Business School practice to report the results of student feedback in the form of class averages for each question on each module, it was important to examine whether this was simply a potential problem or an actual problem. The analysis on a sample of 65 modules revealed that there was indeed an actual problem with presenting student feedback results in the form of class averages as in just under half of the modules (30 out of the sample of 65) a cluster structure was present.

The cluster patterns found in the Business School data support the contention that there are groups of students who are utilising different learning styles. The congruence between the student's learning style and the lecturer's teaching style is shown to influence the lecturer's student feedback ratings in some Business School modules. Lecturers therefore need to be wary of the possibility that although their class average ratings may be at least satisfactory, this does not preclude the possibility that there may nevertheless be a significant number of students for whom the nature of the module and the lecturer's teaching style was not well suited to their chosen learning style. Such students will not judge the quality of the module and the skills of the lecturer(s) to be satisfactory, as the lecturer's aims and objectives when they enter the lecture theatre (manifested through their teaching style) will be incongruent with the student's aims and objectives. As a result the lecturer's teaching style will not particularly facilitate the student to develop their own understanding in the way that the student intends.

A scenario in which nearly all the students on a module consider the module and teaching satisfactory, is rather different from a scenario in which half the students found the module and teaching suited to their learning style whilst the other half found the module and teaching not suited to their learning style. In the first scenario there is no need for the lecturer(s) to consider major changes to the module or their teaching practices (although with only satisfactory ratings they may consider some general improvements). On occasions in which the second scenario reflects reality, there is a need for the lecturer(s) to consider appropriate changes to the module or their teaching practices and then implement them.

The implication for lecturers is that they need to be aware of differences in learning styles being employed by students in the class and strive, wherever possible, to make their teaching as useful as possible for the greatest proportion of a particular cohort of students. To do this lecturers should pay close attention to the spread of the ratings in each module they teach, rather than simply relying on the class average rating for each question, as a guide to the student's perceptions of the quality of the module and the lecturer(s) teaching it. The Business School student feedback system provides an opportunity for lecturers to scrutinise the spread of their ratings for each question on each module as lecturers receive individual sheets, which in addition to reporting the class average rating, also report the proportion of negative responses (i.e. ratings under 3) and the proportion of positive responses (i.e. ratings over 3) for each question on the student feedback form.

Although this information can only provide an indication of the presence of distinct groups of students in the class, the lecturer themselves can use the written comments on the individual student feedback forms (which are confidential to lecturers who retain the forms relating to their own modules) to gain an insight into which aspects of the module or their lecturing the students particularly appreciated or particularly disliked. The wording of the open-ended questions on the reverse side of the student feedback form should facilitate this type of analysis as there are separate boxes that ask the students to state what they liked about the module and how they thought the module could be improved. An examination by the lecturer of the negative written responses might sometimes allow the lecturer to pinpoint the issues raised by the critical students and decide whether a change in either the module design or their lecturing is required.

Departmental managers also need to be aware of the problem of lecturers' student feedback ratings being influenced by distinct groups of students in the class reflecting differences in students' learning styles, whereby a particular teaching style will be appreciated by some of the students in the class but not appreciated by others. Managers should therefore also look at the spread of a lecturer's ratings for a module before discussing the matter with the lecturer and also consider a lecturer's student feedback ratings over a profile of modules. The need to look at the profile of a lecturer's ratings over time becomes important given that a lecturer teaches a different set of students each year. Any one cohort of students may well differ from another in terms of the dominant learning style utilised by the students in the class. Therefore a teaching style that is congruent with the learning style of the majority of students in the class (leading to the lecturer receiving relatively high ratings) may be less compatible with the dominant learning style in the class the following year, which is likely to lead to the lecturer receiving lower ratings. This difference in ratings year on year, despite the fact that the lecturer has neither changed their teaching style or the module content, stresses the importance for student feedback ratings of the interaction between the lecturer and the student. This theme will be developed further in the next chapter.

The analysis in this chapter also indicates that students are considering the skills of the lecturer when answering some of the module-specific questions, particular Q2 and Q3 on the student feedback form. This was far more pronounced in modules taught by a single lecturer, reflecting the fact that in modules taught by more than one lecturer the data on the module-specific questions are the amalgam of students' experiences of different lecturers whose teaching methods and ability to stimulate the interest of the students may differ.

The cluster structure in module B530~97 (see section 7.5) suggests that there may be a contrast effect in the Business School student feedback ratings whereby lecturers' ratings are influenced by the performance of the lecturer(s) with whom they teach on a particular module. This theme is explored in greater depth in chapter 9.

Chapter 8

THE EFFECT OF POSITION

8.1 Introduction

The student feedback form provided centrally by Loughborough University has provision for up to three lecturers to be evaluated on each module. In the modules in which there are either two or three lecturers the 'module leader', i.e. the lecturer who has the overall responsibility for the module, automatically appears as lecturer 1 on the feedback form. In the modules in which there are three lecturers the placing of the remaining two lecturers on the student feedback form is non-systematic. In this chapter research is presented that explores whether there is any relationship between the position that the lecturer appears on the student feedback form and the ratings that they receive. The analysis in section 8.2 seems to suggest that the position that a lecturer appears on the student feedback form does influence ratings. This is worrying since, if there is a causal relationship between position and ratings, the validity of the data collection process and the results based on it would be called into question. A series of interviews was carried out with lecturers whose average ratings are significantly different when they appear in different positions on the feedback form. Typically their scores are at their highest when they appear as lecturer 1 on the feedback form and at their lowest when they appear as lecturer 3. The main findings of these interviews are outlined in section 8.3. The outcome of these interviews suggested that a new variable, the ratio of lecturing hours undertaken by each lecturer on a module, might explain the variation in ratings. The method of creating this new variable is discussed in section 8.4. The effects of this variable are demonstrated in a regression model presented in section 8.5 and this section also includes the rationale of a second new variable used as the dependent variable in the model. In section 8.6 the effect of removing from the analysis those lecturers who always lecture on single lecturer modules is discussed. In section 8.7 it is suggested that the underlying reason why the proportion of hours a lecturer teaches on a particular module influences the ratings they receive may reflect the lecturer having more time in which to develop an interpersonal relationship with the students. In particular, the lecturer will have more opportunity to build rapport with the students and to develop a degree of trust between them and the students. Finally section 8.8 presents a summary of the findings and some conclusions.

8.2 Initial analysis of the data

The student feedback form has provision for up to three lecturers to be evaluated on each module. Therefore in three-lecturer modules a lecturer could appear in any one of three positions on the form. Over the course of all the modules on which they teach a lecturer is likely to vary in the position that they appear. If the position that the lecturer appears on the form influences a lecturer's student feedback ratings there would be cause for concern as this would suggest that a lecturer's ratings were partly explained by a design feature of the form rather than reflecting aspects of lecturing skill. It is therefore necessary to establish the principle that the position that a lecturer appears on the form does not influence their ratings.

To investigate whether or not the position that a lecturer appears on the student feedback form influenced their ratings the means for each lecturer-specific question were computed separately for each position on the feedback form. It was hoped that there would be no statistically significant difference in the means for each position on the feedback form for any of the lecturer-specific questions (Q13-Q18). The average scores for each question by position on the feedback form is presented in Table 8.1. It was calculated using the individual form level master data set (version A - section 5.3.1).

Table 8.1 Average scores for each question at the individual form level by position on the student feedback form

Position	Q13	Q14	Q15	Q16	Q17	Q18	*Total responses*
1	4.19	4.26	3.85	3.89	3.93	3.76	*78572*
2	3.95	3.99	3.62	3.74	3.68	3.51	*35379*
3	3.79	3.65	3.38	3.51	3.38	3.36	*10858*
Total responses	*20980*	*20969*	*20975*	*20969*	*20456*	*20460*	*124809*

The means presented in Table 8.1 were not in line with expectations. Table 8.1 clearly illustrates that the mean score for lecturer 1 to be higher than for lecturer 2, which is in turn higher than lecturer 3. This relationship is repeated for each lecturer-specific question on the feedback form.

The statistical significance of these differences was tested using regression analysis with indicator variables. This process is equivalent to that used in analysis of variance. The regression procedures followed those presented in chapter 5. The regression analysis was run using the module level data set described in section 5.4. As discussed in section 5.4, the regression analysis needed to be conducted using the module level data, as if regression was run on the individual level data the effects of the predictor variables would be swamped by the between respondent variation. By working at the module level the effects of the predictor variables on ratings should be clearer. A separate regression model was created for each question in Table 8.1. In each case the question was used as the dependent variable and two predictor variables one for position 2 (POST2) and the other for position 3 (POST3) were entered into the model. Position 1 was used as a reference group and therefore not entered into the model.

The model for Q17 is presented as equation 8.1. It was chosen for presentation as it represents the most general lecturer-specific question, 'the lecturer overall was effective'.

Brief details of the other models follow the discussion of equation 8.1.

Predictor variables (2): POST2 and POST3.

$$Q17 = 4.000 - .185\ POST2 - .474\ POST3$$

$$\text{t-values:} \quad (137.688) \quad (-3.528) \quad (-5.402) \qquad [8.1]$$

$R^2 = .073$ $F = 17.914$ $(p = .000)$

Given the figures presented in Table 8.1 it was expected that the differences between the positions would prove to be significantly significant. This would be shown if both the position variables (POST2 and POST3) appeared in the final model. This would indicate that POST 2 was different from the reference group POST1 and also different from POST 3, therefore implying that there are significant differences between each of the lecturing positions. Equation 8.1 shows that this is indeed the case. Both POST2 and POST3 prove to be significant in the final model. The t-values for both POST2 and particularly POST 3 are high.

The regression models were similar for each of the other questions presented in Table 8.1. In each case both position variables were statistically significant at the 5% level. The value of the constant ranged between 3.83 (Q18) and 4.30 (Q14). The value of R^2 ranged between 0.041 (Q15) and 0.143 (Q14).

This unexpected finding was particularly worrying. If ratings are not only dependent on the teaching/learning experience but are also a function of the student feedback form itself, this would cast doubt on the validity of the student feedback form at providing information that genuinely reflected the performance of the lecturer on the module.

It is important to establish whether it is simply the position that a lecturer is placed on the feedback form that is causing this relationship, or whether there are alternative explanations that could explain the results presented in Table 8.1.

To investigate the issue of position on the feedback form the ratings for each of the 45 lecturers for each position were calculated. These were then analysed by creating a different table for each question.

Nine of the lecturers appeared in each of the three positions on the student feedback form over the two academic years covered in this research. Using Q17 as an example, four had noticeable falls in their ratings between position 1 and position 2 and a further fall between position 2 and position 3. In another case the lecturer's average ratings for position 2 was fractionally higher than their average ratings for position 1, with their average ratings when they appeared as the third lecturer on the feedback form significantly lower than in either position 1 or 2. In another four cases the average rating for a lecturer who appeared as lecturer 3 on the feedback form was higher than when they appeared as lecturer 2, but both of these were far lower than ratings for lecturer 1.

In summary, many (although not all) of the lecturers who appeared in all three positions on the feedback form over the two academic years, had their highest average scores when they appeared as lecturer 1 and their lowest average scores when they appeared as lecturer 3.

8.3 Interviews with selected lecturers

Four lecturers were selected for interview on the basis of having average scores for Q17 that descended across the three positions on the student feedback form. They were asked if they could offer any explanation as to why their average scores changed significantly between positions on the student feedback form. Various suggestions were offered by the four interviewees as potential explanations for the variation in their ratings. There was no universal agreement between the four, who were interviewed individually, without knowledge of the responses given by previous interviewees. The main explanations offered are outlined below.

a) Effect of module leader. Two of the interviewees believed that the module leader (lecturer 1) has a built in advantage as they were able to set the norms for the module and therefore directly influence the students' expectations of what teaching on that particular module should be like. The ratings of other lecturers who do not conform to these expectations would then suffer. Module leaders are also at a potential advantage, as they tend to give students more detailed advice about the coursework and examinations.

b) Building trust. Two of the interviewees believed that one of the key elements in student feedback ratings was building a relationship of trust with the students. Lecturers who have more exposure to students on a particular module have greater opportunity to achieve this. One of the interviewees discussed the wider issue of visibility and overall significance of a particular lecturer for that set of students in their studies as a whole. For example, Programme Directors teaching on the degree that they direct may have a built in advantage in the ratings.

c) Extent of contact. One of the interviewees stated that when a lecturer teaches a large proportion (or all) of the module they have a greater opportunity to integrate material. They can refer students back to a previous lecture and can therefore incrementally build students' knowledge. Whereas a lecturer who only delivers a small proportion of the lectures on the module has less contact with the students and correspondingly less opportunity to incrementally build the students' knowledge.

d) Difficulty level. One of the interviewees suggested that when a lecturer teaches only one or two lectures on a module they are often specialist topics which may be rather more difficult than other topics on that module. Another interviewee suggested that when a lecturer only teaches one or two sessions on the module, there may be a tendency to push the students a bit harder than they would if they had a longer period in which to develop the student's understanding. In other words, a lecturer who lectures on a small proportion of the module has less time to make an impact on the students may be more inclined to pack more material into the lectures and thus make their lectures more challenging for the student. However research evidence (e.g. Marsh and Dunkin, 1992; Centra, 1993) indicates that difficulty level may be positively correlated with ratings, i.e. greater difficulty → higher ratings.

e) Different levels of attendance. One of the interviewees focused on the implication of attendance. If a module has an overall absenteeism rate of 10% and this is constant across each lecturer on the module, then a lecturer who only teaches one or two lectures is far more likely to be judged without foundation than the 'main' lecturer(s) on the module. There are likely to be a number of students who have had no exposure to the lecturer on the module. The effect would be even stronger if the absenteeism rate were higher than the average for the module in the lecture taken by the lecturer who only takes a single session. This situation may transpire if a student has previously decided that they do not intend to answer a question on that lecture topic should it appear on the examination paper. It may be possible to explain this with the following train of argument:

Lower attendance → Lower student motivation → Lower ratings.

f) Preparation time. Generally the four lecturers who were interviewed did not believe that preparation time per contact hour varied depending on their overall contribution to the module. One interviewee did think that preparation time might be slightly lower, as if one only teaches a single slot one may look to 'pick something off the shelf', but another interviewee believed that preparation time was probably higher if one only teaches a single slot. The other two believed that preparation time was essentially independent of the number of sessions taught on a module.

A close inspection of these explanations for differential ratings between positions on the student feedback form reveals that they each have one important factor in common. Each of these explanations is predicated on the assumption that those appearing as lecturer 3 on the feedback form have the least contact lecture hours with the students. This opens up the interesting possibility that it is the proportion of lecturing hours undertaken by a lecturer on a module rather than their position on the feedback form that accounts for the relationship presented in Table 8.1. If this were shown to be the case it could be concluded that it was not the instrument itself that was influencing student feedback ratings, but rather a variable that can logically relate to the experience of the students studying a module.

8.4 Construction of a new variable: Rat.L.Hrs.

To examine the degree to which the proportion of lecturing hours undertaken by a lecturer on a module influences ratings information needed to be collected on the proportion of lecturing hours taken by each of the lecturers on all of the modules that had more than one lecturer. The data was collected by reference to the teaching allocation sheets prepared by the Subject Area Co-ordinators for each section of the department. These sheets list each module taught within that section and the number of teaching hours that the relevant lecturers were due to teach. These sheets were incomplete or dubious in a number of ways. Firstly, in some modules the lecturers originally down to teach on a particular module did not actually do so and had subsequently been replaced (without the original sheet being amended) by another lecturer. Secondly, on some modules the full number of scheduled hours were not taught and therefore the ratio of teaching hours between lecturers indicated on the original teaching allocation sheets did not match the ratio of lecturing exposure that the two lecturers actually had to the students on the module. Thirdly, some of the Subject Area Co-ordinators had not marked a clear division between tutorial hours and lecturing hours. This was necessary information for this analysis as this investigation is only interested in lecturing. Tutorials at the Business School take a variety of forms and are normally conducted by personnel not included in this research, e.g. Ph.D. students or staff on short-term contracts. As information was available for some ‘tutors’ but not for others, it was decided to only analyse data that referred to lecturing.

To overcome these problems two other sources of information were used. Firstly, in some cases, the data on the teaching allocation sheets could be checked against the module outlines given to the students at the start of each module. On some of these sheets the breakdown of the lecture schedule giving both the lecture topic and the name of the presenting lecturer is given for each session. Where this was not available the individual lecturers concerned were approached and asked to recall their involvement on a particular module. For one section of the department the Subject Area Co-ordinator was approached and asked to provide information that differentiated between lecturing and tutorial hours.

This information allowed a new variable which was termed ratio of lecturing hours (Rat.L.Hrs.) to be created which was the proportion of the total lecturing hours taken by each lecturer on the module. In modules taught by a single lecturer the value of Rat.L.Hrs. corresponded to 1. This data was then entered into the main data, providing the opportunity to analyse the relationship between the proportion of lecturing hours undertaken by an individual lecturer on a particular module and the ratings they received.

8.5 Regression model using Rat.L.Hrs.

It was necessary to check whether POST and Rat.L.Hrs. were highly collinear. If they were they would effectively be providing the same information. Under these circumstances the inclusion of Rat.L.Hrs. in the model coupled with the absence of the two POST variables would fail to adequately illustrate that POST does not explain part of the variation in DIFF. Examination of the data showed that POST and Rat.L.Hrs. were not highly collinear. Out of the 40 modules taught by three lecturers only 8 had a pattern whereby lecturer 1 taught more than lecturer 2 who in turn taught more than lecturer 3. In a greater number of cases (10) lecturer 2 taught the greatest proportion of lecture hours, in five modules lecturing was split equally between the three lecturers and in 3 modules lecturer 3 taught the greatest proportion of lecture hours. In the remaining 14 modules there were a variety of patterns such as lecturer 1 teaching the greatest proportion of lecture hours and lecturers 2 and 3 taught the remaining lectures equally between them.

Out of the 104 modules taught by two lecturers, the lecturing was split evenly in over a third of the modules (37 out of 104) and in 13 modules lecturer 2 taught more lectures than lecturer 1 did. These findings provide reassurance that Rat.L.Hrs. and POST are not sufficiently closely associated to lead to a problem with collinearity when student feedback ratings are modelled against them.

With this new variable available in the data set, a regression model was run with Rat.L.Hrs. and POST2 and POST3 as the predictor variables. Before proceeding with this regression analysis a decision had to be taken as to what the most appropriate dependent variable would be.

Table 8.2 presents the bivariate correlation coefficients of each pair of lecturer-specific questions. An inspection of Table 8.2 shows that there is a high degree of correlation between the variables. The correlation coefficient for every pair of variables is significant at the 1% significance level. The variable that has the highest correlation with each of the other five variables is Q17.

Table 8.2 Bivariate correlation coefficients of each pair of lecturer-specific questions

	Q13	**Q14**	**Q15**	**Q16**	**Q17**	**Q18**
Q13	1.000	.732	.794	.640	.820	.546
Q14	.732	1.000	.765	.798	.828	.513
Q15	.794	.765	1.000	.777	.956	.549
Q16	.640	.798	.777	1.000	.819	.525
Q17	.820	.826	.956	.819	1.000	.587
Q18	.564	.513	.549	.525	.587	1.000

This is not surprising as this question ('overall, the lecturer was effective') represents the most general lecturer-specific question on the feedback form. A general lecturer-specific question on a feedback form is often termed the 'global' question in the student feedback literature (see for example Abrami and d'Apollonia, 1991; Abrami, d'Apollonia and Rosenfield, 1996).

The relationship between Q17 and the other five lecturer-specific questions was further explored using multiple regression. Using Q17 as the dependent variable and the other five lecturer-specific questions as the predictor variables gives the model presented as equation 8.2.

Predictor variables (5): Q13, Q14, Q15, Q16 and Q18.

$$\begin{array}{llllllll} \mathbf{Q17} = & -.369 & +.138\,Q13 & +.177\,Q14 & +.612\,Q15 & +.107\,Q16 & +.05367\,Q18 & \\ \text{t-values:} & & (4.805) & (6.816) & (29.760) & (5.001) & (2.764) & [8.2] \end{array}$$

$R^2 = .945 \quad F = 1540.071 \quad (p = .000) \quad n = 458$

With a R^2 of .945 this model has very strong predictive power. In particular 95% of the variance in Q17 is explained by variation in the other five lecturer-specific questions. Each of the other lecturer-specific questions is statistically significant in explaining the lecturer's rating on Q17. For each of the predictor variables the coefficient is positive. By far the most significant variable in explaining ratings for Q17 is Q15, which asks about the lecturer's clarity. This was also reflected by the very high correlation coefficient between Q15 and Q17 (.956) as illustrated in Table 8.2.

On the basis of the strong relationship between Q17 and the other five questions illustrated in Table 8.2 and equation 8.2, the lecturer's score for Q17 was selected as the basis of this analysis.

For the dependent variable we could have used the 'raw score' for Q17 on the feedback form. The raw score simply provides a measure of how well a lecturer was considered to have performed by the students. An attempt to look at the relationship between Q17 (in its 'raw score' format) and Rat.L.Hrs. would not have provided the type of information that was sought. To understand why this is the case one has to recall the layout of the data set. As discussed in section 5.3.2, the data for each lecturer on each module represents an individual case in the data set and is placed in sequence one on top of another. Consider the following hypothetical example. The first two cases in the data set are as follows:

Case 1: A particular lecturer (lecturer A) scores highly (say 4.4 on Q17) and is positioned as lecturer 1 on the student feedback form.

Case 2: A different lecturer (lecturer B) scores lower than lecturer A (say 3.8 on Q17) and is positioned as lecturer 2 on the student feedback form.

There may seem to be *prima facie* evidence here that the position that a lecturer appears on the student feedback form does influence ratings. The lecturer positioned as lecturer 1 has a substantially higher rating for Q17 than the lecturer positioned as lecturer 2. There is, of course, an alternative explanation, lecturer A may simply be a better lecturer than lecturer B. The difference in their ratings reflecting this difference in performance, rather than being the result of the position that they appeared on the student feedback form. To measure the extent to which a lecturer's student feedback ratings are influenced by their position on the student feedback form the dependent variable needs to compare a lecturer's ratings on a particular module to their overall average for a particular question over all the modules on which they lecture. In other words, measure whether lecturer A is scoring higher than usual by virtue of being positioned as lecturer 1 on the student feedback form. For the variable Rat.L.Hrs. to have practical significance it is necessary to establish that a lecturer's ratings vary directly with the proportion of lecturing they deliver on a module.

Another new variable was therefore created which represented the difference between a lecturer's rating for Q17 on each module and their overall average for this question, based on all the modules in which they lecture. This variable was termed DIFF, i.e. for any given lecturer on a specific module:

$$\text{DIFF} = \text{Q17}_{\text{ma}} - \text{Q17}_{\text{oa}}$$

where: ma = module average and
oa = overall average.

Based on the interview evidence presented in section 8.3 the expectation is that the greater the proportion of lecturing undertaken by a lecturer on a module the higher their ratings will be. In other words, a lecturer should score higher than their average in modules in which they deliver a greater proportion of the lecturing contact hours. Conversely, lecturers would be expected to score below their average in modules in which they only deliver a small proportion of the total number of lecture hours.

These predictor variables, two relating to position (POST2 and POST3) and the other relating to the proportion of lecturing hours delivered by each lecturer on each module (Rat.L.Hrs.). were entered into the regression model presented as equation 8.3.

These were regressed against the dependent variable of difference (DIFF). As before, POST1 is used as a reference group and Q17 is selected as being the most representative of the lecturer-specific questions. The sample size for equation 8.3 is 458. This consists of a separate case for each lecturer on each module, e.g. in a module with two lecturers there will be two cases in the data set. The 458 cases represents 149 cases for modules with one lecturer, 192 cases (2 x 96) for modules with two lecturers and 117 (3 x 39) cases for modules with three lecturers.

Predictor variables (3): POST2, POST3 and Rat.L.Hrs.

DIFF Q17 = - .0641 + .242 Rat.L.Hrs.
t-values: (-1.594) (4.186) [8.3]

$R^2 = .037$ F = 17.525 (p = .000)

POST2 and POST3 do not appear in this model. Hence the model suggests that there is no significant effect on ratings between being placed as the first, second or third lecturer on the feedback form in addition to the effect of Rat.L.Hrs. It does though suggest that the proportion of lecturing hours that each teaches on a module has a significant impact on the feedback ratings.

The value of the constant in the model is very close to zero. It is not statistically significant (t-value = -1.594). This confirms our expectations. Since the model is gauging differences about the mean the differences should sum to zero.

Similar models were found for each of the other lecturer-specific questions on the student feedback form. The position variables were not significant at the 1% level for any of the lecturer-specific questions.

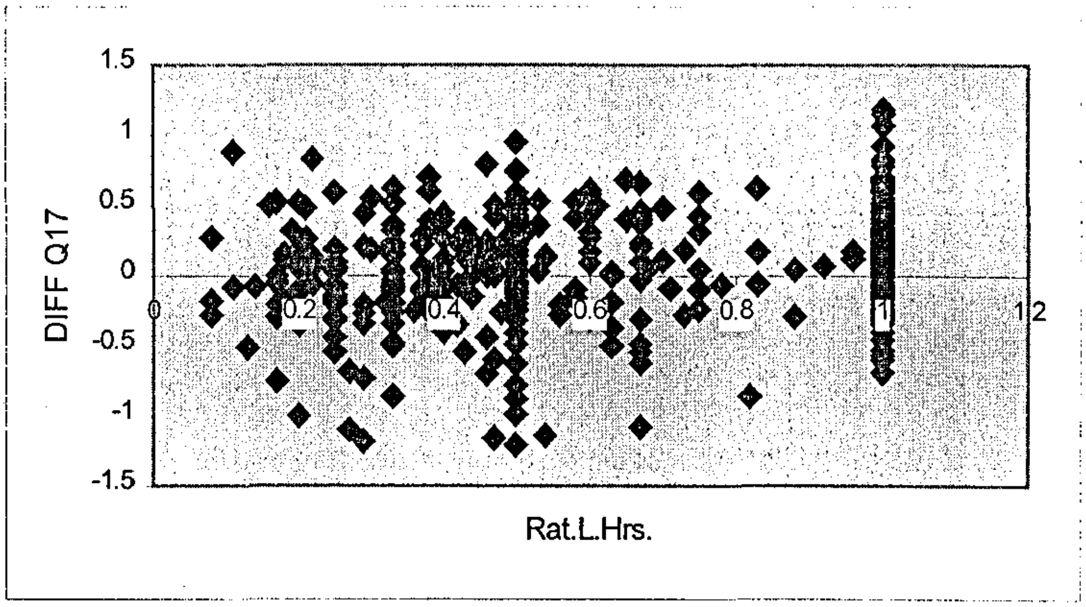
8.6 The effect of removing those who always teach on single lecturer modules

The regression model for DIFF Q17 (equation 8.3) suggested that the position that a lecturer appeared on the student feedback form did not, in addition to Rat.L.Hrs. have a significant impact on student feedback ratings. However the R^2 was very small illustrating that the model only accounts for a very small amount (3.7%) of the variation in DIFF Q17. The following analysis examines the effect on the model (particularly focusing on the R^2) of omitting those lecturers who always lecture alone. The reason for computing this second model was that for lecturers who always lecture on single lecturer modules there is no variation in the variable Rat.L.Hrs. Hence for these lecturers no variation in their ratings can be explained by variation in the ratio of lecturing hours.

By eliminating the lecturers who always lecture alone the value of R^2 should rise. This becomes clear when one examines Figure 8.1. Figure 8.1 presents a scattergram of Rat.L.Hrs. and DIFF Q17. The scale for DIFF runs from -1.5 to 1.5, where 0 represents the average score for a lecturer over all the modules on which they lecture. The actual range of DIFF Q17 is -1.21 to 1.18. The scale for Rat.L.Hrs. runs from 0 to 1.2, with the actual range 0.08 to 1 (the maximum possible ratio).

Figure 8.1 shows that there is a concentration of points on the x-axis in two places, the x-axis values of 0.5 and 1. The former generally occurs for those modules for which the lecturing is split equally in a two-lecturer module. The latter occurs for the modules which are taught by a single lecturer. It would be erroneous to assume that all points corresponding to $x = 1$ add no useful information to the analysis. For lecturers who *sometimes* lecture alone there will be a variation in their x-axis scores and this may explain some of the variation in DIFF Q17. However for those who *always* lecture alone, there will be no variation in their x-axis score and hence no possibility of examining any variation in DIFF Q17. Hence including those lecturers who always lecture alone will distort the model and dilute the strength of the relationship.

Figure 8.1 Scattergram of Rat.L.Hrs. and DIFF Q17 for all module level data



There are four lecturers who always lecture alone and two lecturers who almost always lecture alone. The regression analysis was re-run without these six lecturers. The resulting model presented as equation 8.4 has a sample size of 401 cases.

Predictor variables (3): POST2, POST3 and Rat.L.Hrs.

DIFF Q17 = - .09435 + .325 Rat.L.Hrs.

t-values: (-2.173) (4.880) [8.4]

$R^2 = .056$ $F = 23.815$ $(p = .000)$

The value of R^2 in equation 8.4 compared to equation 8.3, has noticeably increased, although it is still very small. The result reinforces the previous evidence that it is the proportion of lecturing hours undertaken by a lecturer on a module rather than their position on the feedback form *per se* that influences their ratings.

8.7 The relevance of the student-lecturer interaction

The significant influence that the proportion of lecturing undertaken by a lecturer on a module has on lecturers' ratings is particularly interesting as it highlights a student-lecturer interaction in student feedback ratings. The evidence from the interviews conducted with Business School lecturers explicitly recognised that lecturers who teach a greater proportion of the module gain a greater rapport with the students and have more opportunity to build trust with their students.

Interpersonal dynamics between lecturers and students therefore are an influence on lecturers' student feedback ratings. These interpersonal dynamics are therefore a second type of interaction effect that influences lecturers' ratings, alongside the other interaction effect discussed previously namely, the compatibility of the student's learning style and the lecturer's teaching style.

An understanding of the nature of the interpersonal relationship between the lecturer and the student has substantial implications for both effective teaching practices and the understanding of lecturers' student feedback ratings.

The psychologist Carl Rogers (1993) argues that:

> 'the facilitation of significant learning rests upon certain attitudinal qualities that exist in the personal *relationship* between the facilitator [lecturer] and the learner [student]' (Rogers, 1993, 230, emphasis in the original).

Rogers outlines the skills that aid a positive interpersonal relationship between the lecturer and the student as a:

> 'transparent realness in the facilitator [lecturer], a willingness to be a person, to be and live the feelings and thoughts of the moment. When this realness includes a prizing, caring, a trust and respect for the learner, the climate for learning is enhanced. When it includes a sensitive and accurate emphatic listening, then indeed a freeing climate, stimulative of self-initiated learning

> and growth, exits. The student is *trusted* to develop (Rogers, 1993, 241, emphasis in original).

Rogers conceptualisation of good teaching practice stresses the interconnection between teaching and learning. This is reflected in the evidence presented in chapters 6 and 7 that highlighted the importance of the relationship between the lecturer and the student through the compatibility of the student's learning style and the lecturer's teaching style. The need to view good teaching as an interaction between lecturer and student is recognised by Ramsden (1989), who writes:

> 'I think we ought to focus on both students and teachers...The time has passed, if it ever existed, when we could locate students' learning problems inside their heads and carry on teaching them regardless. We, the teachers, need to reflect on what we do that influences the relation between the students and what they learn. It is then our job to act on what *we* have learned about the relation and our part in it' (Ramsden, 1989, emphasis in original).

Viewing teaching as an interaction between teacher and pupil has long been the norm in school education. Until recently there may have been less of an appreciation of the importance of the interaction in higher education, particularly in terms of lecturers' behaviour in lecture theatres. One of the effects of the implementation of student feedback systems in U.K. universities is that it helps to raise the profile of the interaction between the lecturer and the students as the lecturer is faced with the knowledge of how effective the students found their teaching. Ideally, lecturers, as a consequence of the student feedback ratings they receive will consider adapting their teaching and lecture theatre behaviour. In doing so, they are implicitly recognising that neither they nor the students can be considered in isolation, but rather the learning experience for the student directly reflects the interaction between themselves and the lecturer. The implicit focus on this interaction is one of the positive by-products of the process of collecting student feedback data.

The evidence in this chapter allows the relevance of the student-lecturer interaction to be taken a stage further, by stressing the need for lecturers to *motivate* students to develop their cognitive capacities. The lecturer needs to build trust with the students in such a way as the student develops confidence in the lecturer and in their own abilities and consequently their chances of success in the module. Abouserie (1995) has found:

> 'a significant positive correlation between students' self-esteem and their scores on comprehension learning, meaning orientation and deep processing. This suggests that individuals with high self-esteem more frequently involve their self concept in data processing and in adopting deep approaches to studying' (Abouserie, 1995, 24).

Therefore, students' self-esteem has a significant effect on the way they deal with information and with learning situations. Abouserie also found evidence that shows:

> 'a significant positive correlation between students' achievement motivation and their scores on... meaning orientation...and elaborate processing, which suggests that students with high achievement motivation are likely to adopt deep and elaborate approaches in their study' (Abouserie, 1995, 24).

One of the implications of focusing on a student-lecturer interaction is that it places the emphasis on teaching to respond to learning as being:

> 'active and...invariably [involving] change, which is why its most significant outcomes can never be pre-specified with certainty. A consequence of conceptualising learning in this way is that student motivation becomes a major pedagogical concern. If motivation is seen not as a precondition of learning but as a vital and active constituent of learning, then student motivation itself becomes central to the task of higher education' (Nixon, 1996, 10).

Significant outcomes can never be specified with certainty as the lecturer (although possibly teaching the same module with the same syllabus year after year) faces a new set of students on each occasion they teach the module. Once student feedback ratings are understood to reflect the interaction between the lecturer and student it becomes clear why a lecturer's student feedback ratings may change year on year, even for the same module. The students are different, the appropriateness of the lecturer's teaching style for a different set of students may be more of less appropriate than for the previous years students and this in turn is likely to affect the degree to which the lecturer builds rapport with the students and engages their trust.

Therefore, although the lecturer may teach the same module in the same way, using the same handouts and examples in the lecture theatre, there is no reason to expect similar ratings year after year. Any lecturer who does expect similar ratings year after year fails to appreciate the effects of the student-lecturer interaction on lecturers' student feedback ratings that have broadly been discussed in this chapter.

The dynamic, often unpredictable, nature of student-lecturer interactions makes it important that lecturers and departmental managers do not consider any one set of student feedback data in isolation. An important implication of these findings is that lecturers should not get too despondent if one set of student feedback results are unfavourable to them. Furthermore it is important that lecturers do not make rash changers to either the module of their teaching in responses to one set of ratings, especially if these ratings are not in line with the student feedback ratings from comparable modules on which they lectured. Similarly managers must not treat lecturers harshly, potentially knocking the lecturers confidence, as a result of one set of student feedback ratings. The findings in this chapter stresses the need for managers to take a measured approach to judging lecturers in light of the student feedback ratings they receive from the students.

8.8 Conclusions

The evidence in this chapter has shown that the position that a lecturer appears on the student feedback form does not significantly affect the ratings that a lecturer receives. The evidence does suggest that an alternative variable, the proportion of lecture hours taught by a lecturer on a particular module (Rat.L.Hrs.), does have an impact on the ratings that the lecturer receives. The underlying reason why the proportion of lecturing hours that a lecturer teaches on a particular module influences the ratings they receive may be explained by the extra contact with students giving lecturers more time to develop an interpersonal relationship with the students. This is particularly likely to be manifested in the building of rapport and the development of trust between the lecturer and the student. The relevance of Rat.L.Hrs. is an important finding as this variable is a potential bias on student feedback data and has not featured in the student feedback literature. The current evidence suggests that the impact of Rat.L.Hrs. on student feedback data is mild (the value of R^2 when student feedback data is modelled against Rat.L.Hrs. is small). Subsequent analysis will illustrate the extent to which Rat.L.Hrs. continues to have a significance influence on student feedback ratings when other predictor variables are added to the model.

An implication of the evidence presented in this chapter is that lecturers and departmental managers should make allowances for the proportion of hours taught by a lecturer on a module when they examine a lecturer's student feedback ratings. A further implication of the evidence in this chapter, that the position that the lecturer appears on the student feedback form is unimportant, is that there is no need for departmental managers to systematically vary the position that a lecturer appears on the student feedback form. Those who criticise the student feedback process for disadvantaging lecturers who rarely appear as lecturer 1 (or who regularly appear as lecturer 3) are shown to be employing an argument that is ill-founded.

The evidence that the position that the lecturer appears on the student feedback form does not directly influence the student feedback ratings a lecturer receives, supports the validity of the student feedback form showing that in this aspect of instrument design the student feedback form fairly and accurately reflects the judgements of the students towards the module and the lecturer(s) teaching on it.

The research also provides a rationale for why lecturers should not expect to receive the same ratings for the same modules year-on-year even if they teach the module in the same way and use the same materials. Their ratings each year will be influenced by how congruent their teaching style is to the dominant learning style being utilised by students in the class. As the preponderance of various learning styles are likely to change year-on-year lecturers' ratings will too.

The analysis of the effect of Rat.L.Hrs. is particularly important for interpreting the student feedback data collected in the Business School as the lecturing responsibilities are shared in many modules. Another potential influence on student feedback data that derives directly from the sharing of lecturing responsibilities, i.e. the extent to which a lecturer's ratings are influenced by the ratings of the lecturer(s) with whom they lecture on a particular module, is examined in the following chapter.

Chapter 9

CONTRAST EFFECTS IN STUDENT FEEDBACK RATINGS

9.1 Introduction

This chapter explores whether the psychological concept of 'contrast effects' has applicability to student feedback ratings. The central proposition of a contrast effect is that the performance of other lecturer(s) teaching on a particular module influences a lecturer's student feedback ratings. Section 9.2 presents an example from the Business School student feedback data that suggests that a contrast effect in student feedback ratings may be present. Section 9.3 outlines the rationale for the dependent variable and discusses two alternative predictor variables that could be used to measure the contrast effect of lecturers' feedback ratings on each other. In section 9.4 three regression equations, one for modules with two lecturers, another for modules taught by three lecturers and a combined model are discussed. The three models were run on both the alternative predictor variables. Residual plots are used to determine the degree of linearity of the data. Section 9.5 presents the rationale for an extended model, which introduces four additional predictor variables. Section 9.6 presents three regression equations for the extended model. These equations correspond to the three equations presented in section 9.4 and use both of the predictor variables measuring contrast effects discussed in section 9.3. The two predictor variables measuring contrast effects are then contrasted in light of the various regression equations. Section 9.7 discusses the implications of the results of the analysis.

9.2 The concept of contrast effects in student feedback ratings

The body of research in the psychology literature that considers the extent to which evaluations of an individual's performance are made relatively to either the performance of others or to the previous performance of that individual are termed contrast effects. As was outlined in chapter 2, very little attention has been given to contrast effects in higher education. Where it has been utilised the research has generally taken an experimental design (see Murphy et al., 1985). There has been very little application of contrast effect theory to actual performance measures in higher education such as student feedback data. This chapter provides a systematic framework from which to assess the degree to which lecturers' student feedback ratings are influenced by the performance of those with whom they lecture on a particular module.

A search of the BIDS journal article database found no articles within the student feedback literature that dealt with the impact of one lecturers' feedback ratings on another's ratings on the same module. This was initially surprising given the amount of statistical based research conducted on student feedback ratings particularly in the U.S. One possible explanation is that most modules in the US may be taught by a single lecturer. By contrast, many modules in the Business School at Loughborough University are taught by more than one lecturer. As a result, the available data provides a good opportunity to test whether student feedback ratings are influenced by the other lecturers teaching on the same module.

Initial evidence of a contrast effect in lecturers' student feedback ratings was highlighted in the cluster analysis performed on a set of 65 modules presented in chapter 7. A good illustration of a contrast effect in student feedback ratings is provided by a comparison of two modules, A085~97 and B120~97. The comparison of these two modules is facilitated through the fact that they are similar in several ways. They are both quantitative based modules, they are both taught within the 'Management Science' section of the department, they are both taught by three lecturers and they are both compulsory (for slightly different combinations of degree programmes in the department). However, A085~97 is a first year undergraduate module and B120~97 is a second year undergraduate module.

Firstly, consider module A085~97. Table 9.1 presents the minimum, maximum and mean score for each of the three lecturers over the six lecturer-specific questions on the feedback form. The corresponding statistics for the module-specific questions are also presented. Examination of Table 9.1 clearly illustrates that lecturer C is the lowest scoring lecturer on the module. Lecturer C's mean score across the six lecturer-specific questions is 3.44 compared to 3.74 (lecturer A) and 4.10 (lecturer B).

Table 9.1 Descriptive statistics for module A085~97

	Minimum	Maximum	Mean	Std. Dev.
Questions 1-12	3.08	4.13	3.75	0.35
Lecturer A	3.47	4.17	3.74	0.26
Lecturer B	3.75	4.35	4.10	0.24
Lecturer C	**3.18**	**3.87**	**3.44**	**0.27**

The descriptive statistics for module A085~97 can be contrasted with those for module B120~97. Module A085~97 also features lecturer C, but on this occasion lecturer C teaches alongside two other lecturers, D and E. Table 9.2 presents the minimum, maximum and mean score for each of the three lecturers over the six lecturer-specific questions on the feedback form. It can be clearly seen that lecturer C's ratings are far better in B120~97 compared to A085~97. This is both in terms of numerical ratings (mean score of 3.92 in B120~97 compared to 3.44 in A085~97) and also in terms of lecturer C's ratings relative to the other lecturers on the module. In A085~97 lecturer C is the lowest scoring of the three lecturers, whereas in B120~97 lecturer C is the highest scoring of the three lecturers.

Table 9.2 Descriptive statistics for module B120~97

	Minimum	Maximum	Mean	Std. Dev.
Questions 1-12	2.72	4.07	3.42	0.45
Lecturer C	**3.52**	**4.15**	**3.92**	**0.21**
Lecturer D	3.33	4.12	3.74	0.29
Lecturer E	2.28	3.58	3.07	0.51

From their teaching generally, it is know that lecturers A and B tend to obtain good feedback ratings and lecturers C, D and E tend to obtain less good feedback ratings. Hence in A085~97 lecturer C is teaching alongside two lecturers who regularly score higher ratings than C does. However in B120~97 lecturer C teaches alongside two lecturers who generally obtain similar student feedback ratings to lecturer C. In this context any weaknesses of lecturer C are more likely to be highlighted in module A085~97 when teaching with two highly rated lecturers compared to module B120~97 when teaching with two equally rated lecturers.

This example suggests the following general issue. Consider a hypothetical lecturer who is generally considered of average teaching ability. On some occasions this lecturer will teach on a module alongside a lecturer who is generally considered to be substantially better than average, with student feedback ratings regularly in the top quartile of overall performance ratings. On other occasions this lecturer will teach on a module alongside a lecturer who is generally considered to be substantially worse than average, with student feedback ratings regularly in the bottom quartile of overall performance ratings. The issue of interest is whether, and if so to what extent, this lecturer's student feedback ratings will be influenced by the student feedback ratings given to the other lecturer(s) teaching on a particular module.

9.3 The development of the predictor variables used to measure contrast effects

The criterion for illustrating a contrast effect can be stated formally as:

If a lecturer is teaching with someone who is usually perceived to be better than they are, their ratings will be lower than their norm as they are being evaluated relative to the other lecturer(s) on that module. Conversely, if they are teaching with someone who is usually perceived to be worse than they are, their ratings will be higher than their norm as they are being evaluated relative to the other lecturer(s) on that module.

Table 9.3 presents a schematic representation of the data set used in the analysis. It is based on the master module level data set outlined in section 5.4. The data set includes all modules that have two or three lecturers. As the objective is to examine the extent to which the performance of other lecturer(s) affects the ratings of a lecturer, modules that were taught by a single lecturer were not included in the data set. The data set consisted of 15 columns and 296 rows. There were 188 cases (rows) relating to two-lecturer modules and 108 cases (rows) relating to three-lecturer modules. Not every module taught by two or three lecturers was included in this analysis. Modules that had a lecturer who was not a regular member of faculty were omitted from the data set.

Table 9.3 A schematic representation of the data set

Lecturer 1	Various descriptive variables*	Lecturer code	Rat.L.Hrs.	Q17	**Lecturer average Q17**	**DIFF**	**DIFOTL**
Lecturer 2	Same descriptive variables	Lecturer code	Rat.L.Hrs.	Q17	**Lecturer average Q17**	**DIFF**	**DIFOTL**

*Descriptive variables: year, module code, LEV1, LEV2, LEV3, LEV4, BS/NonBS, Reg.Studs.

To run the proposed analysis three new columns (shown in bold in Table 9.3) were added to the data set. To create the variable representing the difference between a lecturer's score on a particular module to their overall average over all the modules on which they lectured (DIFF) it was necessary, as a first step, to add the lecturer's average to the data set. This was computed by using the subtotal facility in Microsoft Excel. The variable DIFF was then calculated by subtracting a lecturer's average score from their score on a particular module.

The next step was to consider the most appropriate predictor variable that could be used to model the dependent variable DIFF. One possibility was to construct a variable that represented the difference between a lecturer's overall average and the overall average of the other lecturer(s) on the module (termed DIFOLAV - difference of lecturer's averages). This would give a measure of which lecturer is generally perceived to be better overall. Modelling DIFF using DIFOLAV would therefore give a measure of the extent to which a lecturer's feedback ratings vary depending on whether they are generally better or worse than the other lecturer(s) on that module.

An alternative predictor variable could be constructed by taking the raw score for a lecturer and subtracting it from the raw score of the other lecture(s) on that module (termed DIFOTL - difference between the two lecturers).

There are two key differences between the variables DIFOTL and DIFOLAV. Firstly, the variable DIFOLAV has the drawback of being constant for any pair of lecturers. In situations in which the same two lecturers teach together on more than one module (as often happens) the value for DIFOLAV will be the same on each occasion. This will be the case despite the fact that the lecturers could be evaluated quite differently to each other on the different modules. The variable DIFOTL, in contrast, is free from this constraint and will vary according to the actual performance of the lecturers on each module in which they teach.

This variation may reveal more about the degree to which the contrast effect influences ratings. This will be particularly so for lecturers whose scales have greater variation. Some lecturers have particularly large standard deviations with scores for Q17 varying considerably.

A second difference between the variables DIFOTL and DIFOLAV is that the variable DIFOTL is measuring the difference between lecturers on a particular module. The information is predominantly describing actual performance on a specific module rather than a general (overall) measure of a lecturer over a range of situations. We would expect the variable DIFOLAV to be weaker as a predictor variable because the information built into the variable DIFOLAV is the amalgam of varying circumstances that can in themselves influence ratings lecturers receive. Students are asked to judge the performance of their lecturers on particular modules, rather than giving their general impression of the lecturer over the range of exposure that the student could potentially have had to them. If students were to give ratings based on general impressions involving experiences not specific to the module being evaluated then the charge that student feedback exercises are simply popularity contests could hold some credence. The variation in ratings for individual lecturers suggests that students *are* discriminating in their evaluations of lecturers in differing circumstances. These issues increase the potential usefulness of the variable DIFOTL. Regression models for both DIFOTL and DIFOLAV were run. The models are discussed in section 9.4.

9.4 Regression models of DIFF modelled against the alternative predictor variables

9.4.1 DIFF modelled against DIFOTL

Having considered two possible predictor variables that could be used as a measure of contrast effects in student feedback ratings, both were used in turn. In this section the results of modelling DIFF against DIFOTL are discussed. The results of modelling DIFF against DIFOLAV are discussed in section 9.4.2.

Three models, one for two-lecturer modules, another for modules with three lecturers and a third combined model were calculated separately. This permits comparisons between modules with two and three lecturers. For modules with three lecturers the variable DIFOTL consists of the average of the scores of the other two lecturers on the module.

The model for modules with two lecturers is presented as equation 9.1.

Predictor variable (1): DIFOTL.

$$\textbf{DIFF Q17} = .06894 \quad + .258\ \text{DIFOTL} \qquad [9.1]$$

$$\text{t-values:} \quad (2.99) \quad (8.579)$$

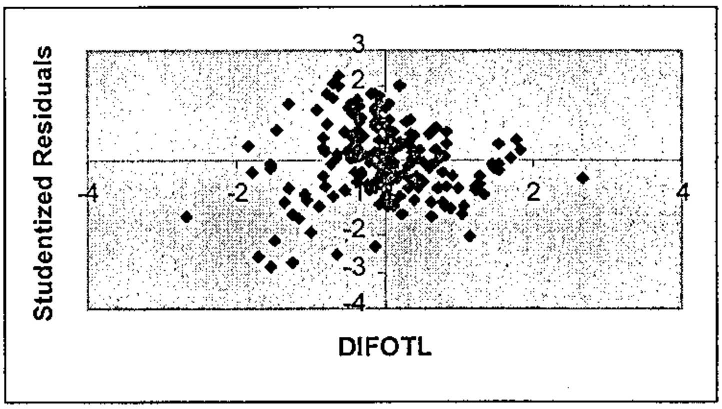
$R^2 = .284$ F = 73.597 (p = .000) n = 188

Equation 9.1 indicates that there is a significant contrast effect in the student feedback ratings for modules that are taught by two lecturers. The model explains nearly 30% of the variance in the dependent variable (i.e. the overall effectiveness of the lecturer). The constant is significant at the 1% level, but the impact of the variable DIFOTL is significantly larger with a t-value approaching 9. This supports the hypothesis that there is a contrast effect in student feedback ratings, whereby the performance of the other lecturer on a particular module has a significant impact on the student feedback ratings that both lecturers receive.

The next step was to plot the residuals against the predictor variable DIFOTL. Residuals are the difference between the observed and predicted values for the dependent variable. In examining the residual plot one hopes to see the residuals distributed randomly. If the residuals are distributed randomly then there is no evidence of a pattern in the values of the predictor variable that had not been reflected in the regression equation. A pattern in the residuals could show that the relationship between the variable DIFF and the variable DIFOTL was non-linear. The studentized residuals were used, as clear rules are available for deciding on which residuals are particularly large.

The scatterplot of the relationship between the studentized residuals and DIFOTL is presented in Figure 9.1.

Figure 9.1 Residual plot of DIFOTL in modules that have two lecturers



The points in the graph are relatively randomly distributed around both axes. From this scatterplot it can be concluded that the relationship between the predictor variable DIFOTL and the dependent variable DIFF is relatively linear.

Equation 9.2 presents the equivalent model for modules that have three lecturers. In this case the variable DIFOTL is the average for the other two lecturers on the module.

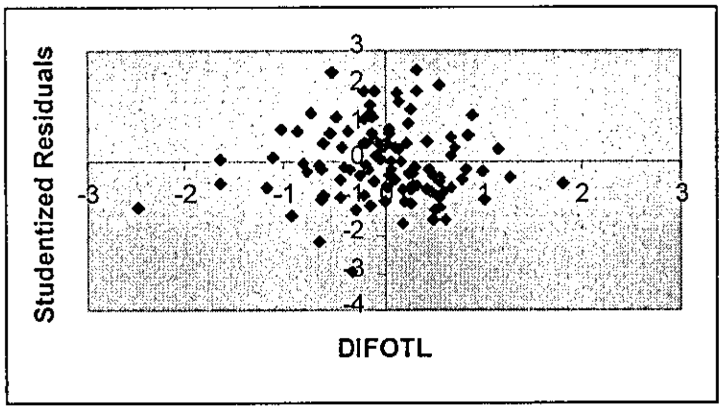
Predictor variable (1): DIFOTL.

DIFF Q17 =	.006736	+ .311 DIFOTL	[9.2]
t-values:	(.209)	(6.066)	

$R^2 = .258$ F = 36.791 (p = .000) n = 108

As in equation 9.1, there is significant evidence that DIFF depends on the predictor variable DIFOTL. Compared to equation 9.1, the t-value for DIFOTL in equation 9.2 is not quite as high. This could simply be a sample size effect reflecting the fact that there are fewer cases in equation 9.2 compared to equation 9.1. At any rate, at 6.1 the t-value for DIFOTL in equation 9.2 remains highly significant. The R^2 is similar and again the F-value indicates that there is significant evidence of a relationship at the 1% level. In contrast to equation 9.1 the constant is not significant.

Figure 9.2 Residual plot of DIFOTL in modules that have three lecturers



As for equation 9.1, the plot of the studentized residuals against the variable DIFOTL shows a relatively random distribution of the points, suggesting that relationship between the predictor variable DIFOTL and the dependent variable DIFF is relatively linear.

Finally, the combined model for modules with either two or three lecturers is presented as equation 9.3. For the three-lecturer modules the variable DIFOTL is the average for the other two lecturers on the module.

Predictor variable (1): DIFOTL.

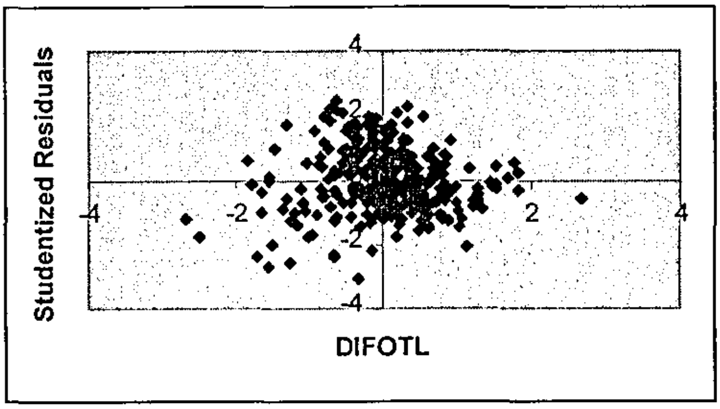
$$\textbf{DIFF Q17} = .04132 \quad + .273\ \text{DIFOTL} \qquad [9.3]$$

$$\text{t-values:} \quad (2.189) \qquad (10.403)$$

$R^2 = .269$ $F = 108.223$ $(p = .000)$ $n = 296$

As with the previous two models, equation 9.3 explains over a quarter of the variation in the dependent variable (i.e. the lecturer's overall effectiveness). The combined model has the largest F-value of all the three equations and hence the largest t-value for DIFOTL of all the three equations. This is a reflection of the fact that equation 9.3, by definition, has a larger sample size than equations 9.1 or 9.2 as it includes both two and three-lecturer modules. The constant is significant at the 5% significance level but not at the 1% level.

Figure 9.3 Residual plot of DIFOTL in modules that have two or three lecturers



In line with equations 9.1 and 9.2 the scatterplot of the residuals illustrate a random distribution of points which suggests that the relationship between the predictor variable DIFOTL and the dependent variable DIFF is relatively linear.

9.4.2 DIFF modelled against DIFOLAV

The dependent variable DIFF was also modelled against the alternative predictor variable measuring contrast effects DIFOLAV. As for DIFOTL, separate models were run for two-lecturer modules and three-lecturer modules. A combined model was also run. There was no evidence, significant at the 1% level, that DIFF was dependent on DIFOLAV for any of these models.

9.5 Expansion of model to include additional predictor variables

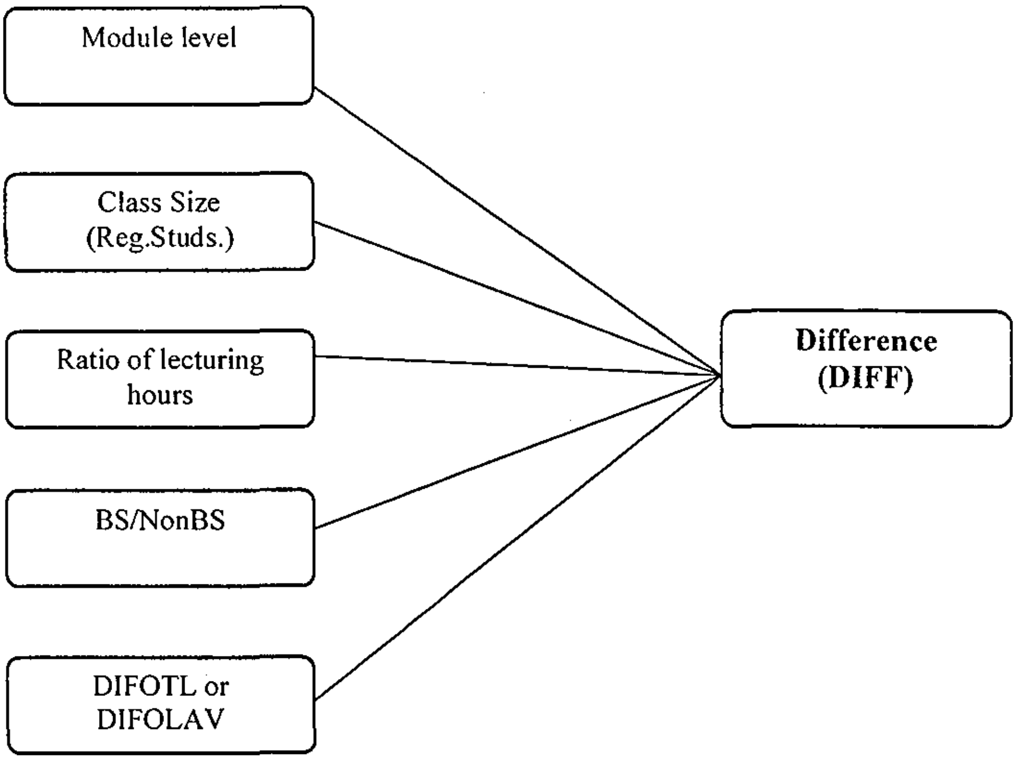
The models presented in section 9.4 proved that there is a contrast effect in the student feedback ratings between lecturers on a particular module. In this section these models are extended to examine how the contrast effect is moderated by other predictor variables that can be postulated as affecting student feedback ratings. Four other predictor variables were introduced into the model. These were:

a) Module level;
b) Class size (Reg.Studs.);
c) Ratio of lecturing hours;
d) BS/NonBS.

Chapter 3 outlined the relationship between module level and class size on student feedback ratings in the published literature on student feedback ratings. Both of these predictor variables have been shown to influence the student feedback ratings that lecturers receive. The ratio of lecturing hours undertaken by the different lecturers on the module is included given the evidence presented in chapter 8. The final additional variable BS/NonBS is dichotomous differentiating between modules taught by Business School staff to Business School students and modules taught by Business School staff to non-Business school students. It is important to include this variable in the model for two reasons. Firstly, the Business School teaches many modules to students in other departments and secondly, the student feedback literature reports significant differences in the average student feedback ratings between academic disciplines (see section 3.6).

This extended model is illustrated diagrammatically in Figure 9.4.

Figure 9.4 Extended model of factors affecting DIFF



This conceptual model is a sub-model of the main model, which will be presented in chapter 10. The dependent variable in this model is again the difference between an individual lecturer's average for an individual module and that lecturer's overall average (i.e. DIFF is used as the dependent variable). The subject area to which the module belonged was not included as a predictor variable in this model. This is because an individual lecturer will regularly teach in only one subject area; i.e. an accountant will rarely lecture on a human resource management module. In chapter 10 the actual rating scores are used as the dependent variable, this allows subject area to be meaningfully used as a predictor variable.

9.6 Regression models of expanded model

9.6.1 Expanded model using DIFOTL as the predictor variable measuring contrast effects

The dependent variable DIFF was modelled against the predictor variables described in section 9.5. In this section the contrast effect predictor variable used was DIFOTL and hence the model obtained can be compared to the models in section 9.4.1. The model for modules with two lecturers is presented as equation 9.4.

Predictor variables (7): LEV2, LEV3, LEV4, Reg.Studs., NonBS, Rat.L.Hrs. and DIFOTL.

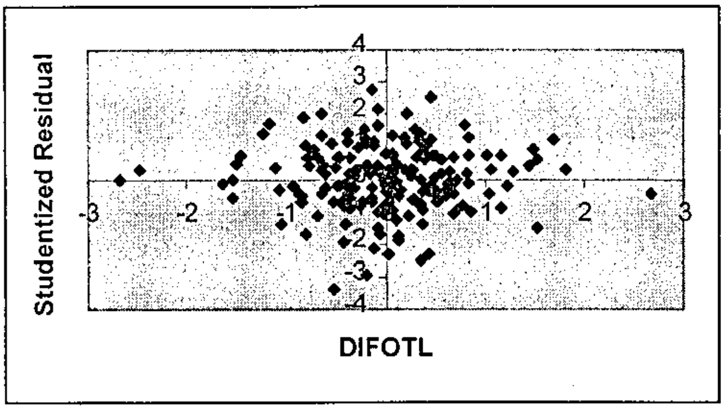
$$\begin{array}{lcccccccc}
\textbf{DIFF Q17} = & .247 & - & .215\ \text{LEV2} & - & .00222\ \text{Reg.Studs.} & + & .258\ \text{DIFOTL} & [9.4] \\
\text{t-values:} & (6.744) & & (-3.938) & & (-4.893) & & (9.412) &
\end{array}$$

$R^2 = .411 \quad F = 42.83 \quad (p = .000) \quad n = 188$

The value of R^2 indicates that the model explains 41% of the variance in the dependent variable. The F-value is significant at the 1% level. Only three of the predictor variables are significant in the model, the other four are not significant at the 5% level. DIFOTL is the variable that has the greatest impact on the dependent variable. The coefficient and the t-value for DIFOTL in equation 9.4 are very similar to the corresponding values in equation 9.1. As equation 9.1, equation 9.4 used only modules taught by two lecturers. In equation 9.1 DIFOTL was the only predictor variable.

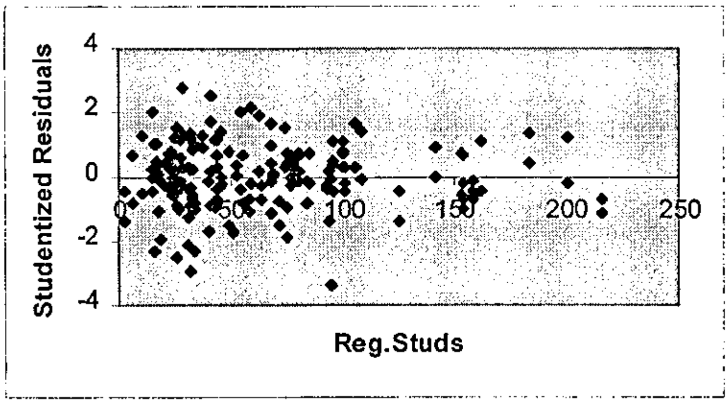
Figures 9.5, 9.6 and 9.7 present the residual plots for each of the predictor variables that are included in the final model. Again the studentized residuals were used. In each case the plots broadly show that the residuals are randomly distributed with no gross outliers. This provides reassurance that the model presented as equation 9.4 does reliably represent the relationship between the dependent variable and the predictor variables.

Figure 9.5 Residual plot of DIFOTL in modules that have two lecturers



The plot of the studentized residuals against the variable DIFOTL (Figure 9.5) shows a relatively random distribution of the points, suggesting that relationship between the predictor variable DIFOTL and the dependent variable DIFF is relatively linear.

Figure 9.6 Residual plot of Reg.Studs. in modules that have two lecturers



The rather different spread of residuals in Figure 9.6 simply reflects the distribution of Reg.Studs. in the data set. There are few residuals plotted for values of Reg.Studs. greater than 100 because there are few modules with class sizes greater than 100. In any case there appears to be no link between the size of the residuals and the class size (i.e. there is no evidence of heteroskedasticity).

Figure 9.7 Residual plot of module level in modules that have two lecturers

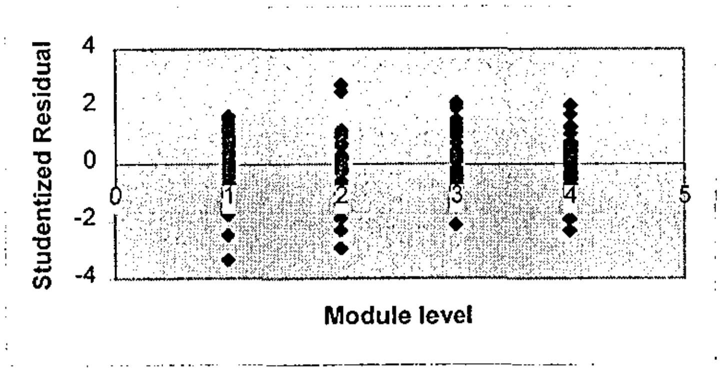


Figure 9.7 plots the residuals for each of the four values for module level. There is a similar distribution of residuals for each of the four values for module level.

Equation 9.5 presents the regression model for modules with three lecturers.

Predictor variables (7): LEV2, LEV3, LEV4, Reg.Studs., NonBS, Rat.L.Hrs. and DIFOTL.

DIFF Q17 = .215 - .213 LEV3 + .373 NonBS -. 0 029 Reg.Studs. + . 312 DIFOTL [9.5]
t-values: (3.915) (-2.976) (2.726) (-4.707) (6.797)

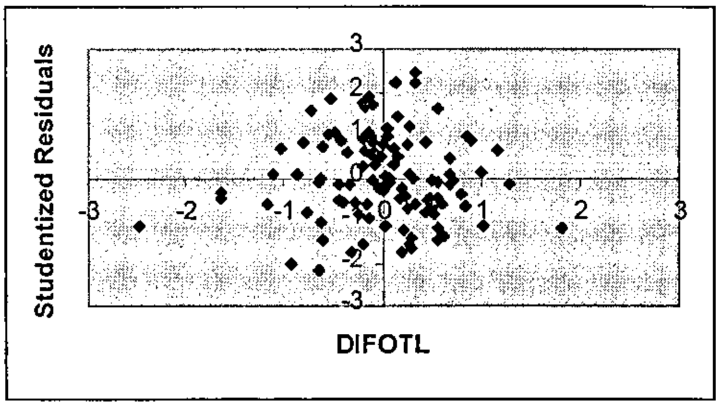
$R^2 = .425$ F = 19.05 (p = .000) n = 108

The value of R^2 indicates that the model explains 42.5% of the variance in the dependent variable. As for equation 9.4, the F-value is significant at the 1% level. In equation 9.5, four predictor variables are significant in the final model. As in equation 9.4, DIFOTL is the predictor variable that has the greatest impact on the dependent variable. The coefficient and the t-value for DIFOTL in equation 9.5 are very similar to the corresponding values in equation 9.2. As equation 9.5, equation 9.2 used only modules taught by three lecturers. In equation 9.2 DIFOTL was the only predictor variable.

The regression model for modules with three lecturers (equation 9.5) displays two key differences from equation 9.4. Firstly, whereas LEV2 (second year undergraduate) appears in equation 9.4, in equation 9.5 this is replaced by LEV3 (final year undergraduate). The second key difference is that in equation 9.5, NonBS (modules taught by Business School lecturers to non-Business School students) is a significant predictor variable of DIFF.

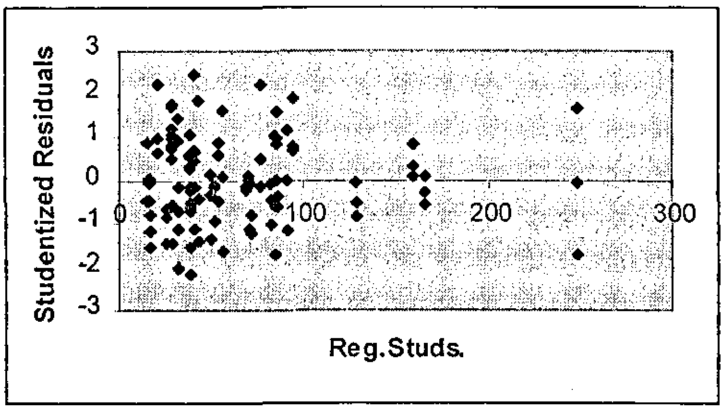
Figures 9.8, 9.9, 9.10 and 9.11 present the residual plots for each of the predictor variables that are included in the final model. Again the studentized residuals were used. In each case the plots broadly show that the residuals are randomly distributed with no gross outliers. Hence there is no evidence that any of the assumptions that underpin the regression analysis have been violated. This provides reassurance that the model presented as equation 9.5 does reliably represent the relationship between the dependent variable and the predictor variables.

Figure 9.8 Residual plot of DIFOTL in modules that have three lecturers



As with the previous residual plots of DIFOTL this scatterplot suggests that the relationship between the predictor variable DIFOTL and the dependent variable DIFF is relatively linear.

Figure 9.9 Residual plot of Reg.Studs. in modules that have three lecturers



The spread of residuals in Figure 9.9 simply reflects the distribution of Reg.Studs. in the data set. There are few residuals plotted for values of Reg.Studs. greater than 100 because there are only four modules taught by three lecturers where the class size is over 100. There appears to be no link between the size of the residuals and the class size (i.e. there is no evidence of heteroskedasticity).

The graph clearly depicts four sets of points that are detached from the main body of points; these represent four different modules. One of these sets the three cases representing the module with the greatest class size (248 registered students) was deleted from the analysis. This allowed an examination of whether any of these residual points had a significant influence on the regression model. The removal of these three cases had a minimal effect on the regression model. This is represented by the fact that the same predictor variables were significant in the regression model with only a slight change in their coefficients. Therefore these points did not significantly influence the regression model.

Figure 9.10 Residual plot of BS/NonBS in modules that have three lecturers

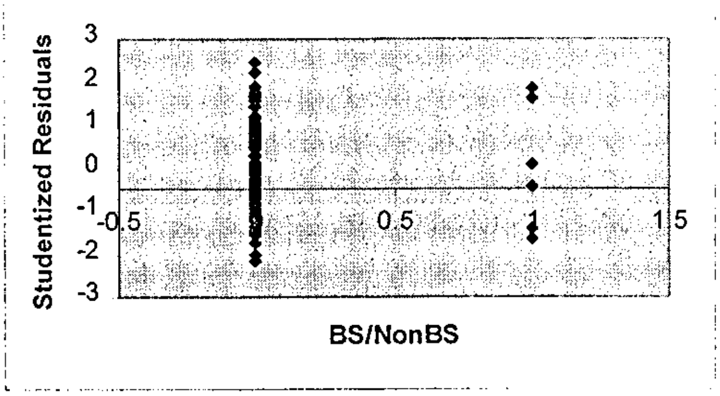


Figure 9.10 presents the residual plot of BS/NonBS in modules that have three lecturers. For both modules taught to Business School students (0 on the x-axis) and those taught to non-Business School students (1 on the x-axis) there is a relatively high degree of symmetry around the x-axis. This illustrates a random distribution of the residual points and suggests that the relationship between the predictor variables BS/NonBS and the dependent variable DIFF is relatively linear.

Figure 9.11 Residual plot of module level in modules that have three lecturers

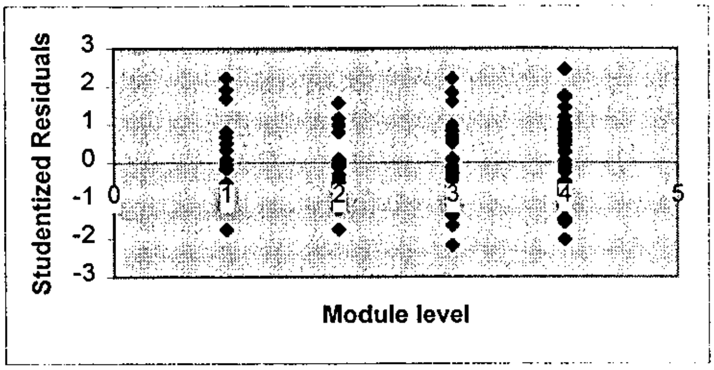


Figure 9.11 plots the residuals for each of the four values for module level. There is similar distribution of the residuals for each of the four values for module level. This

therefore suggests that the relationship between the LEVEL predictor variables and the dependent variable DIFF is relatively linear.

A combined model for modules that have either two or three lecturers is presented as equation 9.6. As in equation 9.3, for the three-lecturer modules the variable DIFOTL is the average for the other two lecturers on the module.

Predictor variables (7): LEV2, LEV3, LEV4, Reg.Studs., NonBS, Rat.L.Hrs. and DIFOTL.

DIFF Q17 = .209 - .155 LEV2 -. 00219 Reg.Studs. + .237 DIFOTL [9.6]
t-values: (7.068) (-3.497) (-5.86) (11.269)

$R^2 = .381$ F = 59.95 (p = .000) n = 296

The value of R^2 indicates that the model explains 38% of the variance in the dependent variable. As for equations 9.4 and 9.5, the F-value is significant at the 1% level. Only three of the predictor variables are significant in the model, the other four are not significant at the 5% level. These are the same variables as in equation 9.4. This is not surprising as equation 9.4 (two-lecturer modules) has a larger sample size than equation 9.5 (three-lecturer modules), 188 cases representing 94 modules compared to 108 cases representing 36 modules. Therefore the combined model equation 9.6 (consisting of both two and three-lecturer modules) would be expected to more closely resemble equation 9.4 than equation 9.5.

In line with both equation 9.4 and equation 9.5, DIFOTL is the predictor variable that has the greatest impact on the dependent variable.

The coefficient and the t-value for DIFOTL in equation 9.6 are very similar to the corresponding values in equation 9.3. As equation 9.6, equation 9.3 used both two and three-lecturer modules. In equation 9.3 DIFOTL was the only predictor variable. The residual plots for each of the three indicator variables significant in equation 9.6

were very similar to the corresponding residual plots for equation 9.4 (i.e. Figures 9.5, 9.6 and 9.7).

9.6.2 Expanded model using DIFOLAV as the predictor variable measuring the contrast effect

In addition to running the expanded model with DIFOTL used as a predictor variable, the expanded model was also run using the alternative contrast effect predictor variable DIFOLAV. To be consistent with the approach taken in modelling DIFF against DIFOLAV, three separate equations were created. The first for modules taught by two lecturers, the second for modules taught by three lecturers and finally a combined model for modules that had either two or three lecturers.

As with the models in section 9.4.2 where DIFF was modelled only against DIFOLAV, there was no significant evidence that DIFF was dependent on DIFOLAV for any of these models. As a result the models are simply summarised in Table 9.4.

Table 9.4 Regression equations of the expanded model using DIFOLAV as the predictor variable measuring the contrast effect

	R^2	F-value	**DIFF Q17 =**	Equation
2 lecturer modules	.128	13.537	.247 - .215 LEV2 - .0022 Reg.Studs.	9.7
3 lecturer modules	.167	6.972	.214 - .213 LEV3 + .373 NonBS - .0029 Reg.Studs.	9.8
Combined model	.126	14.014	.0991 - .153 LEV2 - .0022 Reg.Studs. + .252 Rat.L.Hrs.	9.9

An examination of Table 9.4 reveals that DIFOLAV was not a significant predictor of the dependent variable DIFF (at the 5% level) in any of the equations 9.7, 9.8 or 9.9. In each of these models DIFOLAV was eliminated at the second step of the backward elimination regression process.

9.6.3 Comparison of the variables DIFOTL and DIFOLAV

The importance of DIFOTL in the regression equations presented in sections 9.4 and 9.6, is influenced by the fact that the lecturer's rating for a particular module forms part of both the dependent variable (DIFF) and the predictor variable (DIFOTL). This is in contrast to the predictor variable DIFOLAV, which is constructed by computing the difference between a lecturer's overall average and the overall average of the other lecturer(s) on the module. Equations 9.7, 9.8 and 9.9 suggest that a lecturer's ratings on a particular module have a small effect on the predictor variable DIFOLAV. The usefulness of DIFOLAV as a predictor variable of student feedback ratings will be explored further in chapter 10.

DIFOTL is shown to be an important predictor of the dependent variable (DIFF) in each of the regression equations in which it was used. As discussed in section 9.3, one of the main features of DIFOTL is that it measurers the differences in the ratings between lecturers on a particular module. As a result, DIFOTL is module-specific, restricted to what students actually experience on a particular module. Compared to DIFOLAV, DIFOTL should be less influenced by overall factors, such as a student's experiences of a lecturer in a variety of situations over a period of time. The importance of DIFOTL in the various regression equations is encouraging for advocates of student feedback data. The importance of DIFOTL indicates that variations in lecturer's ratings reflect what happens within lecture theatres, rather than reflecting the generalised feelings of students to their lecturers.

9.7 Managerial implications

This analysis has illustrated that there is a clear contrast effect in the student feedback ratings received by lecturers on modules in which they share the lecturing load with one or two other lecturers. Lecturers' ratings are shown to be significantly influenced by the performance of those with whom they lecture. This has important implications for the appropriate interpretation of student feedback ratings in such circumstances.

In particular the evidence from this chapter shows that an individual lecturer's student feedback ratings may vary considerably for reasons unrelated to the lecturer's performance. Hence, it is important not to place too much weight on the student feedback ratings from any single module where the teaching load has been shared with one or two other lecturers. This conclusion is consistent with the student feedback literature. To increase the generalizability of a lecturer's student feedback ratings Cashin advises:

> 'when making personnel decisions...use ratings from a variety of courses from every term for at least two years, totalling at least five courses' (Cashin, 1995).

The findings in this chapter strengthen the need to be wary of reacting to the student feedback ratings from any one isolated module where the teaching load has been shared with one or two other lecturers.

Sometimes novice lecturers are paired with more experienced lecturers in order to gain experience of the system of lecturing and module management. Departmental mangers are likely to hope that this collaboration will result in the more experienced lecturer acting as a mentor to the novice lecturer and aid the novice lecturer in developing their teaching technique.

Managers need to be aware, however, of the potential implications of this arrangement in situations in which the novice lecturer is perceived by the students to be weaker compared to the more experienced lecturer. This would be particularly important if the novice lecturer scores ratings lower in the jointly taught module compared to the ratings they receive on modules in which they teach alone. It may be possible that that although the novice lecturer returns lower ratings, when teaching with the more experienced lecturer, their teaching had actually improved.

There is an issue here of whether the lecturer's actual performance has deteriorated by virtue of teaching alongside a better lecturer, or whether the performance has stayed relatively constant, with the students rating the performance lower than the lecturer's norm in direct comparison with the other lecturer. From a statistical point of view it would be very difficult to adequately untangle this relationship. What is important is that from a managerial point of view there are clear implications which hold irrespective of whether the actual performance of the lecturer is weaker or not. Student feedback ratings indicating that a lecturer's ratings are lower than their normal average may not conclusively tell us that the lecturer's actual performance is poorer than usual, such ratings do tell us that the students perceive it to be so. Under such circumstances there is a risk that students may lose confidence in the weaker lecturer and disregard the material that they teach the students. As a direct consequence, this could lead to additional pressure being placed on the better lecturer as the students may be more inclined to visit the stronger, rather than the weaker, lecturer. In modules in which examination questions are written by different lecturers to correspond to different sections of the module and students have a degree of choice as to which questions to answer the better lecturer may find that, as a result of the greater confidence that the students have in them, more of their questions will be answered. In extreme cases this has the deleterious implication that that students would not adequately cover the syllabus.

One of the central implications of the analysis presented in this chapter is that in modules in which the lecturing responsibilities are shared the student feedback ratings tend to exaggerate the differences between the lecturers. The lecturer perceived to be better than the other will receive ratings higher than their norm, while the lecturer perceived to be weaker by the students will receive ratings lower than their norm. This stresses the need for departmental managers to compensate for this while looking at lecturers' student feedback results.

For student feedback ratings to be accepted as an important and useful tool in assessing teaching quality and personnel decisions, it is important that all interested parties have confidence in the data itself and the managerial interpretation of it. The extent of the contrast effect in student feedback ratings shown in this chapter makes it imperative that no one module is considered in isolation. Lecturers' student feedback ratings may vary between modules based simply on those with whom they lecture on a particular module. As a result, managers within academic departments and personnel officers should base their judgements on the overall quality/suitability of a lecturer using student feedback ratings for a range of modules encompassing, if possible, a range of situations.

Chapter 10

A GENERAL MODEL OF LECTURERS' RATINGS

10.1 Introduction

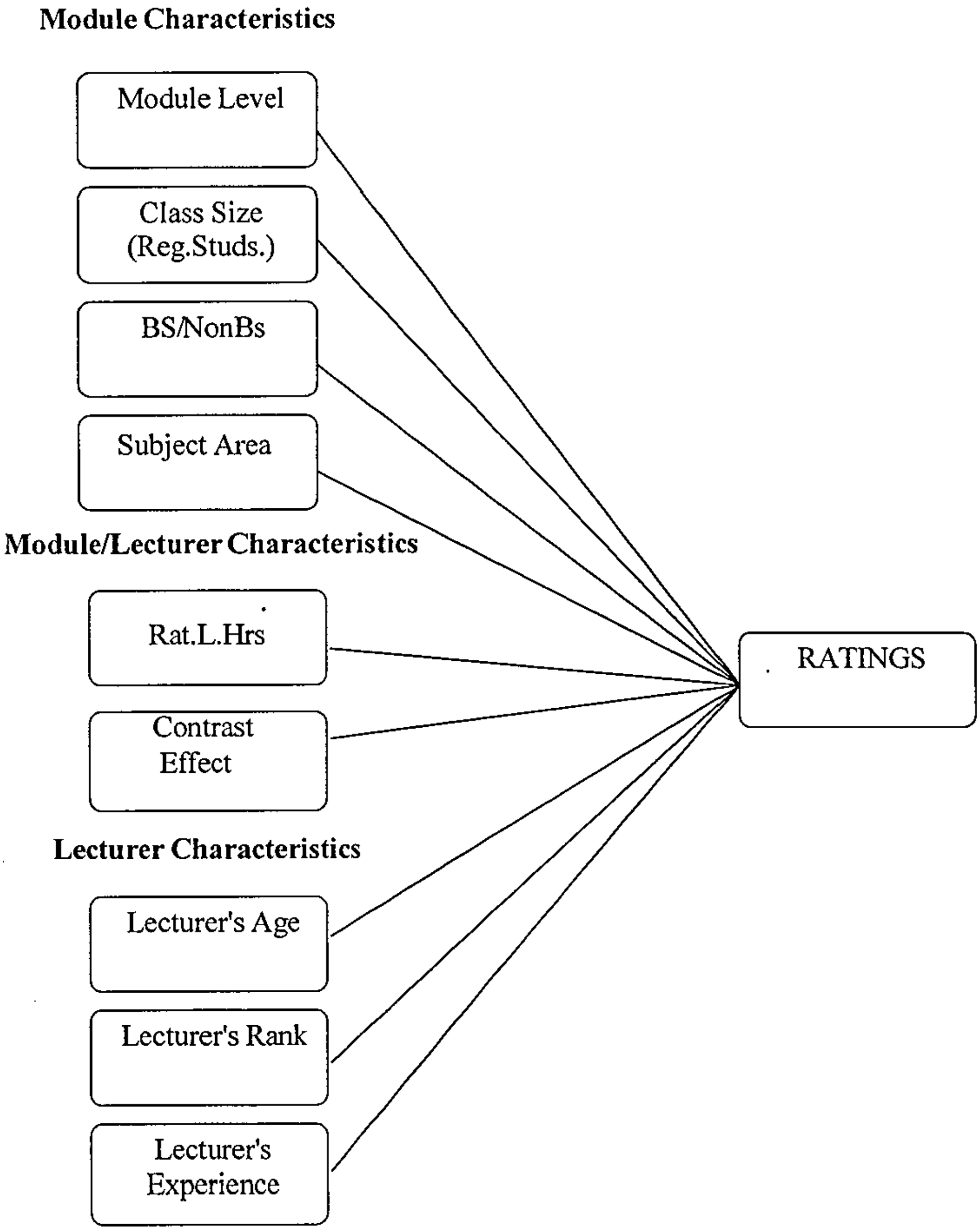
The objective of this chapter is to model lecturers' ratings against a range of predictor variables that have been shown either in the published student feedback literature or in previous chapters of the thesis to influence the ratings a lecturer receives. The aim in conducting this modelling is to determine which of the predictor variables influences lecturers' ratings in the Business School and to measure the relative importance of those that do. The findings can act as a guide for both individual lecturers and managers in the interpretation of a lecturer's ratings, allowing decisions made in light of the data to be better informed. Section 10.2 presents the model and discusses how it has been incrementally developed over the preceding two chapters. The rationale for a change in the dependent variable (compared to the models presented in chapters 8 and 9) is also discussed. Section 10.3 outlines the methods by which the data set used for this model was constructed. Section 10.4 details four different predictor variables that were used to measure contrast effects in student feedback ratings. The advantages and disadvantages of each of these variables are considered. In section 10.5 the expected effect of each of the predictor variables on student feedback ratings is considered. Section 10.6 presents the regression models for lecturers' ratings for each of the six lecturer-specific questions on the student feedback form (Q13-Q18) using one of the predictor variables that measure contrast effects. These models are discussed in terms of the expectations outlined in section 10.4. Section 10.7 presents a comparison of the effects of each of the four possible predictor variables used to measure contrast effects on a lecturer's student feedback ratings. Section 10.8 contrasts the findings from the modelling of the Business School student feedback data to the dominant trends in the student feedback literature.

Finally, section 10.9 presents the conclusions of the analyses and discusses the implication of these for managerial decision making.

10.2 Statement of model

Figure 10.1 presents the predictor variables used to model student feedback ratings. In particular it shows the groupings of the predictor variables into those that describe module-characteristics, those that describe lecturer-characteristics and those that describe a combination of the two. This is an augmented model that has been developed from the model constructed over the previous two chapters.

Figure 10.1 Model of the Loughborough University Business School student feedback data



The model presented as Figure 10.1 has its roots in Chapter 8, which considered whether the position that a lecturer appeared on the student feedback form influenced the ratings they received. Through a set of interviews it was hypothesised that it was not the position that a lecturer appeared on the form that influenced ratings but rather the ratio of lecturing hours undertaken by the lecturer on a particular module. Subsequent statistical analysis on the Business School student feedback data showed this to be the case. As a result the variable Rat.L.Hrs. has been included in the model. Chapter 9 introduced the concept of contrast effects in lecturers' student feedback ratings. It was shown that the lecturer is judged relatively to the performance of the other lecturer(s) in modules where the lecturing load is shared. Added to the model in chapter 9 alongside the ratio of lecturing hours (Rat.L.Hrs.) were other module-specific characteristics, the level of the module (LEVEL), the number of registered students on the module (Reg.Studs.) and whether the module was taught to Business School or non-Business School students (BS/NonBS).

The model presented as Figure 10.1 retains each of these elements built up previously. The use of lecturers' ratings as the dependent variable makes it possible to introduce lecturer-specific characteristics as well. As was outlined in chapter 3, various lecturer characteristics have been shown to influence the student feedback ratings received by lecturers. Research evidence paints an inconsistent pattern of findings, but the three lecturer-specific variables introduced in this chapter are often shown in the published literature to influence a lecturer's student feedback ratings. The three predictor variables introduced into the model are: the lecturer's age, the lecturer's rank and the lecturer's experience.

The dependent variable used in chapters 8 and 9 DIFF would have been unsuitable as the dependent variable in the model presented as Figure 10.1. DIFF measured the difference between a lecturer's average rating on a particular module and their overall average for a particular question. As DIFF measures the relationship between a lecturer's feedback ratings in a particular module and their norm, the lecturer-specific characteristics would not vary with DIFF. Once the actual student feedback ratings were used as the dependent variable there is no longer any problem of lack of variation between the lecturer-specific predictor variables and the dependent variable, this therefore allows the lecturer-specific variables to be added to the model.

10.3 Construction of the data set

This section describes the predictor variables used in the subsequent analysis. The predictor variables used to model contrast effects are described in section 10.4. The data set used in chapter 10 is an extension of the data set used in the previous chapter. In the current chapter the data for modules taught by a single lecturer are added to the data for modules taught by two or three lecturers. In total the data set consisted of 458 cases. The same descriptive variables used in chapter 9 were again used here (i.e. year, semester, module code, LEV1, LEV2, LEV3, LEV4, BS/NonBS, Reg.Studs., lecturer code and Rat.L.Hrs.). To these the following variables were added:

- the ten subject area code variables;
- lecturer's rank;
- lecturer's age;
- lecturer's experience;
- the average rating on the module for the lecturer on a particular question (e.g. Q13);
- the other lecturer's average rating on the module for the question (e.g. Q13);
- the difference between these ratings on the module;
- the lecturer's overall average rating over all the modules on which they lectured for the question (e.g. Q13);
- the overall average rating of the other lecturer over all the modules on which they lectured for the question (e.g. Q13);
- the difference between these overall averages.

This sequence was then repeated for each of the other five lecturer-specific questions.

As in previous chapters, for modules with three lecturers the 'other lecturer's rating' was calculated by taking the arithmetic mean of the relevant rating for the other two lecturers. A data value of zero was entered into the data set in the columns depicting the ratings of the other lecturer and the difference between lecturers for modules in which there was only one lecturer. By inserting a value of zero the case can be used in the analysis. Otherwise there would be a missing value in the data set and SPSS would not include the case in the regression analysis. A value of zero should not affect the coefficients of the resulting equations.

Lecturer's rank was coded as a dichotomous variable, defined as standard and senior lecturers. Senior lecturers included those who were ranked as 'senior lecturer' or higher (i.e. Reader or Professor). In total 28 lecturers were classified as standard and 17 lecturers were classified as senior.

Lecturer's age was coded as a continuous variable.

Lecturer's experience was calculated by the number of years service at Loughborough University. The literature focuses on the number of years lecturing experience, it was not possible to use this specific variable in this research as data for the number of years each lecturer had lectured in total were not available. The number of years service at Loughborough University is a potentially interesting variable as it can be used to reflect whether the degree of experience of lecturing specifically in the system in place at Loughborough University (i.e. a semesterised teaching system, with a general emphasis on applied knowledge) influences a lecturer's student feedback ratings.

10.4 Rationale for choice of predictor variable to measure contrast effects

In this section a number of different variables for measuring contrast effects are considered. It is necessary to consider various possible variables for measuring contrast effects, as the dependent variable used in this chapter is different from that used in chapter 9. In chapter 9 the dependent variable used (DIFOTL) measured the difference between a lecturer's average on a particular module and their overall average on a particular question over all the modules on which they teach. In modelling DIFOTL the lecturer-specific characteristics could not sensibly be included in the model. This reflects the fact that the lecturer-characteristics did not change. In chapter 10 the aim is to use the actual ratings received by a lecturer for a module rather than computing the difference between this rating and the lecturer's overall average (i.e. the dependent variable DIFF used in chapters 8 and 9). This allows the model to be extended to include lecturer-specific characteristics.

Four possible variables that could be used to measure the contrast effect were considered. Each of the four possible variables will be discussed in turn in the context of the following hypothetical example. For a particular question on a module taught by two lecturers, suppose lecturer 1 received a rating of 3.8 and lecturer 2 received a rating of 3.6. Suppose also that the average rating for this question over all the modules on which they lecture is 3.7 and 3.5 for lecturer 1 and 2 respectively.

Option 1: OTLES

This option would predict a lecturer's average rating on a module using the average rating of the other lecturer on this module for the same question. So in our example lecturer 1 who receives a rating of 3.8 for a particular question on this module teaches alongside a lower scoring lecturer who receives a rating of 3.6 for the same question on this module. The central hypothesis in the contrast effects research presented in chapter 9 was that when a lecturer teaches alongside a lecturer who is perceived to be weaker than they are, then that lecturer's rating should be higher than their norm. In using OTLES we do not have a direct measure of how either lecturer has been rated on this module compared to their mean across all modules. However we can be sure of the expected direction of the relationship. For the hypothesis to hold, if lecturer 2's rating had been lower (say 3.4) then we would expect lecturer 1's rating to be higher. Therefore we would expect a negative relationship (i.e. the lower the rating for lecturer 2, the higher the expected rating for lecturer 1). The drawback of OTLES is that it allows no comparison of how two or more lecturers on a module compare to each other. The result of this is that OTLES is unlikely to explain much of the variance in lecturers' ratings.

Option 2: DIFOTL

This option would predict a lecturer's average rating on a module using the difference between the lecturers' ratings on a particular module. This was the variable used to measure contrast effects in chapter 9. So in our example the model for lecturer 1 (rating 3.8) would include the difference of the two lecturer's ratings (i.e. 3.8 - 3.6 = 0.2). The bigger the difference in ratings between the two lecturers the bigger the corresponding value of DIFOTL will be. Following the logic discussed under OTLES (i.e. when a lecturer teaches alongside a lecturer who is perceived to be weaker than they are, that lecturer's rating should be higher than their norm) we would expect that the greater the magnitude of the ascendancy of lecturer 1 over lecturer 2, the higher the rating for lecturer 1 should be. Therefore we would expect DIFOTL to have a positive coefficient.

Options 1 and 2 are similar in that they are restricted to the information that the respondents have for a particular module. In other words, they are a measure of the differing perceptions of the students for the two lecturers on the module. Using OTLES or DIFOTL assumes that ratings are not a product of knowledge of, or judgements made about, lecturers in any context other than their lecturing performance on the module under consideration. This can be contrasted with DIFOLAV and OLSFA below.

Option 3: DIFOLAV

This option would predict a lecturer's average rating using the overall differences in the average rating of the lecturers over all the modules in which they teach. So in our example the model for lecturer 1 (rating 3.8) would include the difference in the averages of the two lecturer's ratings over all the modules on which they lecture (i.e. 3.7 - 3.5 = 0.2). For the same reasons as outlined under DIFOTL, it is expected the coefficient of DIFOLAV would be positive.

DIFOLAV is describing something fundamentally different from DIFOTL. Whereas DIFOTL presumes that a lecturer's ratings are based on the judgements of the students on a particular module, DIFOLAV builds into the model the 'overall' evaluations of students about a particular lecturer. So DIFOTL is measuring the difference between the two lecturers on a particular module. DIFOLAV is measuring the difference between the two lecturers in general.

Option 4: OLSFA

This option would predict a lecturer's average rating using the difference between the other lecturer's average rating on a module and their average over all the modules on which they teach. So in our example, lecturer 1's rating on the module (3.8) will be, in part, predicted by lecturer 2's overall average (3.5) subtracted from lecturer 2's rating on this particular module (3.6). Therefore the value of OLSFA would be 0.1. If OLSFA is positive, lecturer 2 is performing better than their average. The variable OLSFA is expected to have a negative coefficient as an improvement in lecturer 2's ratings relative to their norm should (under the concept of contrast effects) be at the expense of the other lecturer on a particular module.

The main problem with both DIFOLAV and OLSFA is that the measure of any contrast effect would be confounded with the other module-specific predictor variables. For example the lecturer's overall average would be based on modules that would vary in respect to level, class size, different combinations of ratio of lecturing hours and whether modules were taught inside or outside the department.

10.5 The expected effect of each of the predictor variables on student feedback ratings

Using the predictor variables presented in Figure 10.1, a model was built for each of the following lecturer-specific question separately.

- Q13 The lecturer was well prepared;
- Q14 The lecturer was enthusiastic about the subject;
- Q15 The lecturer explained the subject clearly;
- Q16 The lecturer encouraged participation and questions;
- Q17 The lecturer overall, was effective;
- Q18 The lecturer could be contacted for advice if needed.

Building a model for each of the lecturer-specific questions enables us to study which predictor variables have the greatest impact on each question. The six lecturer-specific questions each represent a different aspect of the lecturing process and lecturers' responsibilities. Lecturers' overall average ratings tend to differ substantially over the six questions. This phenomenon is encouraging as it suggests that the students are discriminating in their responses to the six lecturer-specific questions, rather than giving a lecturer similar ratings for each lecturer-specific question based on whether they generally like or dislike a lecturer. This analysis is expected to show that the importance of the various predictor variables substantially differs over the six questions.

The expectations regarding each of the predictor variables are outlined below.

Module level

The expectation is that the effect of module level will vary over the six lecturer-specific questions.

Hypothesis 1: *The module level is unlikely to have a particularly substantial impact on ratings for a lecturer's preparation (Q13) or a lecturer's clarity (Q15).*

This is because these lecturing traits are a reflection of the previous effort and developed skill of the lecturer; the level of the module should not particularly affect this. In some ways it may be easier to prepare for first year modules as the material may be more predictable or 'standard'. Alternatively, it may be the case that some first year modules are taught as 'service' modules, or by less experienced personnel. This effect on ratings should not be profound, as the limited amount of data available on personnel who were not regular members of faculty was not included in this research.

Hypothesis 2: *The module level is likely to have a particular substantial impact on ratings for a lecturer's enthusiasm (Q14) and a lecturer's encouragement of participation (Q16).*

Hypothesis 3: *LEV3 (i.e. final year undergraduates) and LEV4 (i.e. taught course postgraduates) are expected to be stronger predictors of a lecturer's enthusiasm (Q14) and a lecturer's participation (Q16) compared to the four other lecturer-specific questions.*

Finalists (who will have spent a year "out" in industry) and taught course Master's students may have a different perspective on university education to first and second year undergraduates. Experience of industry may have changed the focus of finalists and Master's students and this might be reflected in the ratings that they give their lecturers. These students may place greater emphasis on more dynamic aspects of interaction, e.g. encouraging participation and questions and the enthusiasm of the lecturer.

Hypothesis 4: *There is likely to be a particularly strong effect in the predictive effect of LEV4 on a lecturer's availability (Q18).*

This is likely to be a strong negative relationship compared to LEV2 and LEV3 against Q18 and compared to the strength of LEV4 in predicting lecturers' ratings on the other five lecturer-specific questions. For those who are studying part time (particularly those reading for MBA's) opportunity to speak to the lecturer is likely to be limited. This is likely to be reflected in lower ratings for lecturers' availability (Q18).

Reg.Studs.

Hypothesis 5: *Reg.Studs. is expected to be a weaker predictor of student feedback ratings for preparation (Q13) compared to the other five lecturer-specific questions.*

As discussed under level the lecturer's preparation (Q13) reflects the previous effort of the lecturer; the number of registered students should not particularly affect this.

Hypothesis 6: *The size of the beta value for Reg.Studs. is expected to be higher for enthusiasm (Q14) and participation (Q16) compared to the other lecturer questions.*

As discussed under level, the larger the size of the group the more difficult it becomes to constructively involve members of the group in terms of participation. Likewise it may become more difficult to display enthusiasm to the group as a whole the larger it becomes. Personal interaction is normally aided by physical proximity. A significant negative coefficient is expected for Reg.Studs. against both Q14 and Q16. There may be a mild negative relationship between the number of registered students and the lecturer's clarity. It may be a little more difficult to project clarity to large groups where there is less physical proximity between the lecturer and some of the students in the lecture theatre itself.

BS/NonBS

The extent to which students registered in departments other than the Business School vary in their evaluation of the lecturing performance of Business School lecturers compared to Business School students, will depend on the quality of lecturing that the students are used to from lecturers in their own department. This applies to the issues of preparation, enthusiasm, explanation, encouraging participation and overall effectiveness. A definitive expectation can be stated however for Q18 preparation.

***Hypothesis* 7:** *NonBS is expected to be a significant predictor of a lecturer's availability (Q18). Business School students will tend to rate Business School lecturers higher than non-Business School students.*

This is a reflection of the reality that those lecturing on modules taught to non-Business School students are far more geographically separated from the students than lecturers in the students own department. This is likely to put them at an inherent disadvantage on this question.

Rat.L.Hrs.

Hypothesis 8: *Rat.L.Hrs. is not expected to be a significant predictor of preparation (Q13) and clarity (Q15).*

The comments of the academics interviewed (reported in chapter 8) indicated that there was little evidence that academics vary their preparation in relation to the proportion of the module on which they lecture. A similar argument may well hold for Q15 clarity as the ability to express oneself is a skill that should not be expected to vary in relation to the proportion of lecturing hours taught by a particular lecturer.

Hypothesis 9: *Rat.L.Hrs. is expected be a more significant predictor of participation (Q16) and availability (Q18) than the other four lecturer-specific questions.*

When lecturing on a small proportion of a module there may be a tendency to simply "do one's bit" on a module and not be so keen to engage in discussion with the students on a module in which they are not the main (or joint) lecturer. In terms of availability, the question reads: the lecturer could be contacted for advice if needed. If the question was to be answered in the context that it is phrased then students who did not make an attempt to visit a lecturer for advice on a particular module should offer no response. However in most cases a response is entered. In modules in which there are two or more lecturers there is a possibility here of a contrast effect. The lecturer who plays a smaller part (in terms of contact hours) on the module is likely (simply on a pro rata argument) to be contacted less than the main lecturer on the module. The situation is likely to be starker than this however, as attempts to contact the lecturer who has the smaller proportion of the contact hours is likely to be *less* than pro rata. Such a relationship would reflect the fact that much of the contact outside lecturers is likely to involve clarification of the coursework and examination arrangements which is (by definition) the responsibility of the module leader who will normally (though not always) be the lecturer that lectures the greatest proportion of lecturing hours.

Subject area

As discussed in chapter 5, modules in the Business School undergraduate programmes range from more applied, generally qualitative, disciplines (e.g. human resource management and marketing) to more theoretical, often quantitative, disciplines (e.g. accounting and quantitative methods). The impact on the dependent variable (student feedback ratings) for any of the subject areas is likely to change over the six lecturer-specific questions.

Hypothesis 10: *It is expected that lecturer's enthusiasm (Q14) and the lecturer's encouragement of participation (Q16) may be rated higher on modules that are more applied.*

Hypothesis 11: *Lecturer's clarity (Q15) is not expected to be significantly different between different subject areas.*

This expectation reflects the understanding that clarity is a direct skill of the lecturer. There is no obvious reason to explain the subject material directly influencing a lecturer's clarity.

Lecturer's age

The effect of a lecturer's age is not easy to forecast. One expectation can be offered.

Hypothesis 12: *Lecturer's age is expected to be negatively associated with encouragement of participation (Q16).*

Younger lecturers may be able to generate a greater rapport with the undergraduates than the older lecturers, as their experiences of being an undergraduate will be fresher in their minds. Older lecturers therefore may tend to receive lower ratings on encouragement of participation (Q16) compared to the younger lecturers. Older lecturers are more likely to have experience of lecturers as undergraduates who were more formal compared to the Business School norm, in which students' participation during lectures was not encouraged.

Lecturer's rank

Hypothesis 13: *A lecturer's rank is expected to be a significant predictor of lecturer's preparation (Q13).*

Professors may be inclined to use more traditional teaching methods (for example giving references to books, rather than handouts). This could be reflected in lower ratings for their 'preparation'.

***Hypothesis 14:** Senior lecturers are expected to receive lower ratings on lecturers' availability (Q18) compared to the standard lecturers.*

Students may find it more difficult to arrange a time to see senior lecturing staff compared to the standard lecturers.

Lecturer's experience

***Hypothesis 15:** A lecturer's experience is expected to be positively correlated with a lecturer's ratings.*

Lecturers should be able to improve their lecturing performance with experience. This is one of the fundamental aims of collecting student feedback data. Experience should allow a lecturer to become more acutely aware of the most suitable way to teach the students. However, as pointed out in section 10.3 the data available in this research is the number of years a lecturer has been employed at Loughborough University. Therefore this variable is not expected to be a particularly strong predictor of a lecturer's ratings.

10.6 Regression models for lecturers' ratings using the variable DIFOTL to measure contrast effects

In this section models are built for each of the six lecturer-specific questions using the predictor variables presented in Figure 10.1 together with the variable DIFOTL to measure contrast effects. Models using other variables to measure contrast effects are discussed in section 10.7.

The variable DIFOTL was selected from the four variables that measure contrast effects discussed in section 10.4, as it was the predictor variable used to measure contrast effects in chapter 9. Equations for each of the six lecturer-specific questions are presented as equations 10.1 to 10.6. In each of the equations there were 19 predictor variables as listed below. The regression analysis was carried out in SPSS for Windows using backward elimination.

Predictor variables (19): LEV2, LEV3, LEV4, BK, ECON, HRM, IS, MAR, SM, RET, QU, OTHER, Reg.Studs., NonBS, RANK, AGE, EXPERIENCE, Rat.L.Hrs., DIFOTL.

Q13 = 3.979 + .08321 LEV4 + .268 BK + .298 ECON + .104 HRM -.0999 IS
t-values: (2.547) (4.779) (4.858) (3.088) (-2.217)

-. 000604 Reg.Studs. - .00545 Experience + .279 Rat.L.Hrs. + . 456 DIFOTL [10.1]
(-2.040) (-3.902) (6.388) (15.463)

$R^2 = .490$ F = 47.741 (p = .000) n = 458

Q14 = 3.850 + .07787 LEV3 + .232 LEV4 + .298 BK + .335 HRM +.120 MAR -.0991 QU
t-values: (2.388) (6.206) (5.307) (9.798) (2.740) (-2.817)

-. 000825 Reg.Studs. + .386 Rat.L.Hrs. + . 471 DIFOTL [10.2]
(-2.549) (8.876) (18.690)

$R^2 = .635$ F = 86.653 (p = .000) n = 458

Q15 = 3.688 -.09241 LEV2 + .226 LEV4 + .415 BK + .301 HRM - .195 SM - .181 QU
t-values: (-2.255) (4.898) (5.415) (6.508) (-1.985) (-3.856)

-. 001401 Reg.Studs. + .250 Rat.L.Hrs. + .520 DIFOTL [10.3]
(-3.491) (4.279) (21.790)

$R^2 = .617$ F = 80.035 (p = .000) n = 458

$$\begin{array}{lllllll}
Q16 = 3.861 \; -.117\,\text{LEV2} & +.136\,\text{LEV4} & +.443\,\text{BK} & +.502\,\text{HRM} & +.120\,\text{IS} & +.116\,\text{MAR} \\
\text{t-values:} \quad (-3.296) & (3.184) & (6.440) & (12.210) & (2.255) & (2.181)
\end{array}$$

$$\begin{array}{lllll}
+.113\,\text{OTHER} & -.003509\,\text{Reg.Studs.} & -.106\,\text{RANK} & +.307\,\text{Rat.L.Hrs.} & +.484\,\text{DIFOTL} \\
(2.180) & (-9.936) & (-3.208) & (6.132) & (17.051)
\end{array} \qquad [10.4]$$

$R^2 = .635 \quad F = 70.589 \quad (p = .000) \quad n = 458$

$$\begin{array}{lllllll}
Q17 = 3.767 \; -.103\,\text{LEV2} & +.160\,\text{LEV4} & +.387\,\text{BK} & +.277\,\text{HRM} & -.224\,\text{SM} & -.159\,\text{QU} \\
\text{t-values:} \quad (-2.635) & (3.620) & (5.272) & (6.268) & (-2.389) & (-3.541)
\end{array}$$

$$\begin{array}{lll}
-.00177\,\text{Reg.Studs.} & +.321\,\text{Rat.L.Hrs.} & +.504\,\text{DIFOTL} \\
(-4.611) & (5.737) & (18.989)
\end{array} \qquad [10.5]$$

$R^2 = .581 \quad F = 68.945 \quad (p = .000) \quad n = 458$

$$\begin{array}{llllllll}
Q18 = 3.445 \; +.0837\,\text{LEV2} & +.106\,\text{LEV3} & +.185\,\text{LEV4} & +.201\,\text{BK} & +.100\,\text{IS} & -.144\,\text{MAR} & -.261\,\text{SM} \\
\text{t-values:} \quad (2.114) & (2.405) & (3.691) & (3.195) & (2.066) & (-3.069) & (-3.345)
\end{array}$$

$$\begin{array}{lllllll}
+.282\,\text{RET} & +.156\,\text{QU} & -.001\,\text{Reg.Studs.} & -.083\,\text{NonBS} & -.088\,\text{RANK} & +.413\,\text{Rat.L.Hrs.} & +.425\,\text{DIFOTL} \\
(4.300) & (4.035) & (-2.687) & (-2.580) & (-2.901) & (8.602) & (11.801)
\end{array}$$

$R^2 = .475 \quad F = 28.675 \quad (p = .000) \quad n = 458$ [10.6]

Equations 10.1 to 10.6 present the regression equations for each of the six lecturer-specific questions using the variable DIFOTL to measure contrast effects. Each of the equations has a large R^2 value indicating that the equations have considerable explanatory power, with the predictor variables accounting for a large proportion of the variance in the dependent variable (i.e. student feedback ratings). The variable DIFOTL is significant for each of the six lecturer-specific questions. DIFOTL has the strongest effect of the predictor variables in each of six lecturer-specific questions. As suggested in section 10.5 the importance of the various predictor variables is expected to substantially differ over the six lecturer-specific questions. This reflects the fact that each of the questions represents a different aspect of the lecturing process and a lecturer's responsibilities. Equations 10.1 to 10.6 show this to be the case. Four predictor variables: DIFOTL, Reg.Studs., Rat.L.Hrs. and taught course postgraduates (rating lecturers differently to undergraduates) are significant in each of the six lecturer-specific questions, but there is considerable variation in which of the other predictor variables are significant over the six lecturer-specific questions.

The impact on student feedback ratings for each of the predictor variables will be discussed in turn. They will be discussed in terms of the hypotheses outlined in section 10.5.

Module level

Module level has a significant effect on student feedback ratings for each of the six lecturer-specific questions. Lecturers tend to receive higher ratings from taught course postgraduates compared to undergraduates for all six of the lecturer-specific questions.

There is no significant difference in the ratings of lecturers' preparation (Q13) between any of the undergraduate levels.

Hypothesis 1: *The module level is unlikely to have a particularly substantial impact on ratings for a lecturer's preparation (Q13) or a lecturer's clarity (Q15).*

Therefore hypothesis 1 is partly supported.

Postgraduate taught course students tend to rate lecturers higher than first year undergraduates in terms of lecturers' encouragement of participation and questions (Q16). Third year undergraduates were also expected to tend to rate lecturers higher than first year undergraduates for this aspect of lecturing. It was thought that having returned from spending a year in industry in which the interaction between supervisors/managers and the student is likely to be rather different to that which the students will experience with individual lecturers in the Business School during their final year of undergraduate study the third year undergraduates would evaluate their learning experience differently. This experience does not seem to greatly influence the ratings that finalists give lecturers for the encouragement of participation. In the two questions in which LEV3 is significant - lecturer's enthusiasm (Q14) and lecturer's availability (Q18) - finalists tend to give lecturers higher ratings compared to first year students. In both cases this difference is significant at the 5% level, but not at the 1% level.

***Hypothesis 2:** The module level is likely to have a substantial impact on ratings for a lecturer's enthusiasm (Q14) and a lecturer's encouragement of participation (Q16).*

***Hypothesis 3:** LEV3 (i.e. final year undergraduates) and LEV4 (i.e. taught course postgraduates) are expected to be stronger predictors of a lecturer's enthusiasm (Q14) and a lecturer's participation (Q16) compared to the four other lecturer-specific questions.*

Therefore hypothesis 2 and hypothesis 3 are partially supported.

Each of the three module level indicator variables are significant predictors of a lecturer's availability (Q18). The ratings received by lecturers for this question sequentially increase through the module levels, being at the lowest for first year undergraduates. This may be explained by the fact that students will over time have built up rapport with an increasing number of academics and therefore be more likely to approach them when they need advice or assistance.

***Hypothesis 4:** There is likely to be a particularly strong effect in the predictive effect of LEV4 on a lecturer's availability (Q18).*

Therefore hypothesis 4 is supported.

Lecturers tend to receive higher ratings for their overall effectiveness (Q17) from students at LEV3 and LEV4, than from students at earlier stages of their degree programme. This mirrors the established pattern in the student feedback literature (Aleamoni, 1981; Braskamp and Ory, 1994; Feldman, 1978). Unexpectedly, second year undergraduates are rating their lecturers' overall effectiveness less favourably compared to first year undergraduates. The difference, although significant at the 1% level is not very large (t = -2.635).

Reg.Studs.

The number of registered students on a module has a significant effect on student feedback ratings for each of the six lecturer-specific questions. For each of the six lecturer-specific questions the effect is negative - so that on average the greater the number of students registered on the module, the lower the ratings received by the lecturer.

Of the two hypotheses relating to Reg.Studs. hypothesis 5 is supported and hypothesis 6 is partly supported.

Hypothesis 5: *Reg.Studs. is expected to be a weaker predictor of student feedback ratings for preparation (Q13) compared to the other five lecturer-specific questions.*

The lecturer-specific question that Reg.Studs. influences the least is lecturer's preparation (Q13). Although significant at the 5% level, the effect of Reg.Studs. was not significant at the 1% level. This supports the expectation that as preparation is fundamentally a skill relating to the prior effort of the lecturer; ratings for preparation should not be particularly influenced by factors reflecting dynamics within the lecture theatre (e.g. the number of students).

Hypothesis 6: *The size of the beta value for Reg.Studs. is expected to be higher for enthusiasm (Q14) and participation (Q16) compared to the other lecturer questions.*

The number of registered students has the greatest impact on the ratings for Q16 - lecturer's encouragement of participation. It becomes increasingly difficult to suitably handle, to the overall benefit of all students present, questions and interjections from individual students as the class size increases. However the number of registered students did not have a particularly large impact on a lecturer's ratings for Q14 - lecturer's enthusiasm. Reg.Studs. is not a significant predictor of lecturer's enthusiasm (Q14) at the 1% level, it is though significant at the 5% level. This evidence suggests that lecturer's enthusiasm is less affected by the number of students than might have been expected.

As reported in section 3.6, it has been suggested that the relationship between student feedback ratings and Reg.Studs. is non-linear (see Centra and Creech, 1976). To examine this possibility the residuals were plotted against the predictor variable - Reg. Studs. This scatterplot is presented as Figure 10.2.

Figure 10.2 Residual plot of Reg.Studs. for lecturer's overall effectiveness (Q17)

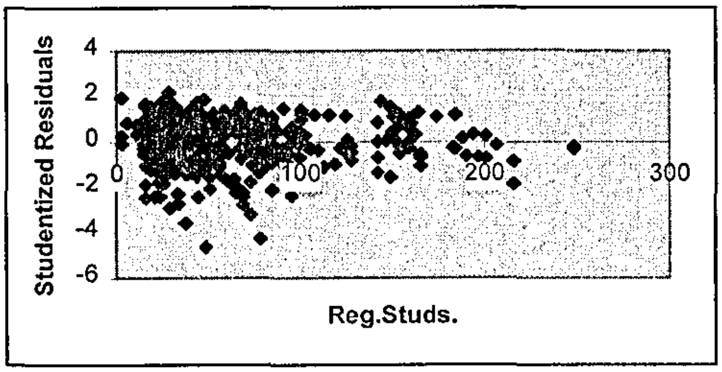


Figure 10.2 offers no evidence that the relationship between the predictor variable Reg.Studs. and the dependent variable (student feedback ratings) is non-linear. The scatterplot is relatively symmetrical around the x-axis suggesting random variation in the residuals.

BS/NonBS

There is no significant difference between the ratings given by students in the Business School and those registered in other departments for Q13-Q17.

The hypothesis relating to BS/NonBS is supported.

***Hypothesis* 7:** *NonBS is expected to be a significant predictor of a lecturer's availability (Q18). Business School students will tend to rate Business School lecturers higher than non-Business School students.*

There was a significant difference for Q18 - lecturer's availability. For this question Business School students tended to rate lecturers higher than students registered in other departments. This is likely to reflect a geographical and possibly a psychological separation between non-Business School students and Business School lecturers. With this one exception this analysis suggests that Business School managers interpreting the results of student feedback data should not consider ratings of lecturers received from students of other departments any differently from ratings received from Business School students. No explicit re-weighting of student feedback data on the basis of a students department need be considered.

Rat.L.Hrs.

The proportion of lecturing hours undertaken by a lecturer on a particular module is a significant predictor variable for student feedback ratings for each of the six lecturer-specific variables.

The two hypotheses relating to Rat.L.Hrs. are only partly supported.

***Hypothesis* 8:** *Rat.L.Hrs. is not expected to be a significant predictor of preparation (Q13) and clarity (Q15).*

Rat.L.Hrs. has the least impact at predicting a lecturer's clarity (Q15).

***Hypothesis* 9:** *Rat.L.Hrs. is expected be a more significant predictor of participation (Q16) and availability (Q18) than the other four lecturer-specific questions.*

The two lecturer-specific questions that Rat.L.Hrs. influences the most are a lecturer's availability (Q18) and a lecturer's enthusiasm (Q14). In modules in which a lecturer only takes a small proportion of the lecturing hours there is a smaller chance that a student will wish to visit the lecturer outside the lecture itself (see section 8.3 for discussion of this).

Subject area

The subject area of a module is a significant predictor of student feedback ratings. Those lecturing on banking (BK), human resource management (HRM) and marketing (MAR) modules tend to receive higher ratings than those lecturing on accounting (AFM) modules, that acted as the comparison group for the subject area indicator variables. In contrast those lecturing on strategic management (SM) and quantitative methods (QU) modules tend to receive lower ratings than those lecturing on AFM modules. Student feedback ratings of those lecturing on economics (ECON) and retailing (RET) modules tended to not be significantly different to those lecturing on AFM modules. It is interesting to note that the lecturer's availability (Q18) does not tend to follow these patterns. There seems to be little similarity between the effect on student feedback ratings of subject area for Q18 and the other five lecturer-specific questions. For example, lecturers on QU modules tend to receive higher ratings for Q18 than lecturers on AFM modules; this is in contrast to the most of the other lecturer-specific questions.

Hypothesis 10: *It is expected that lecturer's enthusiasm (Q14) and the lecturer's encouragement of participation (Q16) may be rated higher on modules that are more applied.*

Hypothesis 10 is supported by equations 10.1 to 10.6. Marketing modules are significantly positively associated with Q14 and Q16, but not with Q13, Q15 or Q17. Human resource management modules are significant predictors of lecturers' ratings in five of the six lecturer-specific questions (the exception being a lecturer's availability - Q18). The largest beta values for HRM are found in Q14 and Q16.

Hypothesis 11: *Lecturer's clarity (Q15) is not expected to be significantly different between different subject areas.*

Hypothesis 11 is not supported. Four of the subject areas are shown to be significant predictors of lecturers' ratings.

Lecturer's age

Age is not a significant predictor of student feedback data for any of the six lecturer-specific questions. This suggests that the age of the lecturer does not systematically affect their student feedback ratings.

Hypothesis 12: *Lecturer's age is expected to be negatively associated with encouragement of participation (Q16).*

Therefore hypothesis 12 is not supported.

As with Reg.Studs. there is evidence in the student feedback literature (see Braskamp, et al., 1985) suggesting that there may be a non-linear relationship between the lecturer's age and the student feedback ratings they receive (see section 3.6). To examine this possibility, the residuals from the model for Q17 were plotted against the predictor variable - lecturer's age. This scatterplot is presented as Figure 10.3.

Figure 10.3 Residual plot of lecturer's age for lecturer's overall effectiveness (Q17)

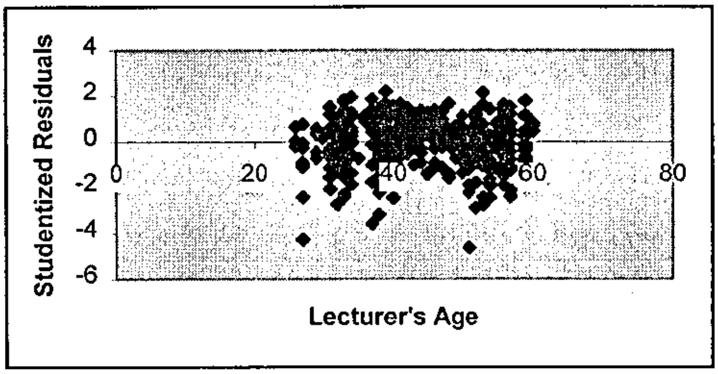


Figure 10.3 offers no evidence of a non-linear relationship between the predictor variable, lecturer's age and the dependent variable (student feedback ratings). The scatterplot is relatively symmetrical around the x-axis suggesting that the relationship between student feedback ratings and lecturer's age is linear.

Lecturer's rank

Rank is a significant predictor variable of student feedback ratings for two of the six lecturer-specific questions, Q16 (encouraged participation) and Q18 (availability). In both cases the relationship is negative, illustrating that senior lecturers (including professors) receive lower ratings than standard members of lecturing staff.

Hypothesis 13: *A lecturer's rank is expected to be a significant predictor of lecturer's preparation (Q13).*

Hypothesis 14: *Senior lecturers are expected to receive lower ratings on lecturer's availability (Q18) compared to the standard lecturers.*

Therefore hypothesis 13 is not supported but hypothesis 14 is supported. Students often find it more problematic to visit senior lecturers (particularly professors) compared to standard lecturers. Often students will have to arrange a suitable time to speak to a senior lecturer with that lecturer's secretary, whereas students can often simply 'knock on the door' of a standard lecturer. Some students may consider the most senior lecturers less approachable because of their status. This does not necessarily imply that the senior lecturer is 'less contactable' but rather 'less accessible'.

Lecturer's experience

Experience is a significant predictor variable of a lecturer's ratings for only one of the six lecturer-specific questions, the lecturer's preparation (Q13). Lecturers who have been teaching at Loughborough University for a greater period of time tend to receive *lower* ratings than lecturers who have more recently joined the department, as indicated by the negative coefficient.

***Hypothesis 15:** A lecturer's experience is expected to be positively correlated with a lecturer's ratings.*

Therefore hypothesis 15 is not supported. Hypothesis 15 was based on the expectation that there would be a familiarity effect in student feedback ratings. Lecturers who have more experience in the dominant student expectations at Loughborough University and the Business School procedures regarding lecturing and module design were expected to receive higher ratings than lecturers who had less experience in these matters. For equations 10.1 to 10.6 this did not prove to be the case. However, of the equations presented in section 10.7 (equations 10.7 to 10.24) in which experience was significant, the coefficient *was* positive, i.e. those lecturers who had been employed at the Business School longer did tend to receive higher ratings than lecturers who had more recently joined the department.

The negative relationship between a lecturer's experience at Loughborough University and the ratings they receive for preparation (Q13), may suggest that lecturers who have been teaching the same module over successive years are less likely to update the material to include current examples and the latest research than lecturers new to the Business School.

DIFOTL

DIFOTL is a highly significant predictor variable of student feedback ratings for each of the six lecturer-specific questions. In each case it is positive and the strongest of the predictor variables. This clearly reflects the importance of contrast effects in student feedback data, based on whom a lecturer teaches alongside on a particular module, as discussed in chapter 9.

10.7 Comparison of the effects of each of the four predictor variables measuring contrast effects

10.7.1 The two main distinctions between the predictor variables measuring contrast effects

Having established the impact of the various predictor variables in the models of student feedback data using DIFOTL, the effect of using DIFOTL as a measure of contrast effects can be contrasted to the effect of using each of the other predictor variables measuring contrast effects presented in section 10.4. As discussed in section 10.4, there are principally two main dimensions that differentiate between the four variables. These are depicted as whether:

- the variable consists of module-specific information (i.e. DIFOTL and OTLES) or a general average of lecturers' student feedback data over all the modules on which they lecture (i.e. DIFOLAV and OLSFA);

- the variable uses data from both lecturers (i.e. DIFOTL and DIFOLAV), or whether the variable only uses data from other lecturer(s) teaching on the module (i.e. OTLES and OLSFA).

10.7.2 Regression models using DIFOLAV as the predictor variable measuring contrast effects

In this section the models including DIFOLAV for each of the lecturer-specific questions are presented. The equations 10.7 to 10.12 correspond to the equations 10.1 to 10.6 in section 10.6. As before in each of the equations there were 19 predictor variables. The same 18 as discussed in section 10.6 and DIFOLAV which was substituted for DIFOTL.

Q13 = 4.126 +.132 LEV4 +.301 BK +.240 ECON +.09167 HRM -.006885 AGE
t-values: (3.956) (4.602) (3.426) (2.353) (-4.140)

+ .353 Rat.L.Hrs. + . 336 DIFOLAV [10.7]
(6.927) (6.590)

$R^2 = .291$ F = 26.393 (p = .000) n = 458

Q14 = 3.916 + .08068 LEV3 + .211 LEV4 + .320 BK + .364 HRM +.133 MAR + .124 OTHER
t-values: (2.175) (4.546) (4.971) (9.248) (2.615) (2.487)

-. 0009371 Reg.Studs. - .003736 AGE + .498 Rat.L.Hrs. + . 456 DIFOLAV
(-2.470) (-2.324) (10.141) (12.275) [10.8]

$R^2 = .521$ F = 48.607 (p = .000) n = 458

Q15 = 3.893 + .254 LEV4 + .446 BK + .315 HRM - .148 QU -. 00138 Reg.Studs. -.006583 AGE
t-values: (4.645) (4.713) (5.518) (-2.531) (-2.769) (-2.742)

+ .305 Rat.L.Hrs. + .525 DIFOLAV [10.9]
(4.150) (11.807)

$R^2 = .407$ F = 38.52 (p = .000) n = 458

Q16 = 3.882 + .134 LEV3 + .288 LEV4 + .390 BK + .445 HRM -. 0027 Reg.Studs. - .0804 RANK
t-values: (3.081) (5.688) (4.975) (9.913) (-6.241) (-1.970)

- .00413 AGE + .374 Rat.L.Hrs. + . 455 DIFOLAV
(-1.973) (6.581) (10.408) [10.10]

$R^2 = .517$ F = 53.352 (p = .000) n = 458

Q17 = 3.904 + .155 LEV4 + .453 BK + .335 HRM + .145 OTHER -. 00188 Reg.Studs.
t-values: (2.849) (5.166) (6.336) (2.157) (-4.013)

- .006679 AGE + .438 Rat.L.Hrs. + . 463 DIFOLAV [10.11]
(-2.992) (6.644) (10.010)

$R^2 = .388$ F = 35.556 (p = .000) n = 458

Q18 = 3.407 + .0955 LEV2 + .133 LEV3 +.197 LEV4 + .182 BK - .174 MAR - .296 SM + .274 RET
t-values: (2.165) (2.729) (3.523) (2.601) (-3.353) (-3.396) (3.771)

+ .158 QU -. 000953 Reg.Studs. - .08907 NonBS - .09466 RANK + .478 Rat.L.Hrs. + . 274 DIFOLAV
(3.699) (-2.288) (-2.489) (-2.788) (9.049) (5.192) [10.12]

$R^2 = .343$ F = 17.858 (p = .000) n = 458

Equations 10.7 to 10.12 are generally similar to the corresponding equations 10.1 to 10.6. The predictor variable measuring contrast effects (in this case DIFOLAV) is significant in each of the six lecturer-specific questions. Rat.L.Hrs. is also significant in each of the six lecturer-specific questions as is Reg.Studs., with the exception of lecturer's preparation (Q13). The R^2 values for equations 10.7 to 10.12 are noticeably lower than the corresponding R^2 values for equations 10.1 to 10.6.

The principal difference between the two sets of equations is that in equations 10.7 to 10.12 the lecturer's age is a significant predictor variable of student feedback ratings for each of the lecturer-specific questions, except Q18 (lecturer's availability). This is in clear contrast to the equations using DIFOTL, in which age is not a significant predictor of student feedback data for any of the six lecturer-specific questions. In each case the lecturer's age is negatively associated with ratings, implying that older lecturers receive lower ratings compared to younger lecturers.

What is particularly interesting is the reason why the change of predictor variable from DIFOTL to DIFOLAV has lead to a change in the significance of lecturer's age on student feedback data. Other than changing the variable measuring contrast effects, no other change was made either to the data set or the predictor variables entered into the regression equation.

The main reason for the lecturer's age being significant in some models in which DIFOLAV was used as the predictor variable measuring contrast effects is that variations in lecturer characteristics are built into DIFOTL and account for the very high impact of DIFOTL (i.e. high t-value) in equations 10.1 to 10.6.

As with DIFOTL, DIFOLAV is a significant predictor variable in each of the six lecturer-specific questions. The effect of DIFOLAV is not as strong as DIFOTL. This is reflected in the t-values for DIFOLAV being lower than the corresponding values for DIFOTL in each of the six lecturer-specific questions and the R^2 being lower for each of equations 10.7 to 10.12 compared to the corresponding equations 10.1 to 10.6. It should be noted that the t-values for DIFOLAV are however still strong ranging from 12.275 (Q14) to 5.192 (Q18) and the lowest R^2 in equations 10.7 to 10.12 still accounts for almost a third of the variation in the lecturer's ratings (.291).

At first sight the effect of DIFOLAV in equations 10.7 to 10.12 appears to contradict the results reported in chapter 9. In chapter 9 it was established that while DIFOTL was a significant predictor of lecturers' ratings, DIFOLAV was not a significant predictor of lecturers' ratings for any of the six lecturer-specific questions. The reason why DIFOLAV was not significant in chapter 9, is that chapter 9 primarily incorporates module-specific characteristics into the model. The variable DIFOLAV is confounded with many of the module-specific characteristics by virtue of its construction (i.e. aggregating a lecturer's student feedback ratings obtained from a variety of modules with differing module characteristics, such as class size and module level). In chapter 10 the model is extended to include lecturer-specific characteristics which allows DIFOLAV to become a significant predictor of student feedback ratings.

10.7.3 Regression models using OTLES as the predictor variable measuring contrast effects

Equations 10.13 to 10.18 use the variable OTLES to measure contrast effects. As before a separate equation was run for each of the six lecturer-specific questions. Other than the change in the predictor variable measuring contrast effects the other 18 predictor variables were the same as in equations 10.1 to 10.12.

Q13 = 4.171 +.133 LEV4 +.338 BK +.227 ECON +.08023 HRM -.008179 AGE
t-values: (3.806) (4.966) (3.096) (1.971) (-4.734)

\+ .384 Rat.L.Hrs. [10.13]
(7.238)

$R^2 = .223$ F = 21.528 (p = .000) n = 458

OTLES was not significant in this model. It was the 10th predictor variable to be eliminated. OTLES significance at point of elimination was .221.

Q14 = 3.998 +.142 LEV3 +.310 LEV4 +.334 BK +.285 HRM - .09383 QU - .007795 AGE
t-values: (3.605) (7.702) (4.522) (6.479) (-2.009) (-3.203)

\+ .004975 Experience + .510 Rat.L.Hrs.
(2.119) (8.778) [10.14]

$R^2 = .357$ F = 31.192 (p = .000) n = 458

OTLES was not significant in this model. It was the 9th predictor variable to be eliminated. OTLES significance at point of elimination was .058. OTLES therefore narrowly misses being included in the final model, given that the required significance level for inclusion in the model was set at 5%.

Q15 = 4.319 + .231 LEV4 + .655 BK + .374 HRM +.216 OTHER -. 0017 Reg.Studs.
t-values: (3.483) (5.961) (5.825) (2.596) (-2.993)

- .164 RANK -.007426 AGE - .07141 OTLES [10.15]
(-2.838) (-2.499) (-5.262)

R^2 = .240 F = 17.679 (p = .000) n = 458

Q16 = 4.496 + .197 LEV3 + .344 LEV4 + .481 BK + .442 HRM -. 0026 Reg.Studs.
t-values: (4.022) (6.075) (5.629) (9.001) (-5.365)

- .164 RANK - .0104 AGE + .00631 Experience - .06802 OTLES
(-3.646) (-3.65) (2.356) (-6.55) [10.16]

R^2 = .418 F = 35.774 (p = .000) n = 458

Q17 = 3.912 + .162 LEV4 + .593 BK + .331 HRM + .184 OTHER -. 00195 Reg.Studs.
t-values: (2.715) (5.976) (5.707) (2.472) (-3.813)

- .162 RANK - .00594 AGE + .459 Rat.L.Hrs. [10.17]
(-3.139) (-2.235) (6.329)

R^2 = .267 F = 20.475 (p = .000) n = 458

OTLES was not significant in this model. It was the 9th predictor variable to be eliminated. OTLES significance at point of elimination was .085. As with Q14 OTLES relatively narrowly misses being included in the final model, OTLES would be significant at the 10% level.

$$\begin{array}{llllllll}
\text{Q18} = 3.368 & + .0951\ \text{LEV2} & + .132\ \text{LEV3} & +.214\ \text{LEV4} & + .219\ \text{BK} & + .118\ \text{IS} & - .166\ \text{MAR} & - .269\ \text{SM} \\
\text{t-values:} & (2.099) & (2.629) & (3.728) & (3.038) & (2.124) & (-3.101) & (-3.006)
\end{array}$$

$$\begin{array}{lllllll}
+ .295\ \text{RET} & + .182\ \text{QU} & -.000919\ \text{Reg.Studs.} & - .09185\ \text{NonBS} & - .111\ \text{RANK} & + .519\ \text{Rat.L.Hrs.} & \\
(3.930) & (4.102) & (-2.153) & (-2.504) & (-3.206) & (9.592) & [10.18]
\end{array}$$

$R^2 = .310 \quad F = 15.379 \quad (p = .000) \quad n = 458$

OTLES was not significant in this model. It was the 5th predictor variable to be eliminated. OTLES significance at point of elimination was .184.

There are three interesting differences between the regression equations that use DIFOLAV as the predictor variable measuring contrast effects (i.e. equations 10.7 to 10.12) and the corresponding equations that use OTLES (i.e. equations 10.13 to 10.18).

Firstly, the variable OTLES is only significant for two of the six lecturer-specific questions (Q15 and Q16).

Secondly, there is a greater significance of rank and experience in the models using OTLES as the predictor variable measuring contrast effects compared to those using DIFOLAV. The lecturer's rank is a significant predictor of student feedback ratings for four of the lecturer-specific questions (Q15 - Q18), in contrast to just Q16 and Q18 for the models using DIFOLAV. The lecturer's experience also has a greater impact, being a significant predictor of student feedback ratings for Q14 and Q16. Experience was significant in neither of these in the models using DIFOLAV.

Thirdly, Rat.L.Hrs. is not as important as a predictor of student feedback ratings in the equations using OTLES compared to those using DIFOLAV. Whereas in the equations using DIFOLAV Rat.L.Hrs. is a significant predictor of ratings in each of the six-lecturer specific questions, in the equations using OTLES it is only a significant predictor of ratings for Q13, Q14 and Q18.

These three differences reflect the fact that a greater degree of lecturer-specific information is being reflected in the variable OTLES compared to the variable DIFOLAV. This is expected as only one lecturer's ratings are used in the construction of OTLES, whereas both DIFOTL and DIFOLAV build in a comparison of the lecturers.

10.7.4 Regression models using OLSFA as the predictor variable measuring contrast effects

Equations 10.19 to 10.24 use the variable OLSFA to measure contrast effects. As before a separate equation was run for each of the six lecturer-specific questions. Other than the change in the predictor variable measuring contrast effects the other 18 predictor variables were the same as in equations 10.1 to 10.18.

Q13 = 4.171 + .133 LEV4 + .338 BK + .227 ECON + .08023 HRM -.008179 AGE
t-values: (3.806) (4.966) (3.096) (1.971) (-4.734)

\+ .384 Rat.L.Hrs. [10.19]
(7.238)

R^2 = .223 F = 21.528 (p = .000) n = 458

OLSFA was not significant in this model. It was the 12th (and final) predictor variable to be eliminated. OLSFA significance at point of elimination was .061. Therefore OLSFA narrowly misses being included in the final model, OLSFA would be significant at the 10% level.

Q14 = 3.961 + .162 LEV3 + .320 LEV4 + .361 BK + .320 HRM + .120 MAR - .008032 AGE
t-values: (4.181) (8.009) (4.880) (7.281) (2.032) (-3.293)

+ .005356 Experience + .513 Rat.L.Hrs.
(2.249) (8.909) [10.20]

$R^2 = .357$ F = 31.210 (p = .000) n = 458

OLSFA was not significant in this model. It was the 3rd predictor variable to be eliminated. OLSFA significance at point of elimination was .793.

Q15 = 3.848 + .257 LEV4 + .635 BK + .361 HRM +.179 OTHER -. 0019 Reg.Studs.
t-values: (3.879) (5.806) (5.647) (2.171) (-3.321)

- .160 RANK -.00593 AGE +.391 Rat.L.Hrs. - .264 OLSFA [10.21]
(-2.792) (-2.011) (4.878) (-3.388)

$R^2 = .255$ F = 17.025 (p = .000) n = 458

Q16 = 4.050 + .186 LEV3 + .327 LEV4 + .469 BK + .433 HRM -. 0025 Reg.Studs. - .173 RANK
t-values: (3.775) (5.744) (5.405) (8.700) (-5.152) (-3.820)

- .00925 AGE + .00581 Experience + . 369 Rat.L.Hrs.
(-3.234) (2.133) (5.778) [10.22]

$R^2 = .407$ F = 34.116 (p = .000) n = 458

OLSFA was not significant in this model. It was the 2nd predictor variable to be eliminated. OLSFA significance at point of elimination was .639.

Q17 = 3.912 + .162 LEV4 + .593 BK + .331 HRM + .184 OTHER -. 00195 Reg.Studs.
t-values: (2.715) (5.976) (5.707) (2.472) (-3.813)

- .162 RANK - .00594 AGE + .459 Rat.L.Hrs. [10.23]
(-3.139) (-2.235) (6.329)

$R^2 = .267$ F = 20.475 (p = .000) n = 458

OLSFA was not significant in this model. It was the 9th (and penultimate) predictor variable to be eliminated. OLSFA significance at point of elimination was .142.

Q18 = 3.278 + .114 LEV2 + .162 LEV3 +.257 LEV4 + .232 BK + .126 IS - .162 MAR - .265 SM
t-values: (2.695) (3.745) (5.873) (3.279) (2.298) (-3.080) (-2.998)

+ .307 RET + .178 QU - .07633 NonBS - .108 RANK + .522 Rat.L.Hrs. + . 284 OLSFA
(4.233) (4.114) (-2.147) (-3.177) (9.812) (4.219) [10.24]

$R^2 = .330$ F = 16.833 (p = .000) n = 458

Equations 10.19 to 10.24 that use OLSFA as the predictor variable measuring contrast effects are very similar to equations 10.13 to 10.18 that use OTLES. There are two reasons that help explain the similarity in the results between the equations that use OLSFA as the predictor variable measuring contrast effects and those using OTLES.

Firstly, in both of the variables OLSFA and OTLES only the other lecturer(s) rating (or ratings in the case of OLSFA) are used to construct OLSFA and OTLES. Neither of these two predictor variables directly compares the two lecturer's ratings. Secondly, OLSFA and OTLES are significantly correlated with each other at the 1% level for Q14, Q15, Q16 and Q17. They are significantly correlated at the 5% level for Q18.

10.8 Comparison of the Business School data and the published student feedback literature

Having discussed the effects on student feedback ratings of the predictor variables outlined in Figure 10.1, these effects can now be compared to the dominant findings in the student feedback literature. These comparisons are of particular interest as very little of the student feedback literature originates from the U.K., or uses U.K. data. Most of the published student feedback literature is U.S. in origin, although there are other notable published works emanating from countries such as Australia and Israel.

The effect of class size on student feedback ratings in the Business School data is consistent with the published literature. Of thirty studies on this theme Feldman (1978) found that twenty of them reported a negative relationship between student feedback ratings and class size, i.e. the greater the number of students registered on the module the lower the ratings received by the lecturer. This relationship is also present in the Business School data. There is a consistent negative relationship between student feedback ratings and Reg.Studs. The effect is not significant for all the lecturer-specific questions however. The teaching skills of preparation and enthusiasm are not as influenced by the class size as the other four lecturer-specific questions. The Business School student feedback data does not support the claim that the relationship between student feedback ratings and class size is non-linear, with relatively small and relatively large class sizes gaining higher ratings (see Centra and Creech, 1976; Marsh, Overall and Kesler, 1979).

The Business School student feedback data generally supports the main finding in the published literature that ratings tend to be higher in modules taken later in the degree programme and higher still in taught course Masters modules (Aleamoni, 1981; Braskamp and Ory, 1994; Feldman, 1978). Students on taught course Masters modules in the Business School tend to rate lecturers higher than the undergraduates. However, there is not a clear consistent increase in ratings throughout the progression of a undergraduate programme in the Business School. For in only two of the lecturer-specific questions (Q14 - lecturer's enthusiasm and Q18 - lecturer's availability) is there evidence of final year undergraduates rating lecturers more highly than first year undergraduates. For two of the lecturer-specific questions (Q15

- lecturer's clarity and Q16 - lecturer's encouragement of participation) second year undergraduates are rating lecturers *lower* than first year undergraduates.

Lecturer's rank is shown not to be a strong predictor of student feedback ratings in the student feedback literature. Of 33 studies analysed by Feldman (1983) the majority (21) reported no significant correlation between a lecturer's rank and their ratings. Of the eleven that did report a significant association all but one reported a positive association, i.e. the higher the rank of the lecturer the higher their ratings. Marsh (1987) reports that questions relating to lecturer knowledge and intellectual expressiveness tend to be positively related to academic rank whereas questions referring to encouragement of discussion, openness and concern for students tend to be negatively related to a lecturer's academic rank.

The Business School student feedback data does not generally support these published findings. In each of the questions in which there is a significant association between student feedback ratings and lecturer's rank the association is negative, i.e. those ranked as senior lecturer or professor tend to receive lower ratings than standard lecturers. In two of the lecturer-specific questions (Q13 - lecturer's preparation and Q14 - lecturer's enthusiasm) there is no significant association.

Lecturer's age is generally shown in the published student feedback literature to be a significant predictor of student feedback ratings. Feldman (1983) analysed twelve studies and found that half of them reported a *negative* relationship with ratings, i.e. the older the lecturer the lower the ratings. In the other six studies there was no significant correlation between the age of the lecturer and the ratings they received.

The Business School student feedback data does support the research reporting an inverse relationship between a lecturer's age and a lecturer's ratings. In all but one of the lecturer-specific questions (the exception being Q18 - lecturer's availability) there is a significant negative association between student feedback ratings and the lecturer's age.

There is a mixed pattern of results in the published literature relating to the relationship between student feedback ratings and lecturer's experience. Feldman

(1983) analysed sixteen studies that showed a mixed pattern of results. Eight studies showed no significant correlation between the experience of the lecturer and the ratings they received. Of the seven studies where a statistically significant correlation was found, the majority (5) found significant evidence of a negative relationship between the experience of the lecturer and their student feedback ratings, i.e. the greater the experience of the lecturer the lower their ratings. The remaining study, Centra and Creech (1976) revealed a non-linear pattern.

The Business School student feedback data reveals a significant negative association between a lecturer's preparation (Q13) and lecturer's experience, i.e. lecturers who have been employed at Loughborough University longer tend to receive lower ratings for preparation than lecturers who have recently joined the Business School. In contrast the data reveals a significant positive association between a lecturer's enthusiasm (Q14) and a lecturer's encouragement of participation and questions (Q16) with a lecturer's experience. These positive associations are not that strong, although statistically significant at the 5% level they are not significant at the 1% level.

The most important element in the comparison of the Business School data to the published student feedback literature is the demonstration of what is absent in the existing published student feedback literature. This research contributes to the student feedback literature through the introduction of two additional variables that this research indicates are significantly associated with student feedback ratings that are not covered in the existing literature. Firstly, the ratio of lecturing hours that a lecturer teaches on a particular module is shown to be significantly positively associated with the ratings lecturers receive, i.e. the greater the proportion of lecturing hours a lecturer teaches on a particular module the higher their ratings are likely to be. Secondly, the impact of the performance of the lecturer(s) with whom one teaches on a particular module has been shown to influence the ratings that a lecturer receives.

10.9 Conclusions

The aim of this chapter was to model student feedback data using various predictor variables that have been shown, either in the published literature or in earlier stages of this research, to potentially influence the student feedback ratings that lecturers receive. In studying the effect of these variables on ratings the aim was to be able to highlight variables that particularly impact on the Business School student feedback data. The benefit of generating this information is that it can then be used to inform the judgement of any individual who has the job of interpreting student feedback data. Understanding which variables generally increase or decrease lecturers' ratings, should allow the data to be interpreted more fairly.

Various implications follow from the analyses presented in this chapter. Each of these will be considered in turn.

a) Lecturers' student feedback ratings need to be contextualised

Lecturers teaching taught course postgraduate modules tend to receive higher ratings than lecturers teaching on undergraduate modules. Similarly, lecturers teaching on modules with a smaller number of registered students also tend to receive higher ratings. Therefore, the student feedback ratings on a large first year undergraduate module should not be interpreted on parity with a small taught postgraduate module. Through a greater appreciation of the subtleties of the Business School student feedback data the opportunity is provided for a more sophisticated interpretation of a lecturer's ratings and should encourage a greater degree of confidence in student feedback data by the individual academic and the manager alike.

The significant findings from the previous two chapters were further emphasised in this chapter. In line with chapter 8, lecturers' ratings are shown to be significantly positively related to the proportion of lecturing hours undertaken by a lecturer on a module. Lecturers delivering a higher proportion of lecturing hours on a module tend to receive higher ratings than lecturers who deliver a smaller proportion of lecturing hours. The main lecturer on a module has greater opportunity to set the expectations for the module, as well as being better placed to give students advice about coursework and examinations.

In line with chapter 9, a lecturer's ratings are shown to be relative to those they teach with in modules taught by more than one lecturer. A lecturer will generally receive lower ratings than their norm when they teach alongside a lecturer who is generally ranked higher than they are. This relationship needs to be clearly recognised by individual lecturers and managers when they interpret student feedback ratings.

b) The implications of the significance of class size for departmental managers and the university's admission policy

The research clearly shows that class size significantly influences lecturers' student feedback ratings. Lecturers teaching larger classes tend to receive lower ratings than lecturers teaching smaller classes. Furthermore, statistical evidence suggests that the relationship is relatively linear, implying that ratings are increasingly negatively influenced the bigger the class gets. This has important implications for departmental and university senior managers in terms of the optimal level student enrolment. As discussed in chapter 2, the expansion in student numbers and the consequent increase in the levels of government funding required to finance the expansion has led to a desire for greater efficiency in universities. Universities have needed to reduce unit costs through increased efficiency, Dearing explicitly referred to the:

> 'increasing workloads and outputs at a time of declining unit resources have been a feature across the system' (Dearing, 1997, 14.16).

One of the consequences of this is that class sizes have generally increased. The ramification of effect of class size is clearly that universities need to be cautious about continuing to increase the number of students they enrol onto undergraduate degree programmes. For although each additional student increases the university's income, there is a price in terms of the students educational experience which would seem to deteriorate as the class size increases. Therefore senior university managers need to carefully consider the optimal level of student intake. Departmental managers may need to consider upper limits on class size and consider splitting very large modules into two or ultimately the student experience may suffer to an extent where it may become difficult to attract students, particularly the high-fee paying oversees students to read at any given university.

c) The implications for departmental policy of the ratings lecturers receive from postgraduate students

Students on taught postgraduate courses are generally particularly satisfied with the educational provision made available to them. This is an important result for the Business School, given the increase in postgraduates studying in U.K. universities. As discussed in chapter 2, of all the categories of student that have increased in number over recent years, postgraduate numbers have grown the fastest (Dearing Report, 1997). In 1995/96 postgraduate students represented 14% of U.K. higher education students compared to 6% in 1962/63. The enrolment of students onto taught postgraduate courses is an important source of income for universities, particularly since successive governments have gradually reduced the amount of per-capita state funding they allocate to universities for teaching undergraduate students. Business Schools have a particular niche for generating income from taught-course Masters' programmes in the form of MBA's. The fee structures for MBA programmes are market driven with fee levels ranging considerable between universities. Developing the reputation of offering a MBA programme that is both efficient and educationally effective is potentially financially lucrative. The presence of successful Master level degrees in an academic department has an additional benefit of raising the academic prestige of the university. The ability of the Business School therefore to deliver postgraduate programmes that are considered by the students to positively enhance their education is encouraging.

d) The implications for the lecturing styles employed by lecturers

The modelling of lecturers' ratings provides evidence that the Business School student feedback data may currently disadvantage those lecturers who employ a 'learning paradigm' form of teaching. A significant piece of evidence to support this is that lecturers' student feedback ratings are significantly influenced by the lecturer's age and the lecturer's rank, i.e. older and more senior lecturers tend to receive lower ratings than younger and less senior lecturers. The explanation of this may lie with Kugel's (1993) conceptualisation of the development of lecturers' teaching technique. As discussed in chapter 2, Kugel argues that as a generality lecturers' teaching techniques tend to change over time from initially placing the emphasis on teaching (i.e. on their own performance) through to placing the emphasis on the students learning, ultimately considering the student as an independent learner and their own role being as a facilitator aiding the student to develop their own independent learning. Therefore there is a tendency for older and more senior lecturers to follow a 'learning paradigm' form of teaching which challenges students and is more appreciated by students following a 'deep' approach to their studies.

A tendency for the Business School student feedback data to disadvantage lecturers following a 'learning paradigm' form of teaching may or may not be considered a problem by departmental managers, depending on whether the department wishes to encourage a particular form of teaching. Some suggestions as to how any particular bias against lecturers following a 'learning paradigm' will be discussed in the next chapter.

e) The implications of the significance of contrast effects for teaching practice

The significance of contrast effects in student feedback data is particularly important in terms of the overall concept of teaching quality assessment. As discussed in chapter 2, Henkel (1997) found for some academics:

> 'the assessment exercises enhanced academics' sense of being part of a collective enterprise with colleagues. It made them more aware of the programmes as a whole and of other people's teaching. This enabled a more coherent departmental approach. Sometimes it generated more sharing between colleagues of their approaches to teaching and what they were trying to achieve' (Henkel, 1997, 19).

The significant influence that the ratings of the lecturer(s) with whom a lecturer shares lecturing responsibilities on a particular module has on the ratings a lecturer receives stresses the need for lecturers to be more aware of their colleagues teaching.

Given that teaching as a team is a frequent feature of teaching in the Business School lecturers should be encouraged more to share their lecturing ideas with each other. Furthermore, modules in which lecturing is shared should be organised in such a way as the student receives a consistent approach to the module material as substantial variations in the teaching style used by lecturers on the same module may lead to the students concentrating their efforts on the part of the syllabus, the part of the syllabus taught by the lecturer whose teaching style is most congruent with the students learning style. If lecturers were encouraged to increasingly share lecturing techniques and pool their experiences, student feedback data would take on a positive role potentially generating lecturing formats and module designs that were more innovative rather than acting to:

> 'displace trust and elevating institutional and system management to a dominant position in higher education' (Trow, 1993).

Such a focus on teaching would achieve the initial intention of the introduction of TQA of raising the emphasis of good teaching as being a pivotal aspect of higher education. Therefore student feedback could be used as a mechanism through which to encourage lecturers to more effectively integrate their polled experiences for the positive outcome of student development and ultimately potentially increase their own sense of professional fulfilment.

f) The implications of the analyses for the validity of the data

The modelling of lecturers' student feedback data has clearly illustrated that students do actively discriminate between the lecturer-specific questions on the student feedback form. Each of the predictor variables are shown to influence each of the lecturer-specific questions (i.e. dimensions of teaching) to a different extent. This shows that the ratings that students give lecturers are not simply a reflection of whether the student generally liked the lecturer, but are rather an indication that the students are actually responding to each of the questions on the student feedback form by giving independent consideration of their ratings for each dimension of teaching. This has an important implication for the debate, discussed in chapter 2, regarding what the overriding purpose of assessing teaching quality is. Trow (1993) argued that the increasing monitoring of higher education exists primarily to discipline rather than inform academics. Yet the evidence of this research shows that student feedback does provide significant information that does inform academics in the light of students actively differentiating between lecturers and between various aspects of each lecturer's teaching.

The analysis presented in this research further shows that students are differentiating between module conditions and lecturers in the ratings (i.e. demonstrating criterion-related validity). This is encouraging as it suggests that the student feedback data have a degree of predictive validity (see Sekaran, 1992 for a discussion of validity) implying that the students have the ability to differentiate between lecturers. The ability to differentiate between stronger and weaker lecturers, allied to the strong levels of fit on many of the regression equations (particularly those using DIFOTL as the predictor variable measuring contrast effects) should increase individual lecturer's and managers' confidence in the usefulness of student feedback ratings.

Chapter 11
CONCLUSIONS AND IMPLICATIONS

11.1 Introduction

This chapter draws together the earlier chapters by considering the contributions that this research makes to the published student feedback literature and the guidance that this research can offer individual lecturers and managers in academic departments regarding the most appropriate interpretations of the results of student feedback. Section 11.2 discusses the aim of the thesis and places the research in the context of the changing nature of higher education in the U.K. in the 1990's. Section 11.3 outlines the implications of the existence of systematic interrater variance for both the appropriateness of using factor analysis on student feedback data and the validity of averaging class responses. Section 11.3 also considers the implications of the presence of different learning styles and different teaching styles found in the Business School data. Section 11.4 considers the implications of the research on the effects on student feedback ratings in modules in which the lecturing responsibilities are shared. Two concepts new to the student feedback literature are discussed: firstly, the effect on ratings of the proportion of lecturing hours undertaken by a lecturer on a particular module; and secondly, the extent to which a lecturer's ratings are influenced by the ratings of those with whom they lecture on a particular module. Section 11.5 considers the relevance that the wording of the questions have for both the style of teaching encouraged in the Business School and the implications for the way in which lecturers are evaluated. Section 11.6 outlines the effects that various predictor variables have on the student feedback ratings lecturers receive and compares these to the results from the published student feedback literature. In doing so, it aims to offer guidance to individual lecturers and departmental managers in the interpretation of student feedback ratings allowing decisions based on the data to be made on a more informed basis and to act as a mechanism to validate the Business School results. Section 11.7 discusses a series of recommendations for managerial practice that arise from the research. Section 11.8 considers the future of student feedback ratings in U.K. higher education. Section 11.9 outlines the limitations of the research and offers some suggestions for future research. Finally, section 11.10 presents a few concluding comments on the research.

11.2 The context and aims of the research

11.2.1 The context of the research

Higher education in the U.K. has changed beyond recognition in terms of both the nature of academic work and of the aims of universities since its origin in the Middle Ages. The early colleges in Oxford and Cambridge were self-governing institutions, independent of government control, free to determine their curricula and their managerial practices (Maassen, 1997, 113). This autonomy continued for several centuries until 1850 when a Royal Commission was established to review the operation of the Oxford and Cambridge colleges. This was the first significant challenge to the concept of the separation of universities from government (Willmott, 1995).

Shortly after the end of the First World War government intervention in higher education become more pronounced when it became apparent that the financial positions of the 'new' city based universities in Birmingham, Manchester and Bristol were in a precarious state requiring significant amounts of state funding to keep them viable. The funding sought by the universities was granted, but at a price. Universities would never again be free from government monitoring as the state sought to ensure that their investments were being well utilised. The University Grants Committee (UGC) set up in 1918 to administer the funding to the universities did provide a buffer between the universities and the government, but this institution was not to stand the test of time as the expansion of student numbers over the past 40 years has made the required level of government funding for higher education increase substantially. The boom in number of students entering higher education following the Robbins Report in 1963 was substantial and has impacted on the very nature of academia itself. The numbers of students entering higher education continued to climb steeply throughout the next three decades rising from 618,000 in 1970 to 1,720,000 in 1995 (Trowler, 1998).

The 1990's witnessed a considerable increase in the number of students enrolled in higher education. In 1988 there were around 600,000 full-time students in higher education, by 1995 this number had nearly doubled reaching 1,100,000. Alongside the increase in numbers the diversity of student also increased. Over a similar period the proportion of women on undergraduate degree programmes has increased from 42% to 52%, mirroring their representation within the overall population. The proportion of 'mature' students on undergraduate degrees has increased from 15% to 29%. The proportions of students from working class backgrounds and from the ethnic minorities have also increased.

Alongside the increase in the numbers of students entering higher education and the number of institutions offering undergraduate education has come an increased need for state financial support. The additional cost of higher education has led to a need for universities to consider carefully ways in which they can improve their efficiency in an attempt to reduce unit costs. As discussed in chapter 2, Dearing explicitly recognised the improvements in efficiency that have characterised higher education in recent years:

> 'increasing workloads and outputs at a time of declining unit resources have been a feature across the system' (Dearing, 1997, 14.16).

Governments have also sought to more closely monitor the quality of the product of which increasing amounts of state funding is supporting. As part of the 1992 Further and Higher Education Act the principle of formalised monitoring of quality in higher education was enshrined as a feature of higher education. One element of the measurement of quality in higher education, namely: the student feedback ratings that students give their lecturers, provides the focus of this research.

Despite the premise of the introduction of teaching quality assessments being to readdress the balance between the priority given within academic departments from research towards teaching, the collection of student feedback data in U.K. universities has not been without its critics. The Higher Education Quality Council refer to 'tension' over the collection of student feedback data in many of the 'Academic Audits' that they have conducted in U.K. universities.

This research aimed to explore whether lecturers should have confidence in the ability of student feedback data to play a useful part in improving the student's learning experiences, by providing a valid and worthwhile guide into how lecturers can improve the quality of modules and their own teaching to enhance the learning process. The research also aimed to offer guidance in the presentation and interpretation of the data, highlighting issues that need to be considered when determining action to be taken in light of the data. It was hoped that the research would be able to facilitate a greater understanding of the subtleties of the data, so that decisions based on the data could be taken on a more informed basis.

The findings of this research are particularly important for two reasons:

- the timing of the study;
- the lack of published statistical analysis on student feedback data collected in U.K. universities.

Although universities that systematically collect student feedback are likely to conduct in-house analysis of the data, there has been little published research that systematically analyses the student feedback data collected in U.K. universities. The analysis in this research facilitates a comparison between the student feedback data collected in a U.K. university and the dominant trends reported in the published student feedback literature, which is largely U.S. in origin.

11.2.2 The aims of the research

One of the clearest features of higher education over recent decades has been the increase in its diversity. The proportion of different types of students has increased as has the nature of institutions offering undergraduate education. Higher education has been transformed from a preserve of the intellectually gifted, often privileged youth to a more heterogeneous environment in which students often look towards future employment as a driving motivation for their studies and industry looks to universities as a source, not only of future employees, but also for practical involvement in their current operations. In short, higher education has considerably changed.

Alongside these changes educational psychologists have considerably developed their understanding of the ways students learn and now conceptualise learning into a set of different learning styles which fundamentally differ in the manner in which students approach their studies and what form of teaching they deem most useful. Similarly, the educational psychologists have considerably developed their understanding of the ways lecturers teach and now conceptualise teaching into a set of different teaching styles which fundamentally differ in the aims that the lecturer has when they enter the lecture theatre.

What is highly significant is that these developments in higher education and the associated literature have not formed the basis of published research into student feedback ratings. The issue of diversity resulting in different teaching and learning styles has largely been ignored in the published student feedback literature. This is particularly unfortunate as different learning styles provide a basis for understanding what students are thinking (i.e. what criteria they are applying) when they evaluate lecturers.

The research set out with three principal aims:

1. To examine the extent to which there are groups of students in Business School modules who psychologically perceive the quality of the module and the skills of the lecturer differently, as a result of utilising different learning styles and to consider the consequences of this for student feedback data;
2. To examine whether the psychological concept of contrast effects is applicable to student feedback data, whereby lecturers' student feedback ratings are influenced by the ratings of the other lecturer(s) teaching on a particular module;
3. To examine the impact of external factors on the ratings lecturers receive.

The findings of the research relating to each of these aims will be considered in the following sections.

11.3 The implications of increased diversity in higher education

11.3.1 Assumptions in the published student feedback literature

As discussed in chapter 3, most of the published student feedback literature proceeds on the underlying assumption that it is appropriate to treat a class as a single coherent homogeneous entity. This assumption is manifested in two explicit assumptions in the student feedback literature:

a) There is a direct relationship between teaching and learning (Timpson and Andrew, 1997);
b) There is a 'true' score for each lecturer on a module, with variation of responses merely being random error (Crichton and Doyle, 1975).

The research presented in chapters 6 and 7 explored the validity of each of these assumptions. The findings of this analysis and the implications of the results are discussed in sections 11.3.2 to 11.3.5.

11.3.2 The existence of systematic interrater variance

To examine whether the second of the common assumptions in the student feedback literature outlined in section 11.3.1 was fallacious, research was conducted using both simulated data and actual student feedback data that explored the implications that heterogeneous groups (or clusters) of students in any particular module have for the analysis and interpretation of student feedback ratings. In particular the research questioned the appropriateness of two common practices in the student feedback literature:

- The use of factor analysis on student feedback data;

- The presentation of the results of student feedback data in the form of class average ratings.

The results of this research provide an important contribution to the student feedback literature as they illustrate that both of these practices, so prevalent in the student feedback literature, become rather dubious when systematic interrater variance exists in any particular module.

The analysis presented in chapter 6 challenges the use of factor analysis in modules in which there are distinct groups of students who view the quality of the module and skills of the lecturer(s) differently. A Business School module shown to have a three-cluster structure was subjected to a factor analysis. The KMO value (0.78) produced by the factor analysis suggested that there was a strong factor structure in the module. However when factor analysis was run on each cluster individually the KMO value fell dramatically (0.54, 0.46 and 0.51 for the three clusters respectively). Additional analysis demonstrated that this result could not be explained simply by the reduction in sample size. Random samples were taken from the same data set, with the sample size set to be the same as the average size of the three clusters. The smaller size of these samples lead the KMO values to fall by approximately 17%, a substantially smaller fall in the KMO values than in the three clusters, where the fall was approximately 35%.

Analysis using four sets of simulated data with known cluster structures illustrated that the interplay between factor and cluster structures can lead to misleading results from factor analysis. In particular, high values for the KMO statistic, indicating that the data set has a well-defined structure, can be due to the correlations between cases caused by the cluster structure. In the light of these findings researchers are recommended to initially check for cluster structure in a data set. If a cluster structure is present the results of factor analysis undertaken on the data set should be interpreted cautiously. These results clearly illustrated that heterogeneous groups of students create a potential problem in the analysis and interpretation of student feedback ratings.

One observation that can be immediately drawn from this analysis is that lecturers should be careful not to place too much emphasis on the class average rating as a measure of student satisfaction for a module. Lecturers who receive 'satisfactory' ratings would do well to consider, either through a direct examination of the data, or through their own judgement of the student's reaction to the material in the lecture theatre, whether the 'satisfactory' rating broadly represents the views of the class as a whole, or whether the 'satisfactory' rating masks a situation in which, although a substantial number of students considered the module and teaching to be more than satisfactory, another substantial group of students found the module and teaching to be less than satisfactory.

11.3.3 The implications of differences in students' learning styles

One of the most important findings of this research is the evidence of different learning styles being utilised by students on Business School modules. Entwistle (1994) outlined three different learning styles (see section 2.5). In short, these three learning styles were categorised as a 'deep' approach, a 'surface' approach and a 'strategic' approach. An understanding of these different learning styles is very important as students using different learning styles will approach their studies with different aims and objectives and correspondingly favour a different form of teaching from lecturers.

These differences become crucial for understanding student feedback ratings as students are likely to rate lecturers in terms of the degree to which the lecturer satisfied the particular learning requirements of the student. In other words, if a lecturer is teaching a group of students who differ in the learning style they utilise then the lecturer's teaching will, by definition, satisfy one group of students (i.e. be more congruent with one of the learning styles) rather than the other. For example, students following a 'surface' or 'strategic' approach to their studies are likely to particularly appreciate a well-organised lecturer who provides the students with clearly structured material and frequent printed handouts. These aspects of lecturing are likely to be less highly prioritised by students following a 'deep' approach to their studies who are likely to particularly appreciate (and therefore rate highly) lecturers who stimulate and challenge them and provide them with the type of material that encourages them to develop their own understanding through their own reading.

The presence of different learning styles can also have implications for the relationship between the module level and lecturers' student feedback ratings. Students entering higher education are likely to find that the academic requirements placed on them in their first undergraduate year differ in emphasis to those placed on them in their final year at school. A greater emphasis on independent learning and independent thinking, as well as the requirement in many undergraduate modules for the student to be responsible for collecting together their own material, may make a proportion of students feel uncomfortable and could result in lecturers teaching first year undergraduates receiving lower ratings compared to lecturers teaching students in later stages of their academic studies.

The Business School data does not, however, seem to report this potential problem. There is little systematic difference between lecturers' ratings and the undergraduate levels of study. For elements of teaching in which there is a significant difference in the ratings given by first and second year undergraduates, first year undergraduates tend to rate lecturers higher than second year undergraduates. This can be considered a good result for the Business School as it may constitute evidence that the initial intellectual adjustment process between school and university generally works well for Business School students.

One of the fundamental implications for lecturers planning their teaching, given the presence of different learning styles in a class, is that lecturers need to decide whether the module and teaching style should be developed to reflect the dominant learning style utilised by students on the module or whether teaching styles should remain independent of students' learning style preferences in effect forcing students to adapt to new ways of understanding their subjects. This second option has the disadvantage that some students might struggle to understand and cope with the demands of the discipline. Alternatively, a failure to encourage students to be more active in their approaches to studying potentially denies the better students the opportunity to develop not only their understanding of particular material, but also their general cognitive capacities.

This decision becomes important for student feedback as the design of the student feedback form is likely to go some way to determine how lecturers will react to this dilemma. Some implications of different teaching and learning styles for the questions on the Business School student feedback form are considered in sections 11.5 and 11.8.

11.3.4 The implications of differences in lecturers' teaching styles

The research published in the educational psychology literature, discussed in section 2.6, categorises teaching into two broad styles. These are the 'instruction paradigm' in which the lecturer aims to impart knowledge to the students and the 'learning paradigm' in which the lecturer aims to act as a facilitator in aiding the student develop their own understanding of the material. It is important to realise that these are not simply slight variations in style, but are fundamentally different approaches to teaching that are underpinned by a different philosophy regarding the central purposes of teaching.

Evidence that lecturers employing a 'learning paradigm' form of teaching are disadvantaged by the Business School student feedback form and the implications of this for the questions asked on the form are discussed in section 11.5. For immediate purposes it is important to realise that the presence of different teaching styles (alongside the presence of different learning styles) highlights the dynamic nature of the teaching and learning process. It is therefore important that the student feedback system itself, as well as the interpretation of the student feedback data, adequately reflects this dynamic complexity. The importance of the interaction between the student and the lecturer will be discussed in section 11.3.5.

11.3.5 The importance of the interaction between student and lecturer

One of the most important findings of this research is that it is inappropriate to attempt to understand student feedback ratings simply in terms of *either* the student *or* the performance of the lecturer. What student evaluations of lecturing measure is the combination of the lecturer's performance, the student's preferences and the personal interaction between the lecturer and the student.

This research has pinpointed two types of interaction effect relevant to lecturers' student feedback ratings:

- The interpersonal dynamics between the lecturer and the student;
- The compatibility of the students' learning style and the lecturers' teaching style.

As was discussed in section 8.7, there is evidence from interviews with Business School lecturers that the level of trust between the student and the lecturer is likely to impact on the ratings that the student gives the lecturer.

Trust is known to be a key element of facilitating learning (see Rogers, 1993) and takes time to develop. There is, therefore, good reason to believe that the amount of contact that the students have with a lecturer over the duration of a module is likely to influence how they feel towards their lecturers. This is reflected in the research presented in chapter 8 which showed that the proportion of lecturing hours undertaken by a lecturer on a module significantly influences the ratings that lecturers receive from their students.

An important implication of this is that departmental managers should avoid (wherever possible) asking a lecturer to take a small part of a module when designing their teaching schedules. Such a practice is likely to be disliked by the students and could be damaging to their studies. If the students have less trust in the lecturer, they may as a consequence be less likely to commit themselves to the pursuit of knowledge suggested by that lecturer. As a result, the students knowledge and understanding of these aspects of the module material may not develop to as high a level as it could.

A policy of requiring lecturers to teach on small proportions of a module is also likely to be counter productive for the lecturer concerned as the students are not likely to rate them in a favourable way. A very stark example of this occurred in one module incorporated in the data set used in this research. The lecturer concerned who is among the highest scoring of all lecturers, consistently returning scores exceeding 4.8, returned a score of under 3.

Another issue raised by the importance of trust between the students and the lecturer is the most desirable size of class which to lecture. This is something that carries important implications for senior university managers. The evidence presented in chapter 10 clearly shows that class size significantly influences lecturers' student feedback ratings. Lecturers teaching smaller classes tended to receive higher ratings than lecturers teaching larger classes. The relationship between class size and lecturers' ratings in the Business School student feedback data is shown to be relatively linear, implying that the ratings are increasingly negatively influenced the larger the class size becomes. The larger the size of the class the more difficult it becomes to generate a sense of rapport with the students, the students are likely to feel more remote from the lecturer and consequently their degree of trust for the lecturer is likely to be less than it would be if the class was smaller.

As discussed in chapter 10, this has important implications for departmental and university senior managers in terms of the optimal level student enrolment. Universities need to be careful not to recruit too many students onto degree programmes. Although each additional student brings additional income to the university, there is a price in terms of the student's educational experience, which would seem to deteriorate as the class size increases. Departmental managers may need to carefully consider upper limits on class size and where necessary consider splitting very large modules into two, or ultimately the student experience may suffer to an extent where it may become difficult to attract students, particularly the high-fee paying oversees students, to read in a particular academic department.

These findings argue for lecturers making a deliberate effort to develop a rapport with their students fostering as much trust as they can between themselves and the students. The need for a lecturer to have the confidence of the student is further enhanced when one appreciates the importance motivating the student. A student is likely to need to be adequately motivated to consider utilising a 'deep' approach to their studies.

11.4 The implications of contrast effects in student feedback ratings

Another important contribution that this thesis makes to the student feedback literature is the examination of the effects on student feedback data of modules that are taught by more than one lecturer. The student feedback data in the Business School provides the opportunity to examine these effects as a considerable proportion of the modules are taught by more than one lecturer (144 modules out of 305 - 47%). The research contributes to the student feedback literature by introducing two new variables, not previously reported in the literature, that are shown to influence the student feedback ratings lecturers receive:

- The proportion of the lecturing hours undertaken by a lecturer on a particular module;
- The effect on a lecturer's ratings of the ratings of the lecturer(s) with whom they teach on a particular module.

One worrying aspect of the initial analysis of the data was that lecturers tended to receive higher ratings when they appeared as lecturer 1 compared to when they appeared as lecturer 2 on the student feedback form. Ratings when they appeared as lecturer 2 were themselves higher than ratings when the lecturer appeared as lecturer 3. This was a potential problem as it raised questions regarding the validity of the student feedback form. If an alternative explanation was not found this analysis would be evidence that the ratings lecturers received were influenced by the design of the form, a factor unrelated to the quality of the teaching provided by the lecturers. Interviews with several Business School lecturers suggested that an alternative variable, i.e. the proportion of lecturing hours undertaken by a lecturer on a module, rather than the lecturers position on the form accounted for the systematic difference in lecturers' ratings between the three positions on the student feedback form.

Regression analysis illustrated that once the ratio of lecturing hours was added into the model, the position that the lecturer appeared on the student feedback form was no longer a significant predictor of a lecturer's ratings. The effect on student feedback ratings of the ratio of lecturing hours was shown to be small, accounting for 3.7% of the variance in ratings. This rose to 5.6% once lecturers who always teach alone were removed from the analysis, a necessary adjustment, as these lecturers have no variation in their ratio of lecturing hours.

The effect on a lecturer's ratings due to the ratings of the other lecturer(s) with whom they teach on a particular module was summarised in chapter 9 as:

> *If a lecturer is teaching with someone who is usually perceived to be better than they are, their ratings will be lower than their norm as they are being evaluated relative to the other lecturer(s) on that module.*

Similarly:

> *If a lecturer is teaching with someone who is usually perceived to be worse than they are, their ratings will be higher than their norm as they are being evaluated relative to the other lecturer(s) on that module.*

Regression analysis using all the modules taught by more than one lecturer showed that this contrast effect was present in the Business School student feedback data. The effect was shown to be highly significant. When modelled against DIFF: the difference between a lecturer's average rating on a particular module and their overall average for a particular question, the contrast effect accounts for 26.9% of the variation in DIFF in the model that combined all the modules taught by more that one lecturer.

In light of these findings, it is recommended that little emphasis should be placed on a lecturer's ratings from a single module, in which the lecturing load was shared with one or two other lecturers. The evidence shows that a lecturer's ratings may vary between modules based simply on who they teach with on a particular module. Furthermore, in situations where a lecturer receives student feedback ratings lower than their norm, then these too should not be interpreted as necessarily implying that the lecturer's performance was weaker than their norm. This is particularly so if they were teaching alongside a lecturer who regularly scores higher than they do.

The need to interpret ratings carefully is especially important for novice lecturers. Novice lecturers are likely to be more closely evaluated as part of their probation assessment. Managers should be particularly careful when considering the ratings received by probationary lecturers, especially if they are teaching alongside a highly rated lecturer.

As discussed in chapter 10, the significant influence that the ratings of the lecturer(s) with whom a lecturer shares lecturing responsibilities on a particular module stress the need for lecturers to be more aware of their colleagues teaching. This is particularly important in the Business School given that teaching as a team is common. The sharing of lecturing strategies and ideas should be encouraged. If this were to happen lecturers might develop new, potentially innovative, lecturing techniques and module formats. The suitability and effectiveness of innovative teaching practices could be measured through lecturers' student feedback ratings.

In modules taught by more than one lecturer, it is also important that the teaching is organised in such a way as the student is presented with a consistent approach to the material. Were there to be substantial differences in the teaching styles that students face on a particular module, there is a potential danger that the students will focus their efforts on the part of the module taught by the lecturer whose teaching style most suited their learning style.

11.5 The relevance of the questions on the student feedback form

To develop a student feedback form that aids lecturers in developing their teaching in the Business School, departmental managers need to be conceptually clear on what the most desirable form of teaching is (i.e. whether they wish to encourage a model closer to the 'instruction paradigm' or the 'learning paradigm').

Power (1997) discussed the mechanisms whereby the introduction of an evaluation mechanism will affect the behaviour of those being evaluated. Power argued that those being evaluated are likely to adjust their behaviour to suit their perceptions of the aims of the evaluation. Furthermore, on occasions the audit process can come to dominate the organisation to such an extent, the thinking of those within the organisation becomes reshaped by the audit itself. Once this is appreciated it becomes clear that the nature of the student feedback system and the design of the instrument used to collect the student feedback has important implications for the teaching style used.

As discussed in section 11.3.3, there are three broad types of learning style utilised by students (i.e. a 'surface' approach, a 'deep' approach or a 'strategic' approach). For university level education the 'deep' approach would generally be considered more desirable than a 'surface' approach. It is understandable that some students enter higher education desiring to employ a 'surface' approach, as this is likely to reflect how they successfully accomplished their previous achievements necessary for their degree enrolment. It would not however, be generally considered desirable for a student to exit higher education with a 'surface' approach to learning. Therefore at some point, ideally as early as possible, students need to be encouraged to adapt their learning style towards a 'deep' approach.

The design of the student feedback system and the student feedback form can be used to guide lecturers in the learning style they encourage their students to follow. If one follows the logic of Power's (1997) argument to its conclusion the design of the student feedback form will, rather than may, influence the learning style that lecturers encourage their student to follow. This reflects the likelihood that many lecturers will seek to score as higher ratings as possible, therefore they will look to the questions on the form and adapt their teaching (as far as possible) to reflect the focus of the questions on which they will be rated.

As discussed in chapter 10, the modelling of lecturers' ratings provided evidence that lecturers who employ a 'learning paradigm' form of teaching may be disadvantaged by the current wording of the student feedback form during the period covered in this research. The main evidence for this was that lecturers' student feedback ratings were shown to be significantly influenced by the lecturer's age and the lecturer's rank, with older and more senior lecturers tending to receive lower ratings than younger and less senior lecturers. As discussed in section 10.9, this result may be explained by Kugel's (1993) conceptualisation of the development of lecturers' teaching technique. If Kugel is correct in his argument that as a generality lecturers' teaching techniques tend to adapt from being akin to 'instruction paradigm' to become more akin to a 'learning paradigm' form of teaching as they develop, older and more experienced lecturers will consequently tend to be more valued by students utilising a 'deep' approach to their studies.

The incidence of bias against lecturers who follow a 'learning paradigm' form of teaching may not be considered a problem by department managers, but does clearly reinforce the need for departmental managers to have a clear policy as to the type of teaching they most wish to encourage.

There are several questions on the Business School student feedback form that do reflect aspects of teaching important to lecturers employing a 'learning paradigm' form of teaching favoured by students utilising a 'deep' approach to their studies. Examples include:

- Q3 – The academic content of the module was stimulating;
- Q7 – The subject material in this module was challenging and stimulating;
- Q8 – The subject material in this module was presented at about the right pace;
- Q16 – The lecturer encouraged participation and questions.

Most of the questions though (and in particular the majority of the lecturer-specific questions) reflect lecturing behaviour particularly important to lecturers employing an 'instruction paradigm' form of teaching favoured by students utilising a 'surface' approach to their studies. Examples include:

- Q1 – I was made aware of the module's aims and objectives;
- Q4 – I understood the assessment requirements of the module;
- Q10 – The teaching was well organised;
- Q13 – The lecturer was well prepared;
- Q15 – The lecturer explained the subject clearly.

These questions reflect an 'objectives-driven model' (Kolitch and Dean, 1999). The underlying principle of an 'objectives-driven model' is that:

> 'education is a process intended to bring about certain observable and measurable changes in students' (Kolitch and Dean, 1999, 32).

Kolitch and Dean are not in favour of such an approach arguing that:

> 'In educational practice, however, the objectives are often narrowed to procedural and superficial learning outcomes' (Kolitch and Dean, 1999, 32).

If departmental managers wish to encourage the employment of a 'learning paradigm' form of teaching by lecturers in the Business School, a lecturer-specific question that directly asks the student the extent to which the lecturer aided the student in developing their own understanding would be appropriate. One of the questions:

- Q2 – The teaching methods for this module made me understand the subject;

does begin to reflect these issues, but it would need to be included amongst the lecturer-specific questions to allow, in modules taught by more than one lecturer, data specific to each lecturer could be collected.

The most important issue is whether the aim of the evaluation is to measure teaching quality in terms of set criteria for teaching, or whether it is more beneficial to measure teaching quality in terms of the extent to which the teaching actively facilitated the students intellectual development. This decision is irrecoverably linked to the wider issue of which style of teaching departmental managers consider most appropriate in the specific context of a particular academic department.

Although the encouragement of students to develop a 'deep' approach to their studies may seem to be clearly preferable to the student adopting a 'surface' approach, there is an important caveat. A decision to actively encourage lecturers to promote a 'deep' approach to their studies needs to be taken with consideration to the wider course structure in place in the Business School. A semesterised system with the academic year split into two and individual modules lasting about four months may mitigate against student being able to adopt a 'deep' approach to their studies. Furthermore, (Entwistle and Ramsden, 1983) argue that the nature of module assessment will directly impact on the approach taken to learning. This needs to be borne in mind in the decision over both the prevalence of coursework in Business School modules and its nature.

11.6 Overview of the variables that most influence a lecturer's student feedback ratings

One of the aims of the research was to investigate which variables influence the ratings that lecturers receive. In light of the analyses presented in chapter 10 individual lecturers and managers should take the following statistical findings into account when they are interpreting student feedback ratings. A more sophisticated appreciation of the subtleties of the data should facilitate a fairer interpretation of lecturers' ratings and potentially increase the confidence that lecturers and departmental managers have in student feedback data.

The research provided strong evidence to suggest that the following relationships exist for all the lecturer-specific questions:

- There is a high degree of relativity in lecturers' ratings in modules taught by more than one lecturer. Lecturers are likely to receive ratings below their norm when they teach alongside a lecturer who generally receives higher ratings than they do. The converse is also shown to be true;
- The proportion of lecturing hours undertaken by a lecturer on a module is significantly positively associated with a lecturer's ratings. The greater the proportion of lecturing load undertaken by a lecturer on a module the higher their ratings are likely to be;
- The number of registered students on a module is significantly negatively associated with lecturers' ratings. Lecturers teaching on modules with a larger number of students are likely to receive lower ratings than lecturers teaching on modules with a smaller number of students. There is no evidence that this relationship is non-linear;
- Postgraduate students tend to rate lecturers significantly higher than undergraduate students. The high level of satisfaction of taught course

postgraduate students for the provision made available to them has important implications for the Business School, especially in terms of the viability of financially lucrative taught course postgraduate degree programmes (see discussion in section 10.9).

The research provided strong evidence to suggest that the following relationships exist for some of the lecturer-specific questions:

- For lecturer's enthusiasm (Q14), lecturer's encouragement of participation (Q16) and lecturer's availability (Q18) final year undergraduate students rate lecturers significantly higher than first year students. It is not the case however that the ratings that students give lecturers tend to steadily increase as the students progress through their degree programme. There is evidence that for lecturer's clarity (Q15), lecturer's encouragement of participation (Q16) and lecturer's overall effectiveness (Q17) first year undergraduates rate lecturers significantly higher than second year undergraduates;
- Lecturers teaching in subject areas that have a less technical focus such as banking, human resource management and marketing tend to receive significantly higher ratings than those teaching accounting modules. In contrast, lecturers teaching on strategic management and quantitative methods modules tend to receive significantly lower ratings than those teaching accounting modules;
- Business School students rate Business School lecturers higher on the lecturer's availability (Q18) compared to students in other departments. This reflects the lack of geographical proximity of non-Business School students and Business School lecturers. There is no significant difference between ratings given to Business School lecturers by Business School students compared to students registered in other departments for any of the other lecturer-specific questions;
- A lecturer's rank is significantly negatively associated with the ratings they receive. Those ranked as senior lecturers or professors tend to receive lower ratings on each of the lecturer-specific questions (other than for a lecturer's preparation - Q13 or enthusiasm - Q14) than standard lecturers;
- A lecturer's age is significantly negatively associated with the ratings they receive. Older lecturers tend to receive lower ratings on each of the lecturer-specific

questions (other than availability - Q18) compared to younger lecturers. There is no evidence that this relationship is non-linear;

- A lecturer's experience at Loughborough University is significantly negatively associated with the ratings a lecturer receives for preparation (Q13) and significantly positively correlated with the ratings that lecturer's receives for enthusiasm (Q14) and encouragement of participation and questions (Q16).

One of the limitations of this research is that that the data used was drawn from a single academic department. This constraint on the breadth of the data was deemed necessary at an early stage of the research given the sensitive nature data, in particular potential sensibilities as to how the data might be used. This raises the issue of how generalisable are the results of the Business School data to the wider population. Extensive statistically based student feedback research outside of the U.K. provides a useful point of comparison, through which to validate the statistical findings reported in this research.

The analysis of the student feedback data from the Business School generally supports the findings in the published student feedback literature, which are predominantly U.S. in origin. As discussed in section 10.8 the effects on student feedback ratings of class size (i.e. lecturers teaching large classes tend to receive lower ratings than lecturers teaching small classes), module level (i.e. postgraduate students tend to rate lecturers higher than undergraduate students) and the lecturer's age (i.e. older lecturers tend to receive lower ratings than younger lecturers) illustrated in the Business School data are in line with the dominant trends in the published literature. The published student feedback literature that explores the relationship between the lecturer's experience and the ratings they receive is too contradictory to determine a dominant trend. The only variable shown to effect the Business School data in a way different to the dominant trend in the published student feedback literature is the relationship between the lecturer's rank and the lecturer's ratings. The majority of published studies that explore the effects of lecturer's rank on student feedback ratings have reported an insignificant result. In the published studies which report a significant relationship the association is generally positive, i.e. lecturers of higher rank receive higher ratings. Reasons for the Business School data not supporting these findings are discussed in section 11.5.

11.7 Recommendations for managerial practice

Broadfoot (1995) offers a wary interpretation of the role of 'performance indicators':

> 'Centring on the generation of more and more information in the form of "performance indicators" at every level of the system, the new rationality of the information society is that embedded in the concept of total quality management: the generation of both targets and accounts about performance in relation to those targets. Underlying such procedures is the assumption that the generation of the account *in itself* will ensure the desired effect' (Broadfoot, 1995, p.64, emphasis in original).

This perception of the nature of performance indicators highlights the need to appreciate that the results of any 'performance indicator', in this case student feedback ratings, need to be carefully considered, to be contextualised and if they are to serve the purpose of predicating positive change, be actively used. Above all student feedback ratings must never be considered to 'speak for themselves'.

With this in mind, this section will consider the recommendations for managerial practice that arise out of this research. To begin with, some of the current Business School policies can be praised. The principle of encouraging lecturers to read through the written comments on the student feedback forms is necessary for lecturers to understand whether there are groups of students with different learning styles present in the class. Lecturers can then use this knowledge of the differences in students' learning styles when planning their teaching for a particular module. The presence of different learning styles in Business School modules highlights the need for lecturers to be flexible in their teaching style. Lecturers can also consider whether it is possible for them to structure their material in a way that touches all the learning styles in a single session. Also, the requirement on lecturers to prepare formalised reports for modules that have received less than adequate ratings from the students again encourages an active response to the student feedback ratings received by lecturers.

Bearing these current practices in mind, the following nine recommendations for managerial practice can be offered:

1. *Fine distinctions in lecturers' ratings are not appropriate*

The findings from the modelling of the Business School data reported in section 11.6 showed that various variables (e.g. class size, module level, the proportion of lecturing hours a lecturer takes on a module and who they lecture alongside) influence the student feedback ratings lecturers receive. Lecturers' ratings year to year (even for the same module) are also likely to vary. The implication of this for managerial use of student feedback ratings is that there is little to be gained by placing too much emphasis on small differences in lecturers' ratings.

This is particularly the case for those lecturers who occupy the middle three quintiles of overall ratings. These lecturers are clearly performing at least satisfactorily and can be left to make their own judgements using both the statistical ratings and the written comments they receive from students to develop their module and lecturing. This conclusion reflects the work of the Harvard organisational behaviourist Michael Beer, speaking on the use appraisal practices in U.S. industry, Beer (1990) comments:

> 'the idea of trying to differentiate on some fine-grained system is ridiculous. You can't make those kinds of discriminations on total performance. Total performance is a complex collage of competency, skills, and knowledge. Most people are in the middle, and what we need to do is to identify the really outstanding performers and the really poor performers and try not to pretend that one can make fine-line differentiations objectively' (Michael Beer, quoted in Gabor, 1990, 251-252).

2. *To adequately understand a lecturer's student ratings it is necessary to contextualise the data*

To come to a valid assessment of a lecturer's performance based on the results of their student feedback it is important to understand lecturers' ratings in context of the nature of the module. The influences on student feedback discussed in chapter 10 and summarised in section 11.6 need to be appreciated by departmental managers and university personnel officers.

3. *Personnel decisions should not be made on the performances of lecturers whose ratings fall in the middle quintiles, but can be taken on those returning ratings in either the highest or lowest quintile*

The managerial time required and the ambiguity in the meaning of ratings that are finely differentiated would make decisions taken on lecturers falling in the middle quintiles dubious, potentially reflecting the precise circumstances of a particular module rather than a more general underlying measure as to the quality of a particular lecturer.

However, for lecturers whose scores fall in either the highest or lowest quintile their ratings are distinctive enough to warrant analysis. Lecturers who return very low ratings are either genuinely poor, or have had to operate in circumstances that has made their task significantly more challenging than that faced by other lecturers. The data presented in this research should make it relatively easy to see whether a low scoring lecturer has such grounds for defending their module and lecturing performance. Similarly, it is also important to examine the performance of lecturers who are scoring very highly. In this case the lecturer is either particularly good (and could be promoted as a role model for other lecturers) or is getting high ratings through some other means that Business School managers may well consider undesirable and should be aware of, such as engaging in lenient grading (see for example 'the grading leniency hypothesis', Greenwald and Gillmore, 1997).

For lecturers returning either very high or very low scores it is advisable for managers to take the time to understand why the lecturers are returning these ratings. It should be relatively quick to do so and the lessons that can be gleaned from such analysis are likely to be important to the management of teaching responsibilities in the Business School and allow students' attitudes expressed about lecturers outside of formalised student feedback to be placed in context.

4. *It is inadvisable to create league tables of lecturers' performance*

The ratings that lecturers receive are not absolutes, but rather relative to those they lecture alongside, as well as other influences summarised in section 11.6. It would be erroneous to compare a rating of 3.9 on a compulsory module taught to a large number of first year undergraduate students unfavourably to a rating of 4.1 on a postgraduate module taken by a small number of students.

Once lecturers appreciate the effect of the influences on their ratings, league tables may create resentment amongst lecturers. This may particularly be the case for lecturers who have been lecturing to large classes to students in the early phase of their studies who find themselves ranked below other lecturers who have been predominantly teaching postgraduate students in small classes.

5. *Decision-making should be made at the lowest level possible*

The need for departmental managers to have an awareness of the subtleties in the data makes it important for decisions taken in light of the results of student feedback data to be taken at the lowest possible level. This should give the greatest opportunity for decision makers to have as deep an understanding of 'localised' factors that could have influenced a lecturer's student feedback ratings as possible. This may help build the level of trust that lecturers have in the student feedback process.

6. *Managers should be seen to have a 'lighter hand' with the data*

There are several benefits of managers approaching student feedback with a 'lighter hand'. Firstly, it will reduce the cost of the student feedback system, particularly in terms of demands on senior staff time. Secondly, it would have the effect of potentially strengthening the confidence in the system itself by reducing any feeling in lecturers' minds of the decision-making being remote from them. Any conceptualisation in the minds of lecturers of student feedback being an imposition on them could be allayed by placing more emphasis on lecturers having increased control over their own ratings, bringing student feedback closer to the lecturer.

7. Lecturers need to have more comparative data made available to them

Whilst respecting the confidentiality of each lecturer's student feedback ratings, lecturers should be provided with more comparative data to aid them in understanding the messages being sent by the students.

As discussed in section 11.3, lecturers need to be made aware of whether there are distinct groups of students in the class who view the quality of the module and the skills of the lecturer differently. Such clustering is likely to indicate that there are different learning styles being employed by the students in the class and these styles are interacting with the lecturer's style in such a manner as to make the lecturing a more effective facilitator of learning for some students than others.

There are also other potential advantages with providing lecturers with more comparative information. This may help to allay any anxieties amongst lecturers who perceive that the presence of a student feedback system reduces their control over the output of their teaching. Lecturers must not see the use of a student feedback system as 'taking power away from them'. This is vital if lecturers are going to use the results of student feedback in a positive manner, i.e. to actively improve the quality of their modules and lecturing. Therefore it must be a clear department policy that the 'ownership' of the data must stay with the lecturer, as it is important that student feedback needs to be perceived by lecturers as ‘enabling’ rather than ‘controlling’.

8. Student feedback data should be collected as late in the module as possible

The evidence presented in this research of the presence of different learning styles in Business School modules and the differences in teaching styles employed by lecturers, makes it important for student feedback data to be collected as late in the module as possible. Lecturers who are employing a 'learning paradigm' form of teaching are likely to be disadvantaged if the student feedback data is collected earlier in the module. This is because lecturers employing a 'learning paradigm' form of teaching are likely to be more demanding of students particularly early in the module and may as a result unsettle some of the students, in particular students utilising a 'surface' learning style. However, for the students who are able to respond to the demands made of them, a 'learning paradigm' form of teaching may prove to be more beneficial for them (and therefore the lecturer should command high ratings). The students may not appreciate this until later in the module after they have overcome the initial challenges posed by the lecturer.

9. *Above all, departmental managers need to be conceptually clear on what the student feedback system is trying to achieve.*

As the mere presence of a student feedback system will have an impact on the behaviour of lecturers, it becomes crucial to gear the student feedback system and the questions on the student feedback form to produce the change deemed most desirable.

11.8 The future of student feedback

An indication of the potential future of student feedback systems in U.K. universities was outlined by the current Minister for Higher Education (Rt. Hon.) Baroness Blackstone speaking to the House of Commons Education and Employment Committee on 1 February 2001. Asked by the committee chairman Barry Sheerman:

> 'do we in the university system today give adequate bonuses and rewards, brownie points, whatever you want to call them, to good teachers?'

Baroness Blackstone replied:

> 'The way in which you should encourage people to take their teaching seriously from the funding point of view is by promoting people when they can demonstrate that not only is their research of high quality but so is their teaching. Similarly, where somebody is an outstanding teacher – and we should collect more evidence about the quality of people's teaching, including directly from students, because they are the people who tend to know usually about that – we should be giving additional increments. We should be definitely celebrating high quality teaching at our universities. I think that there is a case for doing more of that than has perhaps been done in the past.'

The phrase 'additional increments' implicitly suggests that the government is seriously considering introducing a form of performance related pay for lecturers in higher education. In an article published in the *Times Higher Education Supplement* (9 February 2001), the sociologist Frank Furedi criticises Baroness Blackstone for contemplating the link between pay and performance and for the belief that students 'are the people who usually know about that' on several grounds. Furedi argues that:

- 'by transferring the relationship of conflict that characterises the market place, politicians threaten to erode further the collaboration and trust between teacher and student that is the prerequisite of an education';
- 'To treat students as umpires in a beauty contest is likely to have a destructive impact on education';
- 'Lecturers who know that their pay increments are closely linked to the approval they receive from their students will learn to avoid teaching practices that might undermine their popularity';
- 'Lecturers will be more interested in communicating what they think students want to hear than what undergraduates need to master a subject'.

This research and the evidence from the published student feedback literature can be used to refute each of these arguments.

The evidence discussed in this research shows that far from eroding trust between students and lecturers, the use of student feedback data is a useful measure of trust between students and lecturers as it reflects the extent to which trust is present in the student-lecturer relationship. Potentially, student feedback can actually be used to build trust. In chapter 8 it was shown that trust between students and lecturers was built up through the student having confidence in the lecturer's abilities. In particular, the student's confidence in the lecturer's ability to offer correct, reliable information and the demonstration of a detailed, thorough, grasp of the material that reassures the students that the lecturer can be relied upon (i.e. trusted) to guide them through the course and the intellectual challenges of the material, leading the students towards achieving their learning goals. In other words, trust (in terms of student feedback ratings) emanates directly from the student's educational experience; lecturers offering students low quality teaching will not be trusted, whereas lecturers offering students high quality teaching will. Therefore, the evidence simply does not support the notion that the use of student feedback data will erode trust between student and lecturer. If anything, the active use of student feedback data (particularly as a diagnostic tool by lecturers) may increase the degree of trust in higher education.

Furedi's second point misunderstands the basis on which students evaluate lecturers. To use the phrase 'beauty contest' implies that Furedi believes that there are no systematic, rational reasons for students rating lecturers how they do, other than their personal liking of the lecturer, or factors specific to the lecturer but independent of their teaching. This research and the voluminous published student feedback literature, prove that this is not the case. There *are* systematic, fundamental bases for students rating lecturers the way they do and these are very much linked to the quality of the lecturer's teaching.

Once the relationship between students' learning styles and lecturers' teaching styles is appreciated Furedi's third argument is shown to be questionable. The way a lecturer receives high ratings from a particular student is by providing a form of teaching that suits the learning style of that particular student. If a lecturer can 'avoid teaching practices that might undermine their popularity' it implies that a lecturer is both aware of and then takes steps to provide the type of teaching that suits the learning styles of the majority of the students in the class. Such an outcome must be viewed as an act of skill, rather than something to be denigrated. To dispel any doubt about this conclusion one only needs to consider that what Furedi is almost certainly alluding to is that to increase their popularity lecturers may make their lectures and assessment mechanisms easier.

There are two pieces of evidence that make Furedi's argument doubtful:

- Firstly, as outlined in chapter 3, the evidence in the published student feedback literature shows that the difficulty of the module is *positively* correlated with the ratings that lecturers receive. Lecturers teaching modules considered more difficult tend to receive higher ratings (Marsh and Dunkin, 1992; Centra,

1993) therefore attempts to attract higher ratings by cynically making modules easier are unlikely to be successful;

- Secondly, if by easier one means challenging students less and providing more structured guided teaching this may appeal students utilising a 'surface' approach to their studies, but will not be appreciated by (or attract high ratings from) students following a 'deep' approach to their studies.

The way to score high student feedback ratings is to respond to the educational needs of a student, or, in terms of average ratings, to be able to offer teaching that responds to (the often diverse) educational needs of as many students in the class as possible.

The evidence in the student feedback literature suggests that Furedi's fourth point (i.e. lecturers will be more interested in telling students what they believe students want to hear rather than what the students need to understand) is fallacious. Ramsden (1992) writes:

> 'in spite of a whole series of attempts to popularise the view that students can be fooled into giving those lecturers who are superficially attractive presenters of wrong content high ratings as teachers - the existence and prominence of these studies is an intriguing phenomenon itself - it is evident from the correctly controlled enquiries that students rarely fall into the trap. They can easily differentiate the empty performer from the good teacher' (Ramsden, 1992, 90-91).

This year a change of emphasis is taking place in the formal monitoring of higher education. A new methodology for academic reviews dubbed 'lighter touch' is being introduced, initially in Scotland (Donald Macleod, *Guardian,* 30 January 2001).

The existing methodology includes the direct observation of teaching in universities. This involves assessors attending lectures and evaluating the performance of the lecturer. This form of teaching assessment is time consuming and therefore expensive. There is also a degree of subjectivity in these assessments, for although the reviewers have a set of criteria with which to evaluate a lecturer the teaching evaluation remains dependent on a particular reviewer. The important point is that if the monitoring and evaluation of teaching in place within the department is at least satisfactory the evaluation of teaching quality through direct observation by ESR evaluators may be unnecessary. Removing the requirement for ESR evaluators to evaluate teaching would release ESR resources and potentially reduce the cost of the ESR process.

The new ESR methodology states:

> 'academic reviewers may not need to make direct observations of teaching where a subject provider can provide good evidence of good quality delivery. Such evidence is likely to come from ... student questionnaires and other arrangements for gathering feedback' (QAA Handbook for Academic Review, paragraph 70).

There is a clear mutual advantage to both QAA and individual universities in this methodology. Academic departments need to ensure that their existing monitoring and evaluation of teaching satisfactory provides a measure of the extent to which teaching positively aids the learning experiences of students. Alongside student feedback ratings this includes the feedback collected from staff-student committee meetings, the use of peer observation of teaching and the use of external examiners reports, which is a useful method of ensuring that teaching has achieved its aims. In placing the focus on developing a good quality student feedback system academic departments will potentially realise the benefits of using student feedback as a key measure of teaching quality. These benefits include:

- Student feedback is quick to administer, involving the minimum amount of lecturers' time in terms of the administrative requirements;
- The system is relatively cheap to run, with small data collection and data analysis costs;
- The data is objective in the sense that all the students on each module have an equal opportunity to express their views and as each student answers the same close ended questions. It therefore becomes possible to generate a reliable

composite picture of both the within-module variations in students attitudes towards the quality of a particular module and between-module comparisons;

- The use of student feedback data keeps decision making close to those who directly manage the teaching within an academic department. This allows a far more interactive process in the student feedback system, in which department managers can actively adjust the process, for example by amending the questions on the student feedback form, to take account of 'localised' (i.e. departmental specific) factors.

The presence of a good student feedback system in an academic department will limit the need ESR teams to directly monitor teaching quality. The ESR teams will simply need to ensure that the student feedback system functions adequately. This 'lighter touch' will not only reduce the potentially cost of ESR visits, but will also allay some of the criticism of the QAA process itself.

Donald Macleod writing in the *Guardian (*30 January 2001) comments:

> 'the QAA reviews do not enhance teaching quality in universities, rather they are designed to "assure" the public of the quality of educational provision in universities, in contrast to "ensuring" good quality teaching'.

Student feedback systems, in contrast, keep the organisation of the process and the design of the methodology (e.g. the choice of questions on the student feedback form) in the hands of those who are most able to 'ensure' good quality teaching - departmental managers - who are best placed to understand the 'localised' factors in a particular academic department. This knowledge should allow them to design an evaluation mechanism that measures and encourages the aspects of teaching which are considered most likely to facilitate positive learning outcomes, within the context of a specific academic department.

11.9 Limitations of the research and suggestions for future research

One of the main limitations of this research is that it is based on the student feedback data collected in a single department at a single institution. Although the data is extensive, it remains rooted both within an academic discipline and the institutional norms of a particular university. It would be interesting to compare the findings of this research to student feedback data collected in:

- different academic departments, in which different subject material may lead to differing approaches from lecturers and differing emphasis in the student's judgements of the quality of the modules and the skills of the lecturers;
- Business Schools in other universities, in which institutional norms, in terms of module design and teaching styles, may differ.

Another limitation of the research was that the data available for this research lacked any student-specific information (such as the student's educational background and personal details). The lack of this information limited the investigation of the reasons for the cluster structures discussed in chapter 7. It was not necessary in chapter 7 to explain why a cluster structure occurs in any particular module. The aim in chapter 7 was to gauge the prevalence of clusters in the student feedback for Business School modules to judge the extent to which systematic interrater variance (discussed in chapter 6) exists. The absence of student-specific information on the student feedback form creates therefore an opportunity for future research. It would be helpful to be able to extend the analysis in chapter 7 to include an analysis of whether particular student-specific characteristics help to explain the basis of students in any particular module adopting different learning styles. An understanding of the indicators of the likelihood of a student following a particular learning style could then allow a lecturer to more accurately judge the likely mix of learning styles present in any particular module and adjust their style of teaching appropriately. In addition to helping to explain the existence of cluster structures, data on student-specific characteristics could be integrated into the regression models presented in chapter 10 to measure the extent to which student-specific characteristics influence the student feedback ratings that lecturers receive.

Further research is needed to determine the frequency of the particular learning styles in Business School modules. To fully measure the frequency of different learning styles it would be necessary to develop a questionnaire that asked questions from which one could deduce a student's preferred learning style and in what circumstances they would utilised an alternative style. Such research could build on the 'Learning Styles Questionnaire' developed by Honey and Mumford (1982). This information would have important implications for the mix of questions on the student feedback form and the issue of whether it is appropriate to have different student feedback forms (i.e. which have a different emphasis of questions) for the different years of study.

For student feedback data to fulfil the purpose of lecturers utilising the results of student feedback to improve their teaching, it is necessary for lecturers to take an active interest in their student feedback results. Lecturers also need to be willing to consider incorporating any constructive comments made by the students. Further research is needed to examine the extent to which lecturers are willing to *use* their student feedback results for constructive module and teaching improvement. Cynics of student feedback data often claim that the process is a mere form filling exercise, often this attitude is accompanied by the use of the phrase 'happy sheets' to describe evaluation forms. Therefore there is a need to understand how lecturers use their student feedback data. This would go some way to determine whether the potential positive purposes of student feedback data are merely rhetoric or have a basis in reality.

Further research could usefully be undertaken to examine the prevalence of the alternative teaching styles amongst Business School lecturers. To do this systematically one would need to design a questionnaire that was filled in by each lecturer. There are two particular pieces of information that could arise out of such a study that would have a direct bearing on the Business School student feedback data discussed in this research:

- Firstly, it would be interesting to know whether lecturers in different academic areas (subject groups) in the Business School had systematically different styles. In other words, whether particular styles of teaching were more prevalent in some academic areas (subject groups) than others;

- Secondly, it would also be interesting to know whether Business School lecturers systematically varied their teaching style depending on which module level they were lecturing to.

Systematic differences in either of these would have significant implications for the design of the student feedback form used in the Business School and in particular the issue of whether it is advisable to use the same student feedback form for all Business School modules, or alternatively, whether different forms would be suitable for different levels of study or different subject areas.

The final suggestion for future research involves a more psychological investigation into the effects of student feedback ratings. It would be very interesting to investigate the extent to which lecturers were able to gauge the quality of their own lecturing and have an accurate understanding of the strengths and weaknesses of their teaching. The way to do this would be to videotape a lecture and ask the lecturer before they watched the recording of the lecture to fill in a questionnaire in which they rated their teaching and specifically reported the lectures strengths and weaknesses. Then the lecturer would be asked to watch the recording of the lecture and then re-evaluate the lecture. The important point is that the lecturer is now able to view their teaching from a different perspective (i.e. that of an observer rather than that of an actor). What would be particularly interesting is whether there was a significant correlation between the similarity of lecturers' self evaluations before and after viewing the video recording of their lecture and the lecturer's willingness to treat their students ratings seriously by using the results to improve their lecturing and module design.

11.10 Concluding comments

Broadfoot (1996) offers the following, rather telling, observation:

> 'however it is done, the use of formalised assessment procedures for this purpose is an extremely powerful policy mechanism for exerting control over the education system. Whoever has the power to determine the criteria against which such assessments are made has the power to influence the priorities pursued by teachers' (Broadfoot, 1996, p.8).

The implications for the use of student feedback are clear. To implement a constructive and useful student feedback system departmental managers need to have a clear vision of what they wish the student feedback system to achieve. Managers must then ensure that the student feedback mechanism and the student feedback form itself actually encourages the type of behaviour from lecturers that the department seeks to achieve. A potential weakness of student feedback forms is that they are implemented without adequate regard for whether they are responsive to alternative teaching and learning styles. These differences need to be fully understood and reflected in student feedback systems to ensure that lecturers and students react to student feedback in the ways anticipated by those who have introduced them, which should have, at its heart, the objective of aiding lecturers in improving modules and their lecturing.

This research aimed to explore whether lecturers should have confidence in student feedback ratings and to offer lecturers and managers guidance in the interpretation of lecturers' ratings.

Two things must be fully appreciated by lecturers and departmental managers:

- Student feedback ratings need to be contextualised;
- Student feedback ratings reflect an interaction between the student and the lecturer, principally reflecting the degree of congruence between the lecturer's teaching style and the student's learning style.

On the positive side, there is clear evidence (discussed in chapter 7) that the students are differentiating between module conditions and lecturers in their responses to the questions on the student feedback form. Furthermore, the modelling of lecturers' student feedback data presented in chapter 10 clearly illustrated that students actively discriminate between the lecturer-specific questions on the student feedback form. Each of the predictor variables were shown to influence each of the lecturer-specific questions (i.e. different dimensions of teaching) to a different extent. This suggests that the ratings lecturers receive do not simply reflect whether the student generally liked the lecturer, but are an indication that the students are actually responding to each of the questions on the student feedback form by giving independent consideration of their ratings for each dimension of teaching.

Lecturers and departmental managers must, however, appreciate the variables that influence lecturers' ratings. The danger in the use of student feedback ratings comes when the results are used carelessly. For example, bad practices may include using ratings from a single module to evaluate a lecturer and simply relying on class averages to judge the quality of the module. Lecturers and departmental managers need to be aware that students rate lecturers relatively. The other lecturers with which they lecture on a module and the proportion of lecturing hours undertaken by a lecturer on a module significantly influence lecturers' ratings. Furthermore, the level of the module, the class size and the subject area of the module also significantly influence lecturers' ratings. Characteristics of the lecturer, namely, the lecturer's age, rank and experience are shown to significantly influence lecturers' ratings for some of the lecturer-specific questions.

Through quantifying the degree of impact that each of these variables have on lecturers' ratings this research provides the basis for a more sophisticated understanding of lecturers' ratings. By taking these influences into account, decisions based on the data will be fairer and provide a more accurate reflection of the quality of the module and the skills of the lecturer.

Student feedback ratings need to be interpreted carefully, the recommendations for managerial practice discussed in section 11.7 should help departmental managers in this regard. Provided that lecturers and managers appreciate the influences on student feedback ratings, lecturers should have confidence in the merit of student feedback data using it as one of their guides in improving the quality of modules and their teaching.

REFERENCES

Abouserie, R. (1995). Self-esteem and achievement motivation as determinants of students' approaches to studying. *Studies in Higher Education, 20*, 19-26.

Abrami, P. C. and d'Apollonia, S. (1991). Multidimensional students' evaluations of teaching effectiveness - generalizability of "N=1" research: Comment on Marsh (1991). *Journal of Educational Psychology, 83*, 411-415.

Abrami, P. C., d'Apollonia, S. and Rosenfield, S. (1996). The dimensionality of student ratings of instruction : What we know and what we do not. In J.C. Smart (ed.), *Higher Education: Handbook of Theory and Research, Vol. 11*. New York: Agathon Press.

Abrami, P. C., Dickens, W. J., Perry, R. P. and Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? *Journal of Educational Psychology, 72*, 107-118.

Aleamoni, L. M. (1981). Student ratings of instruction. In J. Millman (ed.), *Handbook of Teacher Evaluation.* Beverly Hills, CA: Sage.

Anderson, D. R., Sweeney, D. J. and Williams, T. A. (1996). Statistics for Business and Economics. St. Paul, Minneapolis: West.

Atlas, R. S. and Overall, J. E. (1994). Comparative evaluation of two superior stopping rules for hierarchical cluster analysis. *Psychometrika, 59*, 581-591.

Baker, F. B. and Hubert, L. J. (1975). Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association, 70*, 31-38.

Barney, J. B. and Hoskisson, R. E. (1990). Strategic groups: Untested assertions and research proposals. *Managerial and Decision Economics, 11*, 187-198.

Barr, R. and Tagg, J. (1995). From teaching to learning - a new paradigm for

Beale, E. M. L. (1969). *Cluster Analysis.* London: Scientific Control Systems.

Becher, T. (1999). Quality in the professions. *Studies in Higher Education, 24*, 225-235.

Blackburn, R. and Jarman, J. (1993). Changing inequalities in assess to British universities. *Oxford Review of Education, 19*, 197-215.

Blashfield, R. K. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin, 83*, 377-388.

Boyer, E. (1990). *Scholarship Reconsidered.* New Jersey: Carnegie Foundation for the Advancement of Teaching.

Braskamp, L. A., Brandenburg, D. C., and Ory, J. C. (1985). *Evaluating Teaching Effectiveness: A practical guide.* Beverley Hills, CA: Sage.

Braskamp, L. A. and Ory, J. C. (1994). *Assessing Faculty Work: Enhancing individual and institutional performance.* San Francisco: Jossey-Bass Publishers.

Broadfoot, P.M. (1996). *Education, Assessment and Society: A sociological analysis.* Buckingham: Open University Press.

Broder, J.M. and Dorfman, J.H. (1994). Determinants of teaching quality: What's important to students? *Research in Higher Education, 35*, 235-249.

Buckley, M. R., Villanova, P. and Benson, P. G. (1989). Contrast effects in performance ratings: Another look across time. *Applied Psychology: An International Review, 38*, 131-143.

Calinski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3,* 1-27.

Cashin, W. E. (1988). Student ratings of teaching: A summary of the research. IDEA paper No. 20. Manhattan, Kansas: Kansas State University, Center for Faculty Evaluation and Development.

Cashin, W. E. (1995). Student ratings of teaching: The research revisited. IDEA paper No. 32. Manhattan, Kansas: Kansas State University, Center for Faculty Evaluation and Development.

Cashin, W. E., Downey, R. G. and Sixbury, G. R. (1994). Global and specific ratings of teaching effectiveness and their relation to course objectives: Reply to Marsh (1994). *Journal of Educational Psychology, 86*, 649-657.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Centra, J. A. (1973). Effectiveness of student feedback in modifying college instruction. *Journal of Educational Psychology, 65*, 395-401.

Centra, J. A. (1977). Student ratings of instruction and their relationship to student learning. *American Educational Research Journal, 14*, 17-24.

Centra, J. A. (1993). *Reflective Faculty Evaluation.* San Francisco: Jossey-Bass Publishers.

Centra, J. A. and Creech, F. R. (1976). *The relationship between student, teacher and course characteristics and student ratings of teacher effectiveness.* (Project Report 76-1). Princeton, NJ: Educational Testing Service.

Chang, W-C. (1983). On using principal components before separating a mixture of two multivariate normal distributions. *Applied Statistics, 32*, 267-275.

Clarkson, P. C. (1984). Papua New Guinea students' perceptions of mathematics lecturers. *Journal of Educational Psychology, 76*, 1386-1395.

Cobban, A. B. (1988). *The Medieval Universities, Oxford and Cambridge C.1500.* Aldershot: Gower Publishing.

Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta analysis. *Research in Higher Education, 13*, 321-341.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51,* 281-309.

Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society, Series A, 134,* 321-367.

Crichton, L. I. and Doyle, K. O. (1975). Reliability of ratings. Unpublished manuscript, Measurement Services Center, University of Minnesota.

Cronbach, L. J. (1984). Essentials of psychological testing. New York, NY: Harper & Row.

Crumbley, L. D. and Fliedner, E. (1995). Accounting administrators' perceptions of student evaluation of teaching (SET) information. Unpublished manuscript, Department of Accounting, Texas A&M University.

Cullen, D. J., Pearson, M., Saha, L. J. and Spear, R. H. (1994). *Establishing Effective Ph.D. Supervision.* Canberra: Department of Employment, Education and Training, Australian Government Publishing Service.

Dearing, R. (1997). *Higher Education in the Learning Society (The Dearing Report).* London: DfEE.

Douglas, M. (1992). The normative debate and the origins of culture. In M. Douglas (ed.), *Risk and Blame: Essays in cultural theory.* London: Routledge.

Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis.* New York: Wiley.

Entwistle, N. J. (1981). *Styles of Learning and Teaching: An integrated outline of educational psychology for students, teachers, and lecturers.* Chichester: Wiley.

Entwistle, N. J. (1994). *Teaching and the Quality of Learning: What can research and development offer to policy and practice in higher education.* Report of a seminar held 25 November 1994. London: CVCP.

Entwistle, N. J. and Ramsden, P. (1983). *Understanding Student Learning.* London: Croom Helm.

Entwistle, N. J. and Tait, H. (1990). Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. *Higher Education, 19*, 169-194.

Everitt, B. S. (1993). *Cluster Analysis.* London: Edward Arnold.

Fava, J. L. and Velicer, W. F. (1996). The effects of underextraction in factor and component analysis. *Educational and Psychological Measurement, 56*, 907-929.

Feldman, K. A. (1976). Grades and college students' evaluations of their courses and teachers. *Research in Higher Education, 4*, 69-111.

Feldman, K. A. (1977). Consistency and variability among college students in rating their teachers and courses: A review and analysis. *Research in Higher Education, 6*, 223-274.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education, 9*, 199-242.

Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education, 10*, 149-172.

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education, 18*, 3-124.

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education, 21*, 45-116.

Feldman, K. A. (1987). Research productivity and scholarly accomplishment: A review and exploration. *Research in Higher Education, 26*, 227-298.

Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators and external (neutral) observers. *Research in Higher Education, 30*, 137-194.

Feldman, K. A. (1993). College-students views of male and female college-teachers 2: Evidence from students evaluations of their classroom teachers. *Research in Higher Education, 34*, 151-211.

Feldman, K. A. (1998). Reflections on the study of effective college teaching and student ratings : One continuing quest and two unresolved issues. In J.C. Smart (ed.) *Higher Education: Handbook of Theory and Research, Vol. 13.* New York: Agathon Press.

Firth, M. (1979). Impact of work experience on the validity of student evaluations of teaching effectiveness. *Journal of Educational Psychology, 71*, 726-730.

Ford, J. K., MacCallum, R. C. and Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*, 291-314.

Freire, P. (1970). *Pedagogy of the Oppressed.* New York: Seabury.

Frey, P. W. (1978). A two-dimensional analysis of student ratings of instruction. *Research in Higher Education, 9*, 69-91.

Frey, P. W., Leonard, D. W. and Beatty, W. W. (1975). Student ratings of instruction: Validation research. *American Educational Research Journal, 12*, 435-447.

Furedi, F. (2001). Under the hammer. *Times Higher Education Supplement* (9 February 2001).

Gabor, A. (1990). *The Man Who Discovered Quality.* New York: Penguin.

Gibbs, G. Lucas, L. and Simonite, V. (1996). Class size and student performance: 1984-94. *Studies in Higher Education, 21*, 261-273.

Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society A, 159*, 385-443.

Gorsuch, R. L. (1983). *Factor Analysis.* Hillsdale, NJ: Erlbaum.

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research, 25*, 33-39.

Green, D., Brannigan, C., Mazelan, P. and Giles, L. (1994). In S. Haselgrove (ed.), *The Student Experience.* Buckingham: Open University Press.

Greenwald, A. G. and Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52*, 1209-1217.

Grey, C. and Mitev, N. (1995). Management Education: A polemic. *Management Learning, 26,* 73-90.

Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. (1995). *Multivariate Data Analysis: With readings.* New Jersey: Prentice Hall.

Hativa, N. (1995). The department-wide approach to improving faculty instruction in higher education: A qualitative evaluation. *Research in Higher Education, 36*, 377-413.

Hendry, G. and King, R. (1994). On the theory of learning and knowledge: Educational implications of advances in neuroscience. *Science Education, 78*, 223-253.

Henkel, M. (1997). Teaching quality assessments: Public accountability and academic autonomy in higher education. *Evaluation, 3*, 9-23.

Higher Education Quality Council (1994). *Learning from Audit.* London: Higher Education Quality Council.

Hill, F. M. (1995). Managing service quality in higher education: The role of the student as primary consumer. *Quality Assurance in Education, 3*, 10-21.

Honey, P. and Mumford, A. (1982). The Manual of Learning Styles. Maidenhead: Peter Honey.

House of Commons Select Committee on Education and Employment minutes of evidence taken on 1 February 2001. (Comments by Minister for Higher Education (Rt. Hon.) Baroness Blackstone).

Hubert, L. J. and Levin, J. R. (1976). A general statistical framework for assessing categorical clustering in free recall. *Psychological Bulletin, 83,* 1072-1080.

Husbands, C. T. (1996). Variations in students' evaluations of teachers' lecturing and small group teaching: A study at the London School of Economics and Political Science. *Studies in Higher Education, 21*, 187-206.

Husbands, C. T. (1997). Variations in students' evaluations of teachers' lecturing in different courses on which they lecture: A study at the London School of Economics and Political Science. *Higher Education, 33*, 51-70.

Husbands, C. T. and Fosh, P. (1993). Students' evaluation of teaching in higher education: Experiences from four European countries and some implications of the practice. *Assessment & Evaluation in Higher Education, 18*, 95-114.

Kaiser, H. F. (1970). A second-generation little jiffy. *Psychometrika, 35*, 401-415.

Kaiser, H. F. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement, 34*, 111-117.

Kaiser, H. F. (1981). A revised measure of sampling adequacy for factor-analytic data matrices. *Educational and Psychological Measurement, 41*, 379-381.

Kember, D. and Gow, L. (1994). Orientations to teaching and their effect on the quality of student learning. *Journal of Higher Education, 65*, 58-74.

Ketchen, D. J. and Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal, 17*, 441-458.

Kolb, D.A. (1981). Learning styles and disciplinary differences. In A.W. Chickering and Associates (eds.), *The Modern American College*. San Francisco: Jossey-Bass.

Kolitch, E. and Dean, A. V. (1999). Student ratings of instruction in the USA: Hidden assumptions and missing conceptions about 'good' teaching. *Studies in Higher Education, 24*, 27-42.

Kugel, P. (1993). How professors develop as teachers. *Studies in Higher Education, 18*, 315-328.

Lorr, M. (1983). *Cluster Analysis for Social Scientists*. San Francisco: Jossey-Bass.

Maassen, P.A.M. (1997). Quality in European higher education: Recent trends and their historical roots. *European Journal of Education, 32*, 111-127.

Macleod, D. (2001). Trial by ordeal. *Guardian* (30 January 2001).

Marsh, H. W. (1981). The use of path analysis to estimate teacher and course effects in student ratings of instructional effectiveness. *Applied Psychological Measurement, 6*, 47-59.

Marsh, H. W. (1982). SEEQ: A reliable, valid and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology, 52*, 77-95.

Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/ course/ instructor characteristics. *Journal of Educational Psychology, 75*, 150-166.

Marsh, H. W. (1984a). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology, 76,* 707-754.

Marsh, H.W. (1984b). Experimental manipulations of university student motivation and their effects on examination performance. *British Journal of Educational Psychology, 54,* 206-213.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues and directions for future research. *International Journal of Educational Research, 11,* 253-388.

Marsh, H. W. (1991a). A multidimensional perspective on students' evaluations of teaching effectiveness: Reply to Abrami and d'Apollonia (1991). *Journal of Educational Psychology, 83,* 416-421.

Marsh, H. W. (1991b). Multidimensional students' evaluations of teaching effectiveness: A test of alternative higher-order structures. *Journal of Educational Psychology, 83,* 285-296.

Marsh, H. W. (1994). Weighting for the right criteria in the instructional development and effectiveness assessment (IDEA) system: Global and specific ratings of teaching effectiveness and their relation to course objectives. *Journal of Educational Psychology, 86,* 631-648.

Marsh, H. W. (1995). Still weighting for the right criteria to validate student evaluations of teaching in the IDEA system. *Journal of Educational Psychology, 87,* 666-679.

Marsh, H. W. and Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. In J. C. Smart (ed.) *Higher Education: Handbook of Theory and Research.* New York: Agathon Press.

Marsh, H. W. and Overall, J. U. (1980). Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. *Journal of Educational Psychology, 72,* 468-475.

Marsh, H. W., Overall, J. U. and Kesler, S. P. (1979). Class size, students' evaluations, and instructional effectiveness. *American Educational Research Journal, 16,* 57-70.

Marsh, H. W. and Roche, L. A. (1992). The use of student evaluations of university teaching in different settings: The applicability paradigm. *Australian Journal of Education, 36*, 278-300.

Marton, F. and Saljo, R. (1976). On qualitative differences in learning: 1. Outcome and process. *British Journal of Educational Psychology, 46*, 4-11.

Marton, F. and Saljo, R. (1984). Approaches to learning. In F. Marton, D. Hounsell and N.J. Entwistle (eds.), *The Experience of Learning.* Edinburgh, Scotish Academic Press.

McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe*, 384-397.

Meyer, J. H. F. and Parsons, P. (1989). Approaches to studying and course perceptions using the Lancaster inventory: a comparative study. *Studies in Higher Education, 14*, 137-153.

Milligan, G. W. and Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50,* 159-179.

Milligan, G. W. and Cooper, M. C. (1987). Methodology review: Clustering methods. *Applied Psychological Measurement, 11*, 329-354.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: An evaluation. *The Computer Journal, 20*, 359-363.

Mulaik, S. A. (1990). Blurring the distinction between component analysis and common factor analysis. *Multivariate Behavioral Research, 25*, 53-59.

Mulaik, S. A. and McDonald, R. P. (1978). The effect of additional variables on factor indeterminacy in models with a single common factor. *Psychometrika, 43*, 177-192.

Murphy, K. R., Balzer, W. K., Lockhart, M. C. and Eisenman, E. J. (1985). Effects of previous performance on evaluations of present performance. *Journal of Applied Psychology, 70*, 72-84.

Murray, H. G. (1984). The impact of formative and summative evaluation of teaching in North American universities. *Assessment and Evaluation in Higher Education, 9,* 117-132.

Myers, R. H. (1990). *Classical and Modern Regression with Applications.* Boston: PWS-Kent.

Neave, G. (1988). On the cultivation of quality, efficiency and enterprise: An overview of recent trends in higher education in Western Europe, 1986-1988. *European Journal of Education, 23*, 7-23.

Nixon, J. (1996). Professional identity and the restructuring of higher education. *Studies in Higher Education, 21*, 5-16.

Nulty, D. D. and Barrett, M. A. (1996). Transitions in students' learning styles. *Studies in Higher Education, 21*, 333-345.

Nunnally, J. L. (1978). *Psychometric Theory.* New York: McGraw-Hill.

Overall, J. U. and Marsh, H. W. (1979). Midterm feedback from students: Its relationship to instructional improvement and student's cognitive and affective outcomes. *Journal of Educational Psychology, 71*, 856-865.

Overall, J. U. and Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology, 72*, 321-325.

Plewis, I. (1997). Statistics in Education. London: Arnold.

Power, M. (1997). *The Audit Society: Rituals of verification.* Oxford: Oxford University Press.

Punj, G. and Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research, 20*, 134-148.

Quality Assurance Agency for Higher Education (2000). Handbook for academic review. Gloucester: QAA.

Ramsden, P. (1989). Perceptions of courses and approaches to studying: An encounter between paradigms. *Studies in Higher Education, 14*, 157-158.

Ramsden, P. (1992). *Learning to Teach in Higher Education.* London: Routledge.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics, 18*, 321-349.

Remmers, H. H. (1928). The relationship between students' marks and student attitudes towards instructors. *School and Society, 28*, 759-760.

Remmers, H. H. and Brandenburg, G. C. (1927). Experimental data on the Purdue Rating Scale for instructors. *Educational Administration and Supervision, 13*, 519-527.

Reynolds Report. (1986). Academic Standards in Universities. CVCP.

Richardson, J.T.E. (1993). Gender differences in responses to the approaches to studying inventory. *Studies in Higher Education, 18*, 3-13.

Robbins (1963) Higher Education (The Robbins Report). London: HMSO.

Rodin, M. and Rodin, B. (1972). Students evaluations of teachers. *Science, 177*, 1164-1166.

Rogers, C. R. (1993). The interpersonal relationship in the facilitation of learning. In M. Thorpe, R. Edwards and A. Hanson (eds.), *Culture and Processes of Adult Learning: A reader.* New York: Routledge.

Scott, P. (1995). *The Meanings of Mass Higher Education.* Buckingham: Open University Press.

Scott, S.V. (1999). The academic as service provider: Is the customer 'always right'? *Journal of Higher Education Policy and Management, 21*, 193-202.

Sekaran, U. (1992). *Research Methods for Business: A skill-building approach.* New York: John Wiley.

Shor, I. (1996). When Students Have Power: Negotiating authority in a critical pedagogy. Chicago: University of Chicago Press.

Silver, H. (ed.) (1992). *Student Feedback: Issues and Experience.* CNAA Project Report 39.

Smalzried, N. T. and Remmers, H. H. (1943). A factor analysis of the Purdue Rating Scale for instructors. *Journal of Educational Psychology, 34*, 363-367.

Sneath, P. H. A. (1969). Evaluation of clustering methods. In A. J. Cole (ed.), *Numerical Taxonomy* (pp. 257-271). New York: Academic Press.

Snook, S. C. and Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin, 106,* 148-154.

Stalnaker, J. M. and Remmers, H. H. (1928). Can students discriminate traits associated with success in teaching? *Journal of Applied Psychology, 12*, 602-610.

Stevens, J. (1996). *Applied Multivariate Statistics for the Social Sciences.* New Jersey: Lawrence Erlbaum.

Stewart, D. W. (1981). The application and misapplication of factor analysis in marketing research. *Journal of Marketing Research, 18*, 51-62.

Stringer, M. and Irwing, P. (1998). Students' evaluations of teaching effectiveness: A structural modelling approach. *British Journal of Educational Psychology, 68*, 409-426.

Thomas, E. (1993). The first distinguished teaching award in the United Kingdom. In R. Ellis (ed.), *Quality Assurance for University Teaching*. Buckingham: Open University Press.

Thorne, M. and Cuthbert, R. (1996). Autonomy, bureaucracy and competition: The ABC of control in higher education. In R. Cuthbert, (ed.), *Working in Higher Education*. Buckingham: Open University Press.

Timpson, W.W. and Andrew, D. (1997). Rethinking student evaluations and the improvement of teaching: Instruments for change at the University of Queensland. *Studies in Higher Education, 22*, 55-65.

Trow, M. (1993). Managerialism and the academic profession. Paper presented at the 'Quality Debate' Conference, Milton Keynes, 24 September.

Trowler, P. (1998). *Education Policy: A policy sociology approach*. East Sussex: Gildredge Press.

Van Rossum, E. J. and Taylor, I. P. (1987). The relationship between conception of learning and good teaching: A scheme for cognitive development. Paper presented at the American Educational Research Association Annual Meeting, Washington.

Velicer, W. F. and Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25*, 1-28.

Wall, D., Macaulay, C., Tait, H., Entwistle, D. and Entwistle, N. (1991). *The Transition from School to Higher Education in Scotland*. University of Edinburgh: Centre for Research on Learning and Instruction.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*, 236-244.

Watkins, D. and Akande, A. (1992). Student-evaluations of teaching effectiveness: A Nigerian investigation. *Higher Education, 24*, 453-463.

Watkins, D. and Gerong, A. (1992). Evaluating undergraduate college-teachers: A Filipino investigation. *Educational and Psychological Measurement, 52,* 727-734.

Watkins, D. and Regmi, M. (1992). Student evaluations of tertiary teaching: A Nepalese investigation. *Educational Psychology, 12,* 131-142.

Webb, C. (1994). Quality audit in the universities. In D. Green (ed.), *What is Quality in Higher Education.* Buckingham: Open University Press.

Weiler, K. (1988). *Women Teaching for Change: Gender class and power.* New York: Bergin and Garvey.

Willmott, H. (1995). Managing the academics: Commodification and control in the development of university education in the U.K. *Human Relations, 48*, 993-1027.

Appendix 4.1

The student feedback form

Loughborough University

Module Feedback [CODE 96BSB120]

umf-form 25.11.96

Programme Code

N	O	T	
	[0]	[0]	[0]
[1]	[1]	[1]	[1]
[2]	[2]	[2]	[2]
[3]	[3]	[3]	[3]
[4]	[4]	[4]	[4]
[5]	[5]	[5]	[5]
[6]	[6]	[6]	[6]
	[7]	[7]	[7]
	[8]	[8]	[8]
	[9]	[9]	[9]

NEEDED

lease spend a few minutes completing this questionnaire. The results will be used to improve the uality of your module. Normally the person handing out this form will tell you your programme code. lease write it in the boxes opposite and then mark the appropriate numbers below.

odule Title
Not Required *(as given below)*

Your Name (Optional)

rogramme Title
Decision Support Systems

or each of the statements below please mark like this ▬ with black or blue biro or pencil, the box hich best indicates your view.

		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	I was made aware of the module's aim and objectives	☐	☐	☐	☐	☐
2	The teaching methods for this module helped me understand the subject	☐	☐	☐	☐	☐
3	The academic content of the module was stimulating	☐	☐	☐	☐	☐
4	I understood the assessment requirements of the module	☐	☐	☐	☐	☐
5	The teaching rooms for this module were fit for their purpose	☐	☐	☐	☐	☐
6	I found the library support for the module to be satisfactory	☐	☐	☐	☐	☐
7	The subject material in this module was CHALLENGING and STIMULATING.	☐	☐	☐	☐	☐
8	The subject material in this module was presented at about the RIGHT PACE.	☐	☐	☐	☐	☐
9	The subject material in this module was RELATED to BUSINESS situations.	☐	☐	☐	☐	☐
10	The teaching on the module was WELL ORGANISED.	☐	☐	☐	☐	☐
11	The COURSEWORK supported the module objectives.	☐	☐	☐	☐	☐
12	The TUTORIALS linked with the module were useful.	☐	☐	☐	☐	☐

	Lecturer A	**Lecturer B**	**Lecturer C**
	Malcolm King	Dave Coates	Neil Doherty

The lecturer...	Lecturer A: Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Lecturer B: Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Lecturer C: Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
13 was well prepared,	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
14 was enthusiastic about the subject,	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
15 explained the subject clearly,	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
16 encouraged participation & questions.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
17 Overall, was effective.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
18 Could be contacted for advice if needed	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

PLEASE TURN OVER

PLEASE TURN OVER...

Appendix 7.1

Results of the cluster analysis on the 64 modules

Appendix 7.1 Results of the cluster analysis on the 64 modules

Code	Reg.Studs	#cases	Any Clusters	Key Variables	No Difference	Description	Cluster size
A052~96	147	119	TWO	2 3 15 17	4 6 11	Lecturer skill	45/72
A530~96	91	74	Continuum			Availability (Q18) showed greatest difference	
A565~96	91	38	Continuum			Availability (Q18) and Encouraged Participation (Q16) greatest difference	
B070~96	188	133	TWO	2 3 8 15 17		Lecturer skill	110/23
B510~96	75	48	Continuum			Lecturer organisation and preparation seem important	
B555~96	54	44	Continuum			Pace of teaching (Q8) showed grearest difference	
C575~96	46	39	Continuum			Availability (Q18) showed greatest difference	
A004~97	**108**	**145**	**Contiuuum**				
A006~97	284	252	Continuum			Q3-Stimulating showed greatest difference	
A015~97	**100**	**78**	**TWO**	**L1 13-18**	**7 11 L2 13&18**	**A quarter of respondents did not like lecturer 2**	**60/18**
A017~97	**68**	**53**					
A020~97	88	72	Unclear			Dendrogram splits into four "clusters"	
A022~97	96	66					
A030~97	100	69	THREE	2 3 13 15 17		Lecturer skill	10/46/13
A035~97	61	50	TWO	2 3 17 18		Mild lecturer effect	12+38
A052~97	194	138	Continuum				
A057~97	**201**	**113**	**THREE**	**2 3 16&17 L1**		**Mild lecturer skill effect**	**54/27/32**
A060~97	207	116	TWO	2 8 15 16 17	5	Lecturer skill	66/50
A070~97	197	98					
A080~97	**128**	**73**	**Continuum**				
A085~97	**159**	**106**	**TWO**	**L3 13-18**		**Cluster 1 did not like lecturer 3**	**43/63**
A100~97	122	62	TWO	3 7 8		Mild lecturer effect	16/33/13
A105~97	184	87	Continuum				
A110~97	**77**	**45**	**TWO**	**2 3 8 9 12 17 L2**	**15 L3**	**Lecturer skill particularly lecturer 2 (Q17)**	**16/29**
A115~97	**101**	**78**	**Continuum**				
A506~97	190	136					
A525~97	**162**	**105**	**Unclear**			**One fifth of respondents scored lecturer 2 very low (PB 14-17)**	

Code	Reg.Studs	#cases	Any Clusters	Key Variables	No Difference	Description	Cluster size
A526~97	174	104	Continuum			Approx: 10% did not like the lecturer	
A530~97	85	54	TWO	10 13	9 18	Organisation and preparation	32/21
A565~97	149	80					
B005~97	82	61	TWO	2 8 15 17	4 6 9 18	Lecturer skill	29/32
B007~97	82	63	None				
B012~97	79	55	TWO	2 3 7 8 9 10 15-17 L1		Lecturer skill	17/35
B020~97	65	52	None				
B022~97	164	105	Continuum				
B023~97	76	43	Continuum			Q17 Lecturer overall effective showed greatest difference	
B028~97	65	56	THREE	2 3 5 8 10 15 17		Lecturer skill	14/23/19
B040~97	165	109	TWO	2 3 7 10 15&17L1	13, 16&18 L3	Lecturers skill particularly lecturer 1	55/54
B070~97	183	107					
B086~97	103	73	Continuum				
B087~97	126	83	Continuum			Organisation Q10 showed greatest difference	
B100~97	160	103					
B105~97	158	75	TWO	2 3 7 10 12-17		Lecturer skill	42/33
B110~97	85	69	None				
B120~97	85	67	TWO	2 3 8-11 17L1 17L2	13-18 L3	Teaching effect (lecturer 3 constant in both clusters)	45/22
B135~97	104	79	Continuum	2 7 10 15	6	Mild lecturer skill effect	26/53
B510~97	95	73	THREE	10 13 15		Organisation and clarity	
B530~97	94	65	TWO	3 8 13-18 L2		Teaching "dynanism" particularly lecturer 2	
B545~97	98	72	Continuum				
B550~97	68	48	Continuum				
B555~97	59	43	TWO	2 3 17 18		Lecturer skill	23/20
B560~97	78	53	TWO	2 3 15 16 17		Lecturer skill	11+42
C005~97	75	51	Unclear			Dendrogram splits into four "clusters"	
C022~97	66	55	THREE	5 6	10 13 14 16 17	Teaching rooms (Q5) and Library (Q6) show the greatest difference	22/13/20
C050~97	159	99	TWO	9 10 L2 13,15&17 L1 15-17		(Mildish) lecturer effect	33/66

Code	Reg.Studs	#cases	Any Clusters	Key Variables	No Difference	Description	Cluster size
C055~97	85	43	TWO	2 3 8 10 13-17 L1		Lecturer skill	28/15
C072~97	59	52	THREE			C2 lecturer 1 scores drop(cf.C1) C3 lecturer 2 down, lecturer 1 up	20/20/12
C080~97	77	51	Continuum			Q5-Rooms showed greatest difference	
C110~97	110	78	THREE	2 3 7 8 16 17		Lecturer skill	26/13/39
C132~97	78	70	THREE			C2 lecturer 2 scores v.low, lecturer 1 scores high C1 lecturer 2 scores significantly higher, lecturer 2 scores fall	17/22/31
C160~97	79	66	Continuum				
C190~97	73	63	TWO	2 3 7 8 10 13-17		Lecturer skill	48/15
C506~97	63	48	TWO	2 3 10 L1 15-17 L2 15-17	6	Lecturer skill	26/22
C540~97	86	55	Continuum			Questions 3&7 show greatest difference	
C565~97	67	52	TWO	2 3 13 15 17	4 18	Mild lecturer effect	24/28

Appendix 7.2

Results of the Calinski and Harabasz stopping rule

Appendix 7.2 Results of the Calinski and Harabasz stopping rule

Code	K = 6	K = 5	K = 4	K = 3	K = 2	%dif.6-5	%dif.5-4	%dif.4-3	%dif.3-2
A052~96	18.56	20.27	22.13	25.27	26.33	9.21	9.18	14.19	4.19
A530~96	10.75	11.86	13.6	16.96	18.5	10.33	14.67	24.71	9.08
A565~96	8.23	8.31	8.52	9.21	11.31	0.97	2.53	8.10	22.80
B070~96	17.47	18.78	21.03	24.84	33.48	7.50	11.98	18.12	34.78
B510~96	10.37	10.94	11.86	14.08	16.61	5.50	8.41	18.72	17.97
B555~96	8.88	9.81	11.2	13.78	18.76	10.47	14.17	23.04	36.14
C575~96	9.84	10.79	12.22	13.9	17.15	9.65	13.25	13.75	23.38
A004~97	**12.56**	**13.91**	**16.2**	**20.63**	**29.93**	**10.75**	**16.46**	**27.35**	**45.08**
A006~97	36.64	41.51	46.87	55.76	72.62	13.29	12.91	18.97	30.24
A015~97	**7.78**	**8.4**	**9.21**	**10.36**	**13.95**	**7.97**	**9.64**	**12.49**	**34.65**
A017~97	**6.31**	**6.83**	**7.61**	**8.75**	**11.04**	**8.24**	**11.42**	**14.98**	**26.17**
A020~97	6.26	6.82	7.29	7.6	8.08	8.95	6.89	4.25	6.32
A022~97	9.82	10.47	11.7	13.22	17.62	6.62	11.75	12.99	33.28
A030~97	17.45	19.44	21.01	24.15	26.77	11.40	8.08	14.95	10.85
A035~97	7.83	8.39	8.48	9.18	10.75	7.15	1.07	8.25	17.10
A052~97	16.34	18.12	20.79	26.47	34.62	10.89	14.74	27.32	30.79
A057~97	**14.97**	**16.32**	**18.29**	**22.6**	**27.49**	**9.02**	**12.07**	**23.56**	**21.64**
A060~97	15.44	16.97	18.54	21.92	32.89	9.91	9.25	18.23	50.05
A070~97	15.09	16.92	18.93	22.74	26.99	12.13	11.88	20.13	18.69
A080~97	**8.66**	**9.5**	**10.7**	**12.82**	**15.79**	**9.70**	**12.63**	**19.81**	**23.17**
A085~97	**9.41**	**10.56**	**12.34**	**13.96**	**19.68**	**12.22**	**16.86**	**13.13**	**40.97**
A100~97	10.49	11.62	13.04	15.19	16.29	10.77	12.22	16.49	7.24
A105~97	13.62	15.33	16.61	18.17	20.33	12.56	8.35	9.39	11.89
A110~97	**6.88**	**7.42**	**8.23**	**9.23**	**11.75**	**7.85**	**10.92**	**12.15**	**27.30**
A115~97	**11.47**	**12.72**	**14.82**	**18.64**	**24.63**	**10.90**	**16.51**	**25.78**	**32.14**
A506~97	13.38	14.71	17.09	19.28	23.7	9.94	16.18	12.81	22.93
A525~97	**12.31**	**13.67**	**15.41**	**18.66**	**25.13**	**11.05**	**12.73**	**21.09**	**34.67**
A526~97	17.84	19.25	21.39	25.15	30.5	7.90	11.12	17.58	21.27

Code	K = 6	K = 5	K = 4	K = 3	K = 2	%dif.6-5	%dif.5-4	%dif.4-3	%dif.3-2
A530~97	8.64	9.1	9.32	9.95	10.46	5.32	2.42	6.76	5.13
A565~97	8.21	8.81	9.33	10.39	13.32	7.31	5.90	11.36	28.20
B005~97	9.16	10.13	11.02	12.76	18.01	10.59	8.79	15.79	41.14
B007~97	**6.81**	**7.37**	**8.4**	**9.42**	**10.18**	**8.22**	**13.98**	**12.14**	**8.07**
B012~97	**12.17**	**13.48**	**15.69**	**20.47**	**30.08**	**10.76**	**16.39**	**30.47**	**46.95**
B020~97	9.34	9.83	10.5	11.04	10.84	5.25	6.82	5.14	-1.81
B022~97	12.65	13.91	15.07	17.16	18.48	9.96	8.34	13.87	7.69
B023~97	5.64	6.06	6.48	7.29	9.76	7.45	6.93	12.50	33.88
B028~97	13.88	15.47	17.9	22.64	25.99	11.46	15.71	26.48	14.80
B040~97	**12.26**	**13.07**	**14.28**	**16.83**	**20.35**	**6.61**	**9.26**	**17.86**	**20.92**
B070~97	10.04	11.17	12.75	14.43	16.88	11.25	14.15	13.18	16.98
B086~97	**7.34**	**7.95**	**8.72**	**9.75**	**13.58**	**8.31**	**9.69**	**11.81**	**39.28**
B087~97	**12.99**	**14.22**	**16.45**	**19.69**	**23.64**	**9.47**	**15.68**	**19.70**	**20.06**
B100~97	**10.39**	**10.99**	**11.67**	**13.34**	**15.55**	**5.77**	**6.19**	**14.31**	**16.57**
B105~97	10.78	12.27	14.34	18.18	26.8	13.82	16.87	26.78	47.41
B110~97	**6.12**	**6.54**	**7.32**	**7.86**	**8.89**	**6.86**	**11.93**	**7.38**	**13.10**
B120~97	**7.1**	**7.83**	**8.76**	**10.16**	**14.88**	**10.28**	**11.88**	**15.98**	**46.46**
B135~97	10.27	11.42	12.7	15.35	22.2	11.20	11.21	20.87	44.63
B510~97	11.6	12.99	15.43	19.06	21.35	11.98	18.78	23.53	12.01
B530~97	**12.25**	**11.89**	**13.35**	**15.05**	**18.5**	**-2.94**	**12.28**	**12.73**	**22.92**
B545~97	10.17	11.22	13.09	15.64	20.47	10.32	16.67	19.48	30.88
B550~97	10.88	11.9	12.97	13.47	17.79	9.38	8.99	3.86	32.07
B555~97	10	10.87	12.16	14.89	22.16	8.70	11.87	22.45	48.82
B560~97	11.75	12.34	13.78	14.13	17.29	5.02	11.67	2.54	22.36
C005~97	**5.93**	**6.16**	**6.49**	**7.43**	**8.12**	**3.88**	**5.36**	**14.48**	**9.29**
C022~97	10.67	11.18	11.57	13.03	14.08	4.78	3.49	12.62	8.06
C050~97	**12.99**	**14.59**	**16.65**	**18.47**	**24.92**	**12.32**	**14.12**	**10.93**	**34.92**
C055~97	**8.05**	**9.1**	**10.76**	**13.81**	**19.95**	**13.04**	**18.24**	**28.35**	**44.46**
C072~97	**8.95**	**10.01**	**11.89**	**14.98**	**18.34**	**11.84**	**18.78**	**25.99**	**22.43**
C080~97	7.42	7.97	8.79	10.28	12.29	7.41	10.29	16.95	19.55
C110~97	11.25	12.59	14.89	17.98	22.53	11.91	18.27	20.75	25.31

Code	K = 6	K = 5	K = 4	K = 3	K = 2	%dif.6-5	%dif.5-4	%dif.4-3	%dif.3-2
C132~97	**8.84**	**9.49**	**10.49**	**12.31**	**13.65**	**7.35**	**10.54**	**17.35**	**10.89**
C160~97	15.18	17.2	20.11	24.57	30.43	13.31	16.92	22.18	23.85
C190~97	15.4	17.27	19.99	25.98	37.33	12.14	15.75	29.96	43.69
C506~97	**8.37**	**9.28**	**10.91**	**12.93**	**18.9**	**10.87**	**17.56**	**18.52**	**46.17**
C540~97	7.2	8.05	8.38	9.43	11.6	11.81	4.10	12.53	23.01
C565~97	11.93	12.4	13.07	14.37	18.28	3.94	5.40	9.95	27.21