OVERSEAS STUDENTS EVALUATION OF TEACHING EFFECTIVENESS

The construction Of Birmingham Overseas Students Teaching Evaluation Questionnaire (BOSTEQ)

KEVIN MARK BREWER

Master of Science (Business Research Methods)

THE UNIVERSITY OF ASTON IN BIRMINGHAM

December 1993

THE UNIVERSITY OF ASTON IN BIRMINGHAM

OVERSEAS STUDENTS EVALUATION OF TEACHING EFFECTIVENESS: The Construction of Birmingham Overseas Students Teaching Evaluation Questionnaire (BOSTEQ).

KEVIN MARK BREWER

MASTER OF SCIENCE (BUSINESS RESEARCH METHODS)
December 1993

SUMMARY

At a time when higher education institutions are becoming more concerned with quality, the evaluation of teaching effectiveness by students is important. One particular group is often ignored, that is overseas students.

In the USA, student evaluation of teaching effectiveness is common, not so in the UK. Furthermore no reported studies have looked at overseas students only in this area.

Overseas students from many countries at the University of Birmingham were sent a questionnaire asking them to rank the most important characteristics for their "ideal lecturer". The results were analysed by nationality grouping particularly. A small number of significant differences ($p<0.001$) were found. The overall rankings were similar to past studies. The most popular characteristics chosen were "has good knowledge of own subject" and "is well prepared for classes".

The results were used to construct a specific questionnaire for overseas students to use to rate their lecturers, called the Birmingham Overseas Students Teaching Evaluation Questionnaire (BOSTEQ).

The issues around the use of teaching evaluation questionnaires is discussed, particularly their reliability, validity, and whether they are influenced by factors other than teaching.

Key words: overseas students; student evaluation of teacher effectiveness; student rating of instruction

ACKNOWLEDGEMENTS

i) To all the students who responded to the questionnaire, and particularly those involved in the pilot study. Thanks to Lourdes Santos who spent time on the first questionnaire, and to her flatmates.

ii) Among the staff in the English For Overseas Students Unit, discussions and advice from Philip King, Richard Cauldwell, Tony Dudley-Evans, Tim Johns, Martin Hewings, and especially Jonathan Rees. Thanks for access to the list of departments of the overseas students.

iii) Help and advice from my supervisor, Guy Cumberbatch, and others at Aston Business School. Also at the University of Aston, unpublished reports, including Sheila Kirby's were interesting. Elsewhere, the opportunity to talk to Joanna Channell at the beginning was useful, and also the unpublished information she gave me.

iv) To anyone else who I may have forgotten, thanks for your comments. Let us hope higher education will listen to the comments of the consumer (overseas students).

KMB Dec 1993.

CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER 1 INTRODUCTION

In the new world for higher education in the United Kingdom since the 1980s, quality has become a key term.

A key term it may be, but then the problem arises of defining what is quality, and how to measure good/bad quality education. Harvey and Green(1993) offers five possible definitions:

i) Quality as exceptional - "Traditionally, the concept of quality has been associated with the notion of distinctiveness, or something special or 'high class'"(p11). However, this notion is difficult to measure in the context of the empirical tradition, which are mainly quantitative.

ii) Quality as perfection or consistency - This involves conformity to specifications. But again there exists the problem of how to measure or set the specifications.

iii) Quality as fitness for purpose - The extent to which a service fits its purpose. The purpose can either be the wishes of the "customer" or the mission of the institution. In the first case, two further questions arise - who is the "customer": the student or the funder? and are the "customers"(students) able to specify what they want? "Defining quality in higher education as meeting customers' needs does not necessarily imply that the customer is always best placed to determine what quality is or whether it is present"(p19).

iv) Quality as value for money - The problem of how to define "value for money". For example, the use of performance indicators (like staff-student ratios) suggest efficiency, but for the accountant rather than the student.

v) Quality as transformation - This is nearer a good definition of quality in higher education. "Education is not a service for a customer but an ongoing process of transformation of the participant, be it student or researcher"(p24). Involved in this process is empowerment - giving the participant power to influence their own transformation. One aspect of this is student evaluation of the institution, and of teaching.

Student ratings of instruction(SRI) or student evaluation of teaching effectiveness(SETE), as it is known, has a long history in the USA, but there

seems to have been little interest in the United Kingdom. Chapter 4 gives the background to the use of student evaluation. What characteristics the students are looking for in their "ideal lecturer". Does it vary from department to department? If students have faith in their evaluations of teaching, do these evaluations agree with what the lecturers think of teaching?

One argument against the use of the student rating of instruction is that they are not trustworthy. This can mean a number of things to the critics. It may be that the students' judgments are based on something other than an evaluation of just the teaching. It may be that the students do not know the difference between good or bad teaching, or that they are prone to change their minds/evaluations about their teachers. Chapter 5 addresses these issues, and concludes that students' rating of instruction can be trusted to a point.

Even if the critics are unhappy about the use of rating questionnaire, then there are other less studied methods of evaluating teaching (chapter 6).

A concern about student satisfaction with the whole institution is starting to grow in the United Kingdom. Chapter 2 looks at this interest, and also at what the theoretical literature says about student satisfaction.

Overseas students are growing as a percentage of the student body. The University of Birmingham has over 1600 overseas students, making up around one-sixth of the student population. Since the introduction of "full cost fees" in the early 1980s, overseas students pay between two and three times the fees

of home students for the same course. Their expectations and satisfaction with the UK higher education institution has to be linked to their tuition. Traditionally the expectations of UK higher education are of high quality. However, this is not fixed in stone. It may be time to listen to the consumer.

Unfortunately any developing interest in student satisfaction has stopped before reaching the question of overseas student satisfaction. Various studies (details in chapter 3) have questioned overseas students about their feelings on the institutions they attend, and looked at the sort of problems they face. But studies have not asked them specifically about their lecturers. What do overseas students expect their "ideal lecturer" to be like?

This research is attempting to ask this question, and the results of the survey show that there are some differences based on nationality (chapter 7). From the results of the survey, the characteristics chosen for the "ideal lecturer" were used to construct the Birmingham Overseas Student Teaching Evaluation Questionnaire (BOSTEQ) (chapter 8). This is a specific rating instrument of lecturers for overseas students.

As mentioned earlier, Harvey and Green(1993) talk about empowerment as an important aspect of quality. It is hoped that BOSTEQ can be used in this context, to benefit both the students (they can feel that the institutions really want to listen to them), and the faculty (they can discover whether they are giving the "consumers" what they really want).

The use of student ratings of instruction throws up many issues and problems. A number of these are addressed with reference to the literature.

CHAPTER 2 STUDENT SATISFACTION

2.1 Introduction

A number of studies have been conducted on students' general level of satisfaction about their institutions. A well-known annual report comes from the Student Satisfaction Research Unit at the University of Central England in Birmingham. The fourth report surveys all aspects of university life, concluding that "students' perceptions of their educational experiences are much more than their reactions to their courses of study" (Mazelan et al 1992 p33). In other words, the experience of university life, certainly in the United Kingdom, is more than just the course and classes.

This finding links with the Roberts and Higgins(1992) report of an interesting contradiction of students partially dissatisfied with their courses, who would still recommend their institutions. The authors suggest a number of reasons for this:

i)"Having a good time" socially compensates for academic deficiences.

ii)Most students have no experience of other institutions and thus cannot know if they are different/better than their own.

iii)Psychological problem of admitting that the students had made the wrong choice.

iv)The "football supporter syndrome":complaining in the company of fellow supporters,but keeping a positive face to outsiders.

(Roberts and Higgins 1992 pp 65-6)

2.2 Application of Consumer Satisfaction Research to Education

In the USA, interest in student satisfaction has appeared in the marketing literature, and particularly in the area of consumer satisfaction. The principle behind the interest is that " a generally accepted principle is that students are indeed consumers and, therefore, the marketing concept should be practiced in academic institutions" (Barry et al 1982 p109). Thus, the aim of these types of studies are to find out what the consumer (student) wants and what are their expectations about the institution?

Overall, it is accepted from the main studies (Barry et al 1982; Hawes and Glisan 1983; and Hampton 1983) that students' satisfaction/dissatisfaction is linked to their expectations. But these expectations also change over time as experience is gained through consumption. Ortinau,Anderson and Klippel(1987) believe perceptions of college attributes will be influenced by involvement in the course. Students were given questionnaires on the first day of the course, the middle of the semester, and at the end. Their findings were that students do "significantly modify their overall evaluation feelings toward the importance of their expectancy beliefs in a general downward direction as they gain more course experience"(p268). Thus expectations are not static, and their importance towards being fulfilled is changing. Ortinau and Anderson(1986) have developed this idea into a 4 stage model of post-purchase satisfaction. From "new student ownership", "early consumption", "subsequent consumption" to "disposition"(p121). Each stage having its own expectations.

Palihawadana(1992), in a rare British article, asks whether what the author calls the "expectancy-performance paradigm" is a good measure of quality assurance in education. Using students from three Masters courses at a British business school, he tests for expectations and satisfaction. They were given a questionnaire at the beginning of the course, after 6 months, and at the end of the course. The results suggest a curvilinear relationship: "expectation levels were very high at the entering stage, but the outcome level falls significantly below the expectation levels by the middle of the programme and rise slightly again at the third stage"(p70). Accepting slight limitations, the findings are similar to those of Ortinau,Anderson and Klippel(1987), and Palihawadana is confident that the "expectancy-performance paradigm" is useful in education.

2.3 Theoretical Explanations of Student Satisfaction

A small number of studies have tried to explain student satisfaction from a theoretical point of view. An interesting idea comes in the form of the "expectancy violation model"(Burgoon and Hale 1988). Used mainly in communication research, it says that people have expectations about others' behaviour, and "violations of these expectations will stimulate change in arousal thus affecting our communicative interaction with others"(Koermer and Petelle 1991 p342). Koermer and Petelle applying this model to the student rating of instruction, found some support for the model. Students with

high expectations of the teacher, and receiving "high experience"(ie expectations met or surpassed) rated teachers more favourably than low expectations/high experience or high expectations/low experience. This emphasises the importance of met or unmet expectations for student satisfaction.

Another possibility is to see satisfaction as a "fit" between two factors. Morstain(1977) considers satisfaction as a "fit" between faculty orientations and student educational orientations. Satisfied students were found to have a similar orientation to their faculty, and the opposite for dissatified students.

Along similar lines, the work of Holland(1973) has been applied in a small number of studies. Holland, concerned mainly with vocational choices, lists six personality types and six environmental models. Satisfaction or congruency is when the similar type of personality or environment are together. Morrow jr(1971) concentrated on mathematics (classified as "intellectual") and sociology (classified as "social") students. He found that there was a significant relationship between personality type and satisfaction for mathematics students, but not sociology.

Nafziger,Holland and Gottfredson(1975) found that satisfied students resembled the typical student at their college. This supports the hypothesis that the congruency between their personality type, and the typical personality type(or environmental type ie all the students around) is important for

satisfaction.

In a slightly different vein, research has looked at the cognitive styles of the students and their teachers. Reninger and Snyder(1983) found that "field independent" students reported more satisfaction and greater teaching effectiveness with "field independent" teachers. The same was true for "field dependent" students with "field dependent" teachers. But dissatisfaction and less teacher effectiveness was reported if there was a difference in cognitive style between student and teacher. Saracho(1992) examined this further, and suggests that it is the type of teaching that is preferred. "Field dependent" teachers and students tend to have more interpersonal orientation, and "field independent" teachers/students are more content orientated.

CHAPTER 3 OVERSEAS STUDENTS

3.1 Introduction

Firstly, it is worth defining "overseas students". A good general definition comes from Morris(1967): " a student who is normally domiciled outside Britain; who is not a British citizen temporarily living abroad, owing, for example, to parents' occupation; and who has come to Britain in order to complete some course of higher education"(p87). This seems acceptable to use here.

Overseas students have been coming to study in the U.K for hundreds of years. Elsey and Kinnell(1990) report a statement by Henry III in 1231 warning people to behave better towards overseas students. Interest developed early in this century with the Lee-Warner Committee(1922), but it is not until the 1960's that detailed studies are being made about overseas students.

Since then, there has been no shortage of studies of overseas students - their expectations of life in the U.K, and the reality of what they found. Some of the best known studies, but slightly dated now, include Burns(1965), Dunlop(1966), Livingstone(1960), Planning(1954), Political and Economic Planning(1965) report on "New Commomnwealth" students, Sen(1970), and Singh(1963) on Indian students. One study that remains unaffected by the passage of time is the Overseas Student Trust study called "Freedom to Study"(Reed et al 1978). They concentrate on groups of students from

Malaysia, Nigeria and Iraq, using both group discussion, and then some individual interviews. Further studies are referenced in Altbach and Wang(1989),who provide an annotated bibliography.

Of the recent studies, Williams et al(1986) found that 90% of overseas students expressed satisfaction with their courses, and 30% were very satisfied. Postgraduate research overseas students replied "very satisfied" most of any group. Around 75% of all overseas students questioned felt they were getting good value-for-money, while 5% saw their course as poor value. The main reasons for finding the course unsatisfactory were "quality of teaching"; "content of course(too easy/shallow)"; "teaching methods"; and "inadequate supervision"(p46).

The issue of the perception of "value-for-money" is an important one. In a study of female overseas students, it was found that the high cost of tuition created high expectations. One student quoted, said, "for example my fees for last term were £900. I had only 3 lectures and 2 tutorials. I do not think this is just" (Goldsmith and Shawcross 1985 p25). In the same study, it was found that the most satisfactory aspect about their course was the "quality of course and tuition", and the least satisfactory aspect was the "workload too heavy".

The largest recent study of overseas students conducted at the universities of Nottingham and Loughborough did not necessarily find anything dramatically new, but merely highlighted that institutions are still a long way from fully

meeting the needs of overseas students. The authors summarise the main comments by overseas students:

- Pre-arrival communications to students in their home countries were seen as a vital pre-requisite to a proper understanding of the U.K.

- On arrival, suitable permanent accommodation should be available, taking into account more carefully vacation accommodation needs and special dietary requirements.

- Appropriate information/help on arrival to create a favourable first impression, and to avoid life becoming problem-centred.

- Early contacts with academic tutors is crucial.

- The concern of overseas students because of their inadequate English and/or study skills.

- The students generally happy with the quality of their courses, but unsure about the informal teaching methods.

- Having the opportunity to meet and befriend British people seen as important.

(Kinnell 1990 p8-9).

3.2 Overseas Students as Strangers.

Schuetz(1943) in his classic essay describes the social psychological processes that a stranger experiences. Particularly here, a stranger is one who comes from another country or culture. What had made sense to them at home, now makes no sense, and they do not understand the assumed aspects of the new culture.

To the "in-group"(the home culture), the system of knowledge makes sense, "offering ready-made directions for use, to replace truth hard to attain by comfortable truisms, and to substitute the self-explanatory for the questionable"(p501). Schuetz calls this "thinking as usual"(p501), which has certain basic assumptions. For instance, "social life will continue to be same as it has been so far"(p502). The stranger does not share these assumptions, and comes to question what is seen as the unquestionable. The stranger then starts to interpret the new environment through their own "thinking as usual". Not only is the stranger threatened by the new culture, but is also seen as a threat to the new culture by its inhabitants. Thus a period of adaptation occurs.

Various studies have looked at adaptation and its problems (see Altbach and Wang 1989). Livingstone(1960) in an earlier study of overseas students on social welfare training courses, subdivides their experience into 4 main phases,(based on the work of Foundation for Research on Human Behaviour 1956):

i) The "Spectator phase" - typified by curiosity about the new culture.

ii) The "Involvement phase" - with the development of the first negative feelings.

iii) The "Coming-to-terms phase" - the visitor starts making sense of new culture.

iv) The "Pre-departure phase" - a period of anxiety about returning home.

(Livingstone 1960; p3)

Banjo(1975), a Nigerian student, in an article entitled "Cultural Castration",

describes the culture shock faced by African students. "Most overseas students, confronted with the difficulties of adjustment to British culture, react by withdrawing into a protective shell"(p17).

Lee(1984) shows how the self-concept of students from developing countries changes to a "compromised rating" while staying abroad,ie lower than they believe their friends at home would rate them, and higher than they believe home(USA) students would rate them.

3.3 Specific Ethnic/Nationality Group Studies

Some studies have concentrated on particular ethnic or national groups. This is known as "cultural profiling", and certainly at one time this was popular in language teaching. Then it became frowned upon as smacking of racial stereotyping.

A number of the studies have been carried out by nationals themselves - Animashawun(1963) on African students; Singh(1963), Bowers(1978), and Regmi(1987) on students from the Indian sub-continent; and Hawkey and Nakornchai(1978) on Thai students.

Grouping students from South East Asia together, a number of studies have looked at aspects of the culture, or individual cultures. Noesjirwan(1970) argues that Asian students at an Australian university have different attitudes to learning because of their home background. Traditionally seen as having "an

uncritical acceptance of authority, and this includes both the teacher and the textbook"(p393). The author found that Asian students scored higher on factors relating to more dependence on authority, less independence of thought, and more emphasis on memory(p394).

However, Kember and Gow(1991) propose an alternative to the stereotype, finding that Hong Kong students scored similar to Australian students on the Biggs' Study Process Questionnaire. The authors suggest this seriouly questions the image of Hong Kong students as dependent on rote learning, due to an innate tendency. More likely, if it is true, to be a function of teaching practices in their country(p126).

Are "cultural profiles" descriptions of specific ethnic/national groups useful? Bowers(1978) believes that some generalization is possible. Hawkey and Nakornchai(1978) suggest that whatever the staff feel, overseas students may be in favour of them. "Anything that helps in the development of more sensitive, helpful learning programmes must be welcomed"(p78).

On the other hand, "cultural profiling" can easily become cultural stereotyping. Channel(1990) reports a comment by a Senior Technician in Engineering:

> "Indians and Pakistanis not generally engineering minded - Kenyans particularly good engineers - Nigerians seem to be hardest to deal with - most African students somewhat unreliable in terms of planning work,time keeping - Filipinos, Chinese and Malays tend to have a more responsible attitude"(p75).

Obviously, staff need help in identifying cultural traits, but is a thin line before cultural stereotyping comes and influences the relationship with the individual overseas student.

3.4 Overseas Students and Academic Life

3.4.1 Problems

Watt(1978) presents a summary of overseas students on a Systems Analysis postgraduate course at the University of Aston in Birmingham, between 1973 and 1978. There were 50 overseas students: only 13 passed the course without some form of repeat examination; 16 passed after sitting one referred paper; 7 passed after a full resit examination; 11 withdrew because of examination failure; 3 yet to take full resit examination(p39). The results appear damning.

Overseas students have problems with studying in English in many cases. This is an obvious statement. Kinnell(1990) included this fact in the summary of main findings of her research.

Holes(1972) concentrating on students in Transportation and Environmental Planning at the University of Birmingham, found students consistently commenting on study skills problems, as well as language-related problems, which were most important. For lectures: "lack of clarity/audibility";

"vocabulary words hindering comprehension"; and "difficulty in taking notes at the same time as listening" were most reported. Difficulties in tutorials reported most commonly were, "expressing my own ideas in English" and "understanding others discussion".

But are there study skills problems other than those which are language-related? Wright(1985) draws the distinction between "Unitary Learning Source" (characterised by heavy dependence on the lecturer and a limited number of textbooks used), and "Multiple Learning Sources" (characterised by different points of view from various sources)(p279). The former is usually associated with undergraduates, and the latter with postgraduates in U.K institutions. "Multiple Learning Sources" lead to the development of higher intellectual skills. Before the development to "Multiple Learning Sources", the student must have self-confidence and personal maturity. "The overseas student may be much more mature personally in these terms but the cultural change removes him from the environment which reaffirms his personal maturity and sense of identity"(Wright 1985;p285). This leads the students to rely too much on their supervisor, even for the whole period of a dissertation.

Bilbow(1989) prefers to place the blame for problems with study skills, on language problems. Concentrating on lecturers, he argues that language difficulties limit the ability to follow the lecturer, and consequently to apply a "deep" learning strategy to the material. Simply, language difficulties with English force a "surface" learning approach, which is not appropriate for U.K

higher education institutions(1).

3.4.2 Different Methods of Teaching/Study

Overseas students are faced with a varied selection of teaching methods in the U.K, how do they find them?

Sen(1970) compared lectures, practical work, tutorials, seminars, and teaching practice between 5 "continent" groups. Most students were accustomed to the lecture method. More Africans and West Indians considered practical work, tutorials, and teaching practice very useful, while Asian and Middle East students considered them of moderate usefulness. More Middle East students considered tutorials less useful, and placed little value on seminars. More Africans emphasised the lecture method, while Asians did not consider it so useful.

Elsey(1990) summarises the survey of 158 final year overseas students at Loughborough and Nottingham universities, by firstly pointing out that the "geographical areas producing most fee payers have little Western tradition of education"(p50). Out of all the comments, the relationship formed with academic staff was most important: to be "guide, philosopher and friend"(p55) to the student.

---

[1] Details of "deep" and "surface" learning in Entwistle and Ramsden(1983)

CHAPTER 4 STUDENT EVALUATION OF TEACHING (SETE)

4.1 Introduction

The use of student ratings of instruction seems to have started in the 1920s in North America, particularly at the University of Washington. Mueller(1951) made the first nationwide study in the USA, by approaching 1000 colleges to discover how many used student ratings: 296 were using them and planned to continue.

More recently, Cashin(1988) points out that there are over 1300 articles and books dealing with research on student ratings of teaching; the vast majority originating in North America, where student rating of instruction is common practice.

Marsh(1984) outlines the purpose of student evaluation of teaching as fourfold:

> a) As a "diagnostic feedback to faculty about the effectiveness of their teaching"(p707). This is the main use for student ratings of instruction, and the issue here revolves around whether students' judgments are a true assessment of effective teaching.
>
> b) For use in teacher promotion decisions. A controverial question in North American universities and colleges - should teaching be used as the only indicator for promotion purposes?
>
> c) As information for students in selection of courses. Marsh and Roche(1993) feel that the Student Educational Evaluation Questionnaire (SEEQ) is not useful for infering course quality. "Research indicates that the ratings are primarily a function of the instructor who teaches the course rather than the course"(p244).
>
> d) To show the "outcome on a process description for research or teaching".
>
> (Marsh 1984 p707)

4.2 Ideal Lecturer

In the first British published study in recent years, Cooper and Foy(1967) compiled a list of 43 statements relating to how a lecturer may present and organise material, their general personality, and the relationship with the class. Asking undergraduate students from all three years of a Pharmacy course, the most important characteristics chosen were "presents his material clearly and logically", "enables the student to understand the basic principles of the subject", and "can be clearly heard" in that order. Least important were "has a democratic approach", and "is skilful in drawing blackboard diagrams."

In conclusion, the authors say that " the students placed a far higher priority than the staff on the adequate coverage of ground in the lecture course, on attempts to link lecture material to laboratory/practical work"(p185) as the main requirements of an ideal lecturer.

4.3 Summary of Studies.

Research studies have asked students about their ideal lecturer in three main ways -

i) To describe their ideal lecturer.

ii) To suggest what characteristics are important for good teaching.

iii) To describe the best teacher they have had.

Feldman(1976b) in a very extensive study records 19 important characteristics from 72 studies in North America. Depending on the type of study, slightly different characteristics are ranked as most important. For non-structured studies (where the students volunteer the characteristics), "respect for students" is ranked highest when the students are thinking of characteristics important to good teaching; but "clarity" is first, when they are asked to think of the ideal lecturer, and "availability" when asked to think of the characteristics of the best teacher ever had. However, in structured responses studies (where the characteristics are given), "knowledge of subject", "stimulation of interest", and "concern for class progress" are ranked highest in each type of study. Table 4.1 summarises the most important characteristics based on the different types of study.

Taking the average of the ranks in the six different types of study, the most important characteristic is "knowledge of subject", followed by "stimulation of interest", and "clarity of presentation".

Feldman's work is with North American studies only. Unfortunately, there were few British studies found, and certainly in recent years (since the 1970's) there is a lack of published work. Whether it is because the issue is seen as resolved, and no further research is needed, is unclear. Furthermore, there are also hardly any British studies on the use of student evaluation of teaching effectiveness. Table 4.2 shows the five most important characteristics in ranked order of the four British structured studies of the ideal lecturer. The characteristics chosen are similar to the North American studies. The only real

TABLE 4.1

NON-STRUCTURED STUDIES				STRUCTURED STUDIES			
Characteristic	Ideal	Imp	Best	Characteristic	Ideal	Imp	Best
Respect	6.5	1	2	Knowledge	2.5	3	1
Knowledge	5	2	3.5	Interesting	1	4	2.5
Availability	-	7	1	Class progress	4	1	5
Discussion	2	4	7	Enthusiasm	2.5	6.5	6.5
Clarity	1	10.5	5	Organisation	6	5	11
Enthusiasm	3.5	9	11	Challenge	7	8	8.5
Fairness	9.5	6	9.5	Availability	10	10	2.5
Organisation	11	12	6	Discussion	8	12	14
Elocution	12.5	5	12	Respect	11	13	10
Interesting	6.5	3	3.5	Clarity	5	2	6.5

Characteristics most associated with superior teachers and superior teaching, as determined by various procedures (Feldman1976b Table III p253). Characteristics in full in Appendix 4.

KEY: Non-Structured response studies = students choose characteristics.
Structured response studies = students choose from list given.
Ideal = thinking of ideal lecturer.
Imp = thinking of characteristics important to good teaching.
Best = characteristics of best teacher ever had.

TABLE 4.2

1. Clear and understandable explanations.
2. Presents material in well-organised way.
3. Shows thorough knowledge of subject.
4. Can be clearly heard in lectures.
5. Stimulates students to think independently.

Table showing the most important characteristics of the ideal lecturer in British studies. Items must appear in all four studies. Based on Cooper and Foy(1967); Foy(1969); Ramsden(1975); Smithers(1970a).

difference is "stimulation of interest" in the North American studies, replaced by "stimulates students to think independently" in the British studies.

Thus it would seem that presentation skills and knowledge are the most desired characteristics of the ideal lecturer for students generally.

4.4 Cross-cultural Studies

So much of the research on the evaluation of teaching has been North American, and as with many areas in psychology, cross-cultural studies are few. But in the last few years, attempts have been made to test certain rating instruments around the world. In particular, Marsh has compared the Student Educational Evaluation Questionnaire (SEEQ), and the Endeavour Instrument in Australia (Marsh 1981; 1986), New Zealand (Watkins,Marsh and Young 1987), and Spain (Marsh, Touron and Wheeler 1985). Elsewhere, the same two rating instruments have been tested in Papua New Guinea (Clarkson 1984), Nepal (Watkins and Regmi 1992), Phillipines (Watkins and Gerong 1992), India (Watkins and Thomas 1991), and Nigeria (Watkins and Akande 1992). Almost all the studies found that the rating instruments clearly distinguished between the "best" and "worst" lecturers, and that the sixty-three items were seen as appropriate by at least 80% of the students. But there were slight differences in the findings. Papua New Guinean students tended to select qualities related to the individual, rather than the organisational skills of the lecturer. In the Phillipines, "good" lecturers were seen as younger and teaching

major subjects, while "poor" lecturers were teaching lower grade courses and larger classes.

From New Zealand, Clift et al(1989) detail the preparation of the Teaching Performance Profile to use for promotion purposes among lecturers. The authors include separate scales for lectures, course, tutoring, lab supervision, and seminars. The items tended to load on to one global factor for the lecture scale, which included "stimulated interest", and "ability to communicate"(p200).

Also from New Zealand, a study by Watkins(1990) of student ratings used as information for perspective students in the "alternative calendar". The author found the ratings to be stable, but to vary according to size of class, year of course, and grades awarded. This does not make student ratings invalid. "But it does certainly make it difficult to compare the ratings of courses and lecturers across a university"(p19).

However, studies from the Middle East have produced different results to the North American studies. Dehlavi(1987) asked 400 Iranian students to choose their ideal lecturer from 39 characteristics. Regardless of their sex, field of study or academic year, they preferred attributes classed as personal-social (eg "always friendly towards students") over cognitive-intellectual (eg "has a thorough knowledge of subject matter"). The author then tries to explain the finding as reflecting "Iranian cultural values which attach great importance to obtaining 'college degree' rather than learning per se"(p145).

Safi and Miller(1986), in the first study of student evaluation in Kuwait University, found that "punctuality in starting and finishing class" was rated highest, with "feedback on assignments and tests" rated lowest. Other important items were "pleased with students' questions", and "respected and trusted students". These three most important items are not among the highest rated in British and American studies.

Finally, Harnesh-Glezer and Meyer(1991) report a study of community college students in Israel. They build upon a study by Neumann and Neumann(1981), which showed that general satisfaction with college instruction is influenced by content and organisation of lectures; satisfaction with tests and assignments; relationship between students and staff; and satisfaction with the techniques of teaching (Harnesh-Glezer and Meyer 1991 p96). Harnesh-Glezer and Meyer(1991) use an adapted version of a questionnaire developed by Chermesh(1978). This approach prefers a path-structure, where different apects of course, teaching and subject are combined to give a general satisfaction score. Harnesh-Glezer and Meyer found that the major factor in course satisfaction was the teacher, with some influence from interpersonal relations among the students. However, the authors have a word of caution: "Our results show that interpersonal relations between students are related to course satisfaction. This does not necessarily mean that they cause this satisfaction"(p105).

Could these studies be a clue to what characteristics some overseas students

coming to the United Kingdom prefer in their lecturers? Though most cross-cultural studies are similar, there are some differences. It is expected that overseas students will choose some different characteristics to the general conclusion in the literature.

4.5 Differences Among Faculties About the Ideal Lecturer.

Two English studies, Ramsden(1975) and Smithers(1970a) break down the ratings of characteristics of the ideal lecturer into faculty groups. Though there are similarities, there are also significant differences.

Smithers divides his subject areas into Engineering, Physical Sciences, Life Sciences, and Arts/Social Sciences. They all agree that "knows his subject" and "presents his subject so that students can understand it" are most important, while "keeps strictly to the text-books" is least desirable. But Physical Sciences, and Arts/Social Sciences rate significantly lower (X2 $p<0.05$) than the other two faculty groups, the following: "speaks slowly enough for full notes to be taken" and "appears to enjoy lecturing". Life Sciences, and Arts/Social Sciences rate significantly higher, "concentrates on providing the essentials of the subject as a framework for independent study".

Finally, Engineering students rate significantly higher, "appears to like students" than do Physical Sciences and Arts/Social Sciences, and significantly lower than them, "takes his own distinctive line on a controversial matter or where a problem has yet to be resolved." Smithers summarises the difference

as "students of applied science and engineering appeared to look towards lecturers for information, social scientists, for stimulation."(p141)

Ramsden compares 3 faculty groups:(1) science,engineering and mathematics;(2) arts and social science;and (3) professional courses, at what was North East London Polytechnic. He notes three significant differences in their ratings:

i)Science,engineering and mathematics students compared to arts and social science students rate higher on "adjusts his or her pace to the needs of the class"(X^2 $p<0.25$) and "is clear and understandable in his or her explanations"(X^2 $p<0.02$), but lower on "is sensitive to feelings and problems of individual students"($p<0.05$).

ii)Students on professional courses rate lower "adjusts pace" ($p<0.05$) and "sensitive to feelings" ($p<0.01$), and higher on "shows the relevance of his or her subject to the work you expect to do when you qualify"($p<0.02$), and "gives a good factual coverage of the subject matter"($p<0.05$) than science, engineering and maths students.

iii)Arts and social science students rate lower "shows relevance" ($p<0.01$) than students on professional courses.

From the North American studies, Riley et al(1950) found for the arts faculty - "knowledge of subject" and "ability to encourage thought" are most important; for Biological and Physical Sciences - "ability to explain" and "organisation"; and for the Social Sciences - "ability to encourage thought" and

"organisation". While Birney, Coplin and Grose(1960) similarly found slight differences between faculty. Students of the Humanities see "dynamic and stimulating" and "helpful to students" as the most important; students of Natural Science prefer "organised" and "helpful"; Social Sciences faculty suggest "dynamic" and "knowledge of subject" (quoted in Feldman and Newcomb 1969a). The full results of both these studies are in Appendix 6.

Cashin(1990) comparing responses to the Student Instructional Report(SIR), and the Instructional Development and Effectiveness Assessment(IDEA) produces 44 different academic fields from 87 843 classes in 316 colleges and universities in North America. Taking the average ratings on each department, the academic fields were divided into five categories, from high to low. The high rating departments tend to be arts and humanities, except English Literature and Language, and History, which are medium low. The low raters are mostly business, economics, computer science, maths, physical science and engineering. The full details are in Appendix 6.

In a more recent, American study Jennings and Bartling(1991) found significant differences between courses in the following areas: more maths and natural science majors responded "most definitely" to "A professor should try to encourage joint student/faculty research projects" than business majors. Business majors preferred courses related to professional exoeriences, while natural and social science majors wanted courses related to research.

Cashin proposes 6 reasons for the difference in ratings by faculty:

i)Quantitative courses receive lower ratings - students on these courses have less developed verbal skills, and thus these courses are more difficult to teach.

ii)Sequential courses(courses following each other in terms of building on knowledge eg science) receive lower ratings - students have a less solid foundation of knowledge.

iii)How masterable is course content? - certain courses are difficult to define in terms of their content(eg American History), and thus allow greater freedom, leading to higher ratings.

iv)Academic fields vary in academic skills and goals assessed as important.

v)Purpose of the course - difficult courses aimed at weeding out weak students may be rated lower.

vi)Some academic fields are poorly taught.

"Probably the real explanation lies in some combination of the explanations just offered"(Cashin 1990 p119).

Because of faculty differences in ratings, Barnes and Barnes(1993) suggest the same evaluation instrument may not be applicable.

Erdle and Murray(1986) took a different approach, looking at teachers from different faculties to see if their teaching behaviour varied. Trained observers

watched 124 teachers. "Interpersonal orientation" behaviour occurred more frequently with Arts and Social Science teachers than Natural Science teachers, while "task oriented" behaviour occurred more with Natural Science and Social Science teachers than Arts teachers. However, for students, what makes an effective teacher was the same across faculties.

4.6 What Do Students' Think About Student Ratings of Instruction?

Costin et al(1971) used a questionnaire with 404 psychology students at the University of Illinois. The findings were generally favourable - only 14% agreed that "ratings are a waste of time"; 49% were willing to spend more time on rating classes; and 64% wanted to rate more classes. The full results are in table 4.3.

Overall, then, students feel positive about their rating of instruction. But faculty are not so convinced.

4.7 Do Students and Teachers Agree on What is Good Teaching?

"Students strongly believed that student evaluations do measure teaching effectiveness and that they were qualified to evaluate their teachers, while faculty only slightly agreed with these statements"(McBean and Al-Nassri 1982 p278).

TABLE 4.3

	% OF RESPONSES				
Items	Strong agree	Some agree	Neutral	Some disagree	Strong disagree
1.The forms will affect most teacher's future performance	3	42	14	24	11
2. The forms will affect most teacher's departmental status or advancement	2	12	12	31	28
3. Ratings are a waste of time	4	10	9	42	33
4. I would like to be able to rate more classes	32	32	25	5	5
5. I would be willing to spend time outside of class to rate courses	17	32	16	13	17
6. In retrospect, I would rate a number of courses differently now than at the time I took them	6	27	13	33	16
7. My ratings usually agree with those of the rest of the class	1	23	16	12	2
8. I generally rate a professor higher than he deserves, since there are so few good professors	2	10	10	29	46
9. A professor who tells good jokes will get a good rating	3	24	15	30	25

Table showing students' opinions concerning student rating of instruction (N=404). (Costin et al 1971 p 523).

Getting students to evaluate their classes has overlooked the question of whether the staff regard the same things as important as the students. The easiest way to discover this information is to ask students and staff to rate the same lecturer, and then find the correlation between the two groups. The higher the positive correlation between the students' and staff ratings, the more they agree.

Ramsden(1975) asked 11 members of staff along with their classes to rank 23 statements for importance with the ideal lecturer. A Spearman rank correlation coefficient of 0.715 ($p < 0.01$) was found between student and staff rankings. But the key differences were for staff: most stress on being "well prepared for classes" and "stimulating students to think independently" - and for students: more emphasis on "ability of lecturer to stress important material".

Feldman(1988) in a detailed literature review looked at 31 studies from North America. The average correlation between faculty and students was $r = .71$ ($p < .001$). However, this correlation of the general similarity between faculty and students disguises differences. Looking at 22 individual characteristics of teaching, Feldman found differences on rating of importance, of seven. Table 4.4 shows the top ten ranked characteristics for students and staff. Students rank "sensitivity" as number one (number three for staff), followed by "preparation" (number four for staff). "Knowledge of subject" was most important for staff, and "enthusiasm" next. More details of this study appear in Appendix 8.

TABLE 4.4

STUDENTS	FACULTY
1.Teacher's sensitivity to, and concern with, class level and progress	1.Teacher's knowledge of subject
2.Teacher's preparation:organisation of course	2.Teacher's enthusiasm(for subject of for teaching)
3.= Teacher's knowledge of subject	3.Teacher's sensitivity to, and concern with, class level and progress
3.= Teacher's stimulation of interest in course and its subject matter	4.Teacher's preparation:organisation of the course
5.Teacher's enthusiasm(for subject or for teaching)	5.Clarity and understandableness
6.Clarity and understandableness	6.Intellectual challenge and encouraging of individual thought(by teacher and course)
7.Teacher's availability and helpfulness	7.= Instructor's fairness;impartiality of evaluation of students;quality of examination
8.Teacher's concern and respect for students; friendliness of teacher	7.= Teacher motivates students to do their best;high standard of performance required
9.Perceived outcome or impact of instruction	9.Teacher's availability and helpfulness
10.Instructor's fairness; impartiality of evaluation of students;quality of examination	9.Teacher's concern and respect for students;friendliness of teacher

Importance (rank order) of various instructional dimensions for students and faculty (Feldman 1988 p311).

The most comprehensive, and most recent report is by Williams,Loder and Fry(1993), which questioned 2357 students, 2309 academics, and 1222 administrators in the UK. It was a general survey of life in higher education, but one question asked what characteristics contribute to effective teaching. The results are detailed in table 4.5, which shows the slight differences in rank

order of characteristics between students, faculty and administrators. In this case, students rate "knowledge" highest, while the academics and administrators prefer "enthusiastic about subject". Large differences in ranking appear on "challenge students' thinking" (ranked seven for students, but ranked two for academics); and "detail and prompt feedback"(ranked three for students, and only six for academics).

TABLE 4.5

Characteristics	Students	Academics	Administrators
Knowledge of subject	1	3	2
Enthusiastic about subject	2	1	1
Detailed and prompt feedback	3	6	3
Good presentational skills	4	5	4
Consciously transmit knowledge to students	5	7	7
Encourage active participation by students	6	4	5
Challenge students' thinking	7	2	6
Put knowledge of Learning Theory into practice	8	-	9
Provide reading lists to guide student reading	9	-	-
Set students regular assessed work	-	8	8

Ranked items are those rated "very important".
Characteristics of lecturers contributing to effective teaching as assessed by 3 sources. (Williams et al 1993 Table 1).

Generally there is agreement between staff and students, and certainly agreement about which teachers are good, but the disagreement is over why they are rated as good (Hildebrand and Wilson 1970).

CHAPTER 5 SETE - CAN WE TRUST THEM? ISSUES WITH THE USE OF STUDENT RATINGS OF INSTRUCTION.

5.1 Can Students Judge For Themselves?

Marris(1964) says students "are still the best judges of a course of lectures, if only because they are generally the only people who listen to them"(quoted by Cooper and Foy 1967 p182). Similarly Riley et al(1950) conclude that "the students' construct of 'good teaching' is closely relevant to the effectiveness of a teacher in reaching the students"(Quoted in Flood Page 1974 p29).

Yet not everybody would agree. Bryant(1967) is scathing: "Most undergraduate students, after all, are not yet fully mature. They do not understand what they can get, should get, or will need from a college education"(Quoted in Flood Page 1974 p25). He suggests that students evaluate courses based on what is "fun" or "dull", not what is learned.

Cooper and Foy's checklist of the ideal lecturer was objected to on the basis that " student opinion is worthless"; students seek different characteristics at different times/classes; and the characterisitcs observer in the lecturer are based on the interaction with that group (Foy 1969).

So,in summary, the main arguments against student ratings of teaching are:

i) Students' decisions/evaluations are influenced by factors other than just the lecture (this is the issue of whether student ratings of instruction are biased).

ii) Students do not know what is a good lecture/teaching (this is the question of the validity of student ratings of instruction).

iii) Students change their minds over time (this is concerned with the reliability of student ratings of insruction).

These three issues are at the crux of the question of whether student evaluation of teaching effectiveness can be trusted. These points are now examined in detail.

5.2 Factors Affecting Student Evaluation of Teaching.

Because of the varied reasons for using student evaluation of teaching, the timing of the administration of the instrument varies. Thus extraneous variables become important. An expectation or evaluation is open to many influences. There is a fear that the instrument, especially one administed well into a course, will measure something other than simply the student's feeling about the lecturer/course.

And this undermines the usefulness of the instruments, says this fear. Marsh(1984), however, believes there is a "witch hunt for potential biases"(p730). Dunkin and Barnes(1986) are not afraid: "The usefulness of student evaluation does not depend on their being free of such influences, so much as the ability to take account of them"(p769).

Here are some of the main factors that could influence the students' ratings of the lecturer/course: actual/expected grades; class size; prior subject interest;

instructor rank/experience; sex of instructor/student; instructor expressiveness; characteristics of the course; student's personality; reasons for rating; administration of ratings.

1.ACTUAL/EXPECTED GRADES. Generally classes expecting or possessing higher grades give higher ratings. This is sometimes known as the "grading bias hypothesis". A number of studies support this (see Arubayi 1987), yet others also contradict (eg Bendig 1953).

Brown(1976) uses multiple regression analysis to conclude that grades do bias student ratings. Grades only accounted for 9% of the variance, but this is more than the other variables (eg class size, course level).

However, the findings are not always consistent. Cohen(1981) embarked on a meta-analysis(1) of 41 studies on this question, and was able to reject the null hypothesis of no relation between course rating and grades. [1]But within the 41 studies were variations in findings. Marsh(1984) discusses possible reasons for the findings. But the positive relationship does not always exist across all situations. For example, Anikeef(1953) found a stronger relationship between expected grade and lower the level of the class. Also there are differences across all apects of teaching. Echandia(1964) looked at accounting students: those who received higher grades rated the lecturer as better organised and as having clearer presentation than those with lower grades, but there was no difference on the lecturer's ability to motivate students.

---

[1]See Glass (1974; 1978) for more details on process of meta-analysis, and McCallum (1984).

Feldman(1976a) remains undecided after reviewing over 200 correlations; concluding that "it cannot be said that grades tend to bias evaluation. But neither can it be concluded that they do not"(p100).

2.CLASS SIZE.Flood Page(1974) feels that "either class size makes no difference,or that larger classes tend to give worse ratings" (p58). Feldman(1978) reviewed 50 studies of which 17 showed no significant relationship between class size and student ratings. The other 33 showed either a small negative correlation, or a curvilinear relationship(ie higher ratings to teachers of small and large classes compared to medium sized classes). The author attempts to explain the curvilinear relationship on the basis that increased resources are given to larger classes, or particular instructors are chosen who can teach large classes well, or instructors see the large class as a challenge and put more effort into preparation.

In a further review, Feldman(1984) found 2 studies with significant positive correlations, 22 studies with no relationship, 22 with a small negative relationship, and 8 showing a curvilinear relationship. Ignoring the curvilinear relationship, the average correlation was only r= -.09. Feldman then compares the studies showing the relationship between individual characteristics of teaching and class size (details of this study in Appendix 9). Most characteristics showed no relationship, except a negative correlation of class size with presentation of subject matter, and communication. Feldman concludes that "class size has been found to be related more frequently and

with greater strength to those instructional dimensions involving teachers' interactions and interrelations with students" (1984 p77).

Frey(1978) testing two dimensions of student rating ("skill" and "rapport") against class size found a strong, negative relationship between class size and ratings of rapport, while the skill factor showed a weak, positive relationship. This agrees with Costin et al's(1971) statement that the relationship "may vary according to the particular aspect of teaching performance that the student is asked to rate"(p521).

3.PRIOR SUBJECT INTEREST. Marsh and Cooper(1981) looked at the correlation between the student rating of the instructor, and their prior subject interest, using 511 undergraduates in Southern California. The correlation was 0.2 ($p<0.01$) for overall rating, but varied for different dimensions of teaching.

Marsh(1982a) examines 16 student/course instructor characteristics, and found that prior subject interest was the variable with the largest impact on ratings. But concludes here that "lecturers actually are more effective at teaching when working with motivated students, and that this more effective teaching is accurately reflected in the student ratings"(p85).

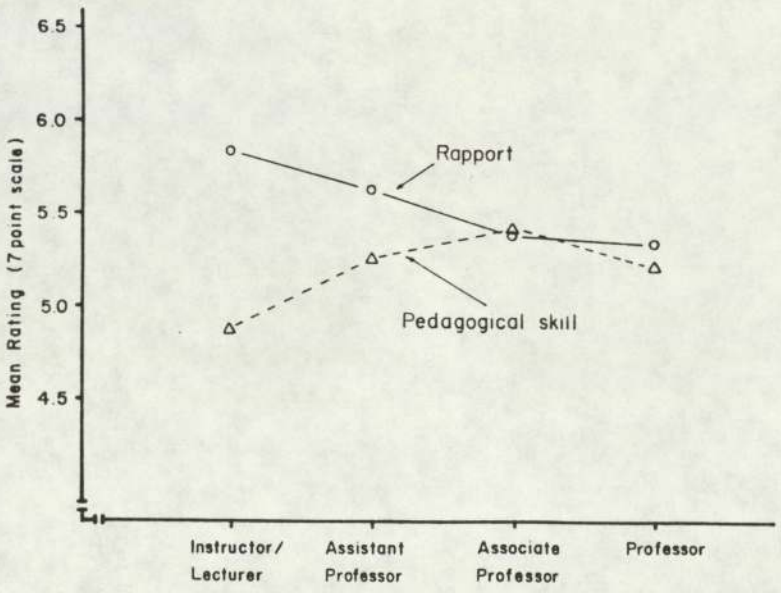
Prior subject interest is a bias, but not specifically to student ratings of instruction; for example, students with high prior subject interest usually do well in course examinations also.

4.INSTRUCTOR RANK/EXPERIENCE. Here there are mixed findings, but probably little effect(Marsh 1985). Frey(1978) reports the concern that younger instructors will get higher ratings because students identify more closely with them. Some evidence supports this (eg Clark and Keller 1954; Guthrie 1949, 1954).

However, Arubayi(1987) reports studies showing that professors receive higher ratings than lecturers (eg Downie 1952; Gage 1961).

Frey provides an answer to this contradiction at Northwestern University, Illinois using his two dimensions of student ratings. The ratings on the rapport factor decreased steadily with rank/age, while the skill factor showed the opposite trend. This can be seen in figure 5.1.

FIGURE 5.1



Ratings of "pedagogical skill" and "rapport" as function of teacher's academic rank (Frey 1978 p84).

Feldman(1983), in another of his extensive reviews, compared a number of studies under 3 headings - academic rank, age, and instructional experience. Table 5.1 shows a summary of the studies found by Feldman, and the type of correlations these studies found.

TABLE 5.1

Type of study	Studies finding significant positive correlation	Studies finding significant negative correlation	Studies finding no correlation	Studies finding other patterns
ACADEMIC RANK	10	1	21	1
AGE	0	6	6	0
INSTRUCTION EXPERIENCE	2	5	8	1

Table showing the number of studies found by Feldman(1983) showing the different relationships between seniority and teaching effectiveness.

These 3 distinctions of rank, age, and instructional experience help to account for the mixed findings. The relationship of academic rank to teaching effectiveness evaluation has more significant positive correlations suggesting that the higher the rank, the more positive the student rating of the instruction. While age has no significant positive correlations, suggesting that older lecturers are not rated more positively than younger lecturers.

5.SEX OF INSTRUCTOR/STUDENT. Sex of the student: Doyle and Whitely(1974) felt it was generally unrelated or trivial. Arubayi(1987) reports

the conclusion from findings that female students rate more favourably than male students; and they rate female lecturers more highly than male lecturers. Aleamoni and Thomas(1977) report no relationship between sex of rater and rating of faculty.

Sex of the instructor: Feldman(1992) reviewed 14 experiments producing 485 analyses, and found that for overall evaluation, there was no difference in the ratings based on gender of the lecturer. Then he examined the individual characteristics of teaching. Again, generally no difference, but if there was a difference, male teachers received higher ratings. In the second half of the article, Feldman(1993) reviewed classroom studies finding no general difference, but this time, if there was a difference it favoured women. The average correlation was only r=+.02 between the sex of the instructor and overall evaluation of teaching. Martin(1984) found that instructors who fitted social stereotypes received better evaluations. Developing this idea, D'Agostino and Dill(1988) noted that behaviours classed as friendliness towards students produced higher ratings for female instructors, but not for male. But overall, male professors were rated as more effective than female professors. The authors conclude that "male and female instructors will earn equal SRI(student ratings of instruction) for equal professional work only if the women also display stereotypical feminine behaviour"(p344).

6.INSTRUCTOR EXPRESSIVENESS. Instructor's expressiveness, sometimes known as the "Dr.Fox paradigm", in that students give high ratings

to entertaining lecturers even though the content is nonsense. Based on the original work of Naftulin, Ware and Donnelly(1973), who introduced an actor as Dr. Myron.L.Fox to give a lecture to a group of educators and mental health professionals. He was entertaining, but spoke deliberate nonsense. Naftulin et al suggested that the lecturer's expressiveness can "seduce" students into believing they have learned something.

Abrami,Leventhal and Perry(1982) compiled a meta-analysis of the studies on the "Dr.Fox paradigm", finding inconsistencies; concluding that "instructor expressiveness had a substantial impact on student ratings but a small impact on student achievement"(p446), while lecture content had the opposite relationship. The methodology of the original experiment by Naftulin et al has been criticised heavily by Frey(1978) and Marsh(1984).

7.CHARACTERISTICS OF THE COURSE. Workload: Marsh(1984) quotes his own earlier research(1982b), where two courses given by the instructor were compared. The course perceived as having the heavier workload or being more difficult was rated higher. However, Marsh does not believe this causes a bias to student ratings.

Reason for course: Research has compared optional against compulsory courses, with teachers of the latter being rated lower sometimes. Those students taking the subject as a major tend to give more positive ratings of the lecturer than non-majors (Feldman 1978).

Class level: Feldman(1977) notes an inconsistency in results of studies on class level and ratings. He suggests it is due to failure to take account of other factors. Marsh and Overall(1981) looked at the contribution of course level (ie undergraduate or postgraduate), and of course type in determining evaluations of teaching. The former was not statistically significant, while the latter was, but accounted for no more than 2-3% of variance on ANOVA analysis. The effect of a specific instructor accounted for five to ten times as much variance on the same analysis.

8.STUDENT'S PERSONALITY. Rezler(1965) administered the Purdue Rating Scale for Instruction, and the Edwards Personal Preference Schedule(which assesses student needs). They found several significant correlations:

- Male students with high needs for "nurturance", heterosexual relations, exhibitionism, and dominance rated male teachers higher.
- Female students with high needs for "succorance", heterosexual relationships, and exhibitionism rated all teachers lower (Quoted in Flood Page 1974 p44).

Smithers(1970b), working at the University of Bradford, has looked at students' scores on the Eysenck Personality Inventory (Eysenck and Eysenck 1964) and Rokeach's dogmatism scale (Rokeach 1960), and their expectations of the lecturers. Significant differences ($p<0.05$) were found on nine of the 50 items. Extraverts expected the lecturer to be "entertaining and confident" compared to introverts; low scorers on neuroticism are less concerned about "speed of lecture" and "lecturer setting a standard" compared to high scorers.

Neurotic introverts are less concerned about the "lecturer taking own line on controversial issues", and want less "use of non-textbook material" compared to other students. High dogmatism scorers have significantly higher expectations on 4 items compared to low scorers: "keeps to point"; "thoroughly prepares for lecture"; "provides duplicated notes of lecture"; and "organises blackboard work clearly".

Other studies have found differences in student ratings based on differences in authoritatianism (Freehill 1967); and general personality profile (eg Rees 1969; Yonge and Sassenrath 1968). Flood Page(1974) concludes that "there is some kind of slight effect, but not one of any practical importance"(p50). Feldman(1977) feels it is difficult to generalise:"Direction and content differences seem dependent on the nature of the rating items, the specific personality or related characteristics measured, differences in experiences and other attributes of the student, and the particulars of the courses and teachers"(p244).

9..REASONS FOR RATING.Ratings being used for promotion purposes are generally higher. Tetenbaum(1977) asked 414 students to evaluate their instructors, and they were divided into three conditions. The difference being the supposed purpose of the ratings - for promotion purposes; to improve quality of teaching; or to aid future course selection. The three conditions produced different means, and then slight variations in the factor analysis. Feldman(1979),in an extensive review, concludes that the ratings are higher for

"official" purposes (eg promotion), but the studies are limited, so caution is needed.

10.ADMINISTRATION OF RATINGS. Anonymous vs identified: Generally felt that identified ratings are higher, but Feldman(1979) emphasises the context in which students identify themselves. For example, when students were asked to identify themselves to explain their evaluations later, the ratings were always higher, than when identified but "only for research purposes" (Sharon and Bartlett 1969).

Who administers evaluation questionnaire: Kirchner(1967) found significant differences in ratings, between when the instructor or neutral observer administers the evaluation session. Presence of the instructor while being evaluated leads to higher ratings. Other factors include the instructor's demeanour during the rating if they are present; presence of the instructor's colleagues (produces higher ratings); and rapport between ratee and rater (Doyle 1983).

When evaluation questionnaire administered: Not important if "(1) the students are asked to rate typical performance; (2) they have had sufficient opportunity to observe the instructor; (3) the evaluations do not take place at the same time as special events like holidays, perhaps, or examinations that might influence the data"(Doyle 1975 p78). Frey(1976) feels that ratings administered during final exams are generally lower than those during term. Feldman(1979) points

out that the few studies that have compared the timing of the evaluation have not found any differences.

Rating format: Feldman(1979) includes 3 variables related to the format of the rating instrument that could influence the results -

i)The instructions given to the students on how to fill in the rating form.

ii)The items used ("stimulus variables").

iii)The response options available. Follman et al(1974) offered 3 groups of students different responses - "degree of agreement" with statement; degree to which improvement needed in characteristic given; and ordered categories (eg "excellent", "average"). The first two produced higher ratings(non-significant).

Feldman(1979) details other variations in rating formats that can influence the level of student ratings:

- that lead to higher ratings: the use of "degree of agreement" rather than disagreement; dropping unfavourable response items but keeping the same number of items; and positive phrasing of the "stem".

- that seem to have no effect: amount of information about the trait being assessed; offering only positive/neutral response items; varying the order of response categories; using negative numbers as response items; or the type of person to imagine (eg "ideal teacher" or "best teacher you have had")(p156-7).

There are a number of other possible biases to the student rating of instruction, but these are seen as less important in the literature. Details of these can be found in the Appendix 5.

5.3 Can We Trust Student Ratings of Teachers?

Opinions vary over the faith to put in student ratings, particularly because of the infinite number of background variables that could bias the ratings.

An interesting study is reported by Miklich(1969) from the University of Hawaii. He compares two groups he had to teach - one he knew well, the other he was teaching for the first time. For the latter he took pains to explain the examinations. The student ratings from the two groups showed a significant difference: "Fairness of Grading" was rated higher by the new group. This seems to suggest that the students were responding to the teacher's behaviour.

Marsh, who has written extensively in this area, believes in the system of student ratings, as long as expectations are not too high. Most studies, he reports, have found a correlation of 0.30 or less between student ratings and particular variables (Marsh 1984 p742).

Marsh(1982a), later concludes " that none of the suspected biases to student ratings seems actually to have much impact''(p87).

Furthermore, Dunkin and Barnes(1986) finish their literature review reasonably confident that students can perceive and rate their teaching.

So background variables do not invalidate the idea that students can tell what is a good lecture. But whether student ratings are valid and reliable, which are important issues before we can trust them, will be reviewed next.

5.4 Reliability of Student Ratings.

Do students change their minds over time,or maybe vary in the ratings from class to class due to inconsistency?

Here reliability refers to the fact that the ratings will measure the same score everytime, ie the same lecturer producing the same quality lecture on two occasions will receive the same rating by the same student.

Doyle(1975) lists the sources of reliability errors:

i) Computational error eg putting the wrong instructor's name on ratings summary.
ii) Rater's task -ie problem with nature of the questions used.
iii) Environment - physical or social environment.
iv) Rater - lacks motivation or memory problems.
- Halo effect: overall impression influences specific rating items.
- Leniency error: tendency to rate higher when known that ratings being used for promotion purposes.
- Central tendency: inclination for mid-point on scale.
- Proximity error: rate adjacent items similarly.
- Contrast error: projection of own deficiencies on to ratee.
- Logical error: rating traits that "ought" to go together

(Doyle 1975 p34).

The first study of reliability came from Guthrie (1927). 285 psychology students ranked lecturers at the University of Washington, and then again 2 weeks later. A correlation of r=0.89 was found.

Foy(1969) followed up his study with Cooper (Cooper and Foy 1967), due to objections about the original findings on an ideal lecturer. A different group of students used the same check-list as the first study, and there was a correlation of 0.93 between the two ratings (1 in 2000 possibility of a chance correlation as high as that). This seems the most straightforward evidence of the reliability of an instrument. Arubayi(1987) reviews a number of studies, "from what is available in the literature it appears that student ratings are reasonably reliable''(p269).

The reliability of individual instruments obviously is an important requirement before general use. Bradbury and Ramsden(1975) detail a reliability retest of the North East London Polytechnic student feedback questionnaire between 7 to 14 days after the original use. The reliability coefficient was 0.77 or above. Certainly for the well-established questionnaires, reliability coefficients are as expected - eg Marsh(1982a) testing the reliability of SEEQ finds correlations of between 0.74 - 0.90 using intraclass correlations (random half of class correlated to other half), and coefficient alpha between 0.88 - 0.97. Table in Appendix 7 shows a selection of recent studies, and the reliability coefficients they found. The overall reliability varies between .54 and .98.

Overall and Marsh(1980) found a significant correlation between ratings in the final year of a course, and one year afterwards; using both class-average

responses (median r=0.83) and individual student responses (median r=0.58).

Another way of looking at reliability could be to compare two groups cross-sectionally. Drucker and Remmers(1951) compared current undergraduates with alumni of 10 years, for ranking of ideal lecturer, using the Purdue Rating Scale for Instruction (Remmers 1960). Of 10 items, there was agreement on 7, including the first 4: "presentation of subject matter", "interest in subject", "stimulating intellectual curiosity", and "liberal and progressive attitude".

Centra(1974) adapted this study to look at overall assessment of teaching between current students and alumni (of five years). There was a significant correlation (r=.75) between the two groups on the rating of "best" and "worst" lecturers in the department.

So students' idea of what constitutes a good teacher remains similar as they grow older.

Braskamp et al(1985) make a number of conclusions about the reliability of SRI, which are detailed in table 5.2. Overall, though, with larger classes, student ratings of instruction are reliable.

TABLE 5.2

CONCLUSIONS	STUDIES ON WHICH CONCLUSIONS BASED
1.Student agreement on global ratings are sufficient high if class greater than 15 students.	Crooks and Kane(1981) Feldman(1977,1978) Marsh and Overall(1981) Marsh,Overall and Kesler(1979b)
2,Students are consistent in their global ratings of the same instructor at different times in the course.	Centra(1980)
3.An instructor's overall teaching performance in a course can be generalised from ratings from 5 or more classes taught by the instructor in which at least 15 students were enrolled in each class.	Crooks and Kane(1981) Kane,Crooks and Gillmore(1976)
4.The same instructor teaching different sections of the same course receives similar global ratings from each section.	Overall and Marsh(1979) Shingles(1977)

Table showing generalisations about the reliability of SRI (Braskamp et al 1985; table 4.4 p42).

5.5 Validity of Student Ratings

Do students know a good lecturer, ie are student ratings actually measuring good teaching? Here validity means that the ratings are an assessment of teaching quality, not other factors; like class size or personality of student.

McBean and Al-Nassri(1982) noted that "students strongly believed that student evaluations do measure teacher effectiveness ... while faculty only slightly agreed"(p278). This statement can be said to show face validity. Some

would argue, though, that this is only valid as an indicator of student satisfaction.

Costin et al(1971), in an early review of the literature, suggest determining the validity of student ratings as a "match" between "students' subjective criteria" and "faculty members' goal in teaching"(p513). But the question is then, what is the basis on which students make their judgments? Consistently 3 items appear in studies that Costin et al review - knowledge, and interest in subject, and preparation. However, this approach is difficult in practice, because other items are also important to students, and faculty and students disagree over the relative importance of each item.

So the approach to establishing validity has concentrated on criterion validity.

5.5.1 Objective Validation: Criterion Validation

This concentrates on the relationship of ratings with other objective measures. The most common measure used is student learning (usually defined as the grade in the course examination).

In a now famous study in "Science", Rodin and Rodin(1972) found a negative correlation between the amount learned from classes, and their rating of the teacher. They used a subjective rating of the lecturer, and an objective measure of the amount of calculus learned. The conclusion of $r = -0.75$ correlation threatened the validity of students' evaluation ratings.

But subsequent studies have consistently found positive correlations. Frey(1978) lists a number of problems with the Rodins study - for example, study based on teaching assistants rather than teachers who gave the main lectures. Further on in his article, after reviewing the studies since Rodins, Frey points out the need to study the "regular instructors", and to use " a rating form which emphasises the appropriate teaching traits"(p75). Marsh(1984) spends time to highlight methodological weaknesses with the Rodins study.

At the end of his meta-analysis of 41 studies, Cohen(1981) found a mean correlational index of 0.43 between student ratings and performance in examinations.

In another meta-analysis, McCallum(1984) examined 12 studies which used a global item evaluation of the instructor or course, and the correlation with student achievement. The average correlation was .064 for "course" and .101 for "instructor"(p155).

Doyle and Whitely(1974) used a beginners French course taught in 12 separate sections, with a common examination. There were significant correlations between level of specific ratings, and scores in the examination. When mean section ratings were used, the correlations were very small. The conclusion is that some items, but not all, are correlated to student learning.

Frey(1978) in testing the validity of the two dimensions of skill and rapport, correlated each with examination scores. Using a course divided into multiple

sections, taught by different instructors, but with a common syllabus, textbook, and examination. The median correlations are different: for the "skill" factor, it was r=0.81 but for "rapport" it was r=0.29. "The two rating factors are clearly not the same in their ability to indicate which teachers where most effective in preparing their students for the final examination"(p87).

What is effective teaching measured in terms of student learning is an unresolved issue. Doyle(1975) feels that there is "a tendency for the instructors' expositional clarity or presentation to relate to student learning as a measured by fairly traditional course examinations"(p65).

Scriven(1981), however, states that "The best teaching is not that which produces the most learning, since what is learned may be worthless"(p248). The Instructional Development and Effectiveness Assessment (IDEA) (Hoyt 1973) treats student learning as the primary measure of teaching effectiveness, by including a section for the student to report their learning progress. Thus the criterion measure of effective teaching is part of the rating instrument.

Obviously this is open to criticism, but Cashin and Downey(1992) point out that "students who report learning more tend to score higher on an external examination"(p568), and there is support for validity of self-reports generally (eg see Balk et al 1989).

Benton(1992), in a little known literature review of 31 studies correlating student achievement to ratings, is confident that "student evaluations of instruction are tapping into an important dimension of teaching"(p40). But later

admits that more research is needed as the significant correlations range from -.75 to +.96.

Doyle(1983) lists his problems with using a student achievement test as the criterion for establishing the validity of student ratings of instruction.

i) some characteristics of teaching are not linked to test scores - eg"clarity" and "rapport".
ii) it is assumed that the relationship is a linear one and thus the Pearson product-moment correlation can be used. But it is possible that it is a non-linear relationship between student achievement and student ratings of instruction.
iii) which unit of analysis should be used: a)pooled within-class analysis (individual ratings in each section of the course, and average across course); b)between-sections analysis (mean ratings of evaluation items across course); c)total-class approach (individual ratings). Doyle prefers the first approach.
iv) if subjects are randomly divided into sections of the course,then the generalizability of findings are limited

(Doyle 1983 p57-8).

(See Appendix 7 for more details of student achievement tests as a means of establishing validity).

The main alternative to final grade is to use students' gains in knowledge. But there are problems in how to measure the gain.Marsh and Overall(1980) tried to combine both criteria. They used final examination grade, ability to apply course material, and inclination to pursue the subject further. The first is seen as a cognitive criterion, while the other two are self-reported affective criteria. The students used were taking a course in computer programming. The authors, accepting methodological weaknesses, feel that more than one construct must be used to establish validity. "Therefore, because there is no universally accepted criterion of effective teaching, the validation of any

teaching effectiveness measure must focus on a wide range of indicators"(p474).

Obviously, the higher the correlation, the better for validation. But validity will be specific to a particular situation, and "must always be evaluated in relation to a situation as similar as possible to the one in which the measure is to be used"(Thorndike and Hagen 1977 p69).

5.5.2 Construct Validation

For some researchers, criterion validity is not a satisfactory method to establish the validity of student ratings of instruction, because effective teaching is a construct. Then for them construct validation is the best method. The main aim is to correlate multiple indicators of effective teaching. For example, student ratings and various criteria assessed for convergent and discriminant validity.

Howard et al(1985) use this method to establish teaching effectiveness using student ratings, colleagues ratings, teacher self-ratings, former-student ratings, and trained observers. Ratings by current and former students were most effective. Gaski(1987) takes this study to issue.

A number of criteria are used under the heading of the Multi-Trait Multi-Method(MTMM) approach (Campbell and Fiske 1959). The use of a number

of methods to measure one trait/construct allows correlations to be made; thus producing a MTMM matrix. It allows the estimation of variance due to traits or methods, and of unique or error variance.

It is possible to show convergent validity (correlation between items that should go together) and divergent validity (small or no correlation between items that should not go together). This method allows the research to estimate the effects of bias; for example, method bias: large correlation between variables because of the method used.

The main criteria used are self-evaluation by the lecturer, colleagues' evaluation, external observers, administrators, former students' evaluations, and the research productivity of lecturers.

1.LECTURER SELF-RATING. There is a general tendency for instructors to rate themselves more favourably than their students do. But there is agreement on instructor's strengths and weaknesses. Centra(1972) found differences also between faculties: instructors in natural sciences rated effort needed for their course less than did the students, while education, business, home economics, and nursing instructors were the opposite. Marsh(1982a), quoting his own studiies, finds correlations of r=0.41 for undergraduate ratings, and r=0.39 for postgraduate ratings, with lecturer's self-evaluation. Marsh(1984) is confident that this method demonstrates "acceptable validity", and also at undergraduate and postgraduate level (p723).

Feldman(1989a) makes the comparison based on individual characteristics of

teaching. Current students' and lecturers' self-evaluation are most similar in "stimulation of interest" and "availability and helpfulness", but less similar on "clarity of course objectives" and "intellectual expansiveness". Also lecturers rate themselves higher on "feedback", "friendliness", and "sensitivity" towards students.

2.RATINGS BY COLLEAGUES. In their early literature review, Costin et al(1971) find correlations between 0.30 and 0.63 for students' ratings and colleagues ratings. But in most cases, colleagues' ratings are not based on sitting through the lecture, but on "student hearsay, on the observation of the presumed effects of instruction ... and on inferences from their personal acquaintances (with the colleagues)" (Guthrie 1949 p113).

Ballard,Reardon and Nelson(1976) found correlations that range from 0.62 to 0.84. Studies based on colleagues actual visitation to the classroom are limited.

Furthermore, there is the problem that the presence of an observer can change the classroom situation - for example, by effecting the performance of the lecturer. Murray(1980) feels peer ratings are "less sensitive, reliable and valid" (p45) than student ratings.

3.OBSERVATION BY EXTERNAL OBSERVERS. Murray(1980) feels that student ratings "can be accurately predicted from outside observer reports of specific classroom teaching behaviours" (p31). The feeling is that trained observers are best, and particularly if they concentrate on specific behaviour

(eg clarity-related behaviour: number of false starts or halts in speech, redundantly spoken words, and tangles in words) (Marsh 1984 p726).

4.ADMINISTRATORS' VIEW. Cotsonas and Kaiser(1962) used clinical students in a medical school, and compared their ratings with departmental administrators. The former tended to stress the attitude towards students, and teaching skill, while the latter stressed knowledge. The authors suggest that the administrators noted the knowledge of the lecturer, and then assumed the other abilities ("halo effect"). It would also seem that the administrators took into account more than just classroom behaviour, but their general judgments about the lecturer.

5.RETROSPECTIVE RATINGS OF ALUMNI. Graduating students were asked to nominate "most outstanding" and "least outstanding" lecturers in their departments. Then undergraduates were asked to rate the nominated lecturers. Results indicated that the "most outstanding" lecturers were rated higher than the "least outstanding". A correlation of r= 0.82 between graduates' and undergraduates' choices of most and least outstanding (Marsh 1977).

Gaski(1987) suggests caution when using former students' ratings for validity purposes because "the similarity between the student and former student teaching evaluations can be explained if the primary determinant of the former student ratings is former students' recollection of the assessment they made when they were current students of the given instructor one or two years earlier"(p329).

6.RESEARCH PRODUCTIVITY. Blackburn(1974) suggested research and effective teaching were opposites. For example, McDaniel and Feldhusen(1970) found significant negative correlation between first authorship of books and students' ratings of teaching. But a significant positive correlation between second authorship of professional articles and rating of teaching.

Marsh(1984) finds no correlation or a small positive correlation between the two. "Although these findings seem to neither support nor refute the validity of student ratings, they do demonstrate that measures of research productivity cannot be used to infer teaching effectiveness or vice versa"(p729).

Feldman(1987), in another extensive review, looks at 43 studies of research productivity and overall teaching effectiveness, and finds a weak positive correlation. But when correlated with specific teaching abilities, there is a strong significant positive relationship with "knowledge of subject", and "preparation for classes".

7.OTHER CRITERIA. Marsh(1987) briefly mentions other criteria for assessing the validity of students' ratings - enrolment in advanced courses of the same subject; instructor enjoyment of teaching; open-ended comments; whether students pursue the subject further (eg Marsh and Overall(1980) computer students who rated lecturer highly were more likely to join local computer club).

Feldman(1989a) undertook a detailed literature review of the North American studies comparing overall ratings of teaching effectiveness made by current

and former students, lecturers' colleagues, administrators, external (neutral) observers, and teachers' self-evaluation. The results are summarised in table 5.3.

Feldman concludes that there is similarity between various raters, in this order: current students and colleagues; current students and administrators; colleagues and administrators (similar in relative assessment, but not in absolute assessment); self-evaluation and current students; self-evaluation and colleagues. For the other relationships, there are not enough studies to determine.

TABLE 5.3

Method Used	Current Students	Former Students	External Observers	Colleague	Adminis-trators
Current Students		+.69 (6)*	+.50(5)*	+.55 (14)*	+.39 (11)*
Former Students			+.08(1)	+.33(1)	no cases
External Observers				-.12(1)	no cases
Colleague					+.48(5)*
Administr-ators					

(*=significant correlation p<0.001 two-tailed). The number in () is number of studies found.

Table showing a summary of the studies found by Feldman(1989a) showing a correlation between different methods of assessing teaching effectiveness.

The question of establishing validity has become a methodological issue

debated in the literature, particularly around the use of criterion validity (established through multi-section courses) or construct validity (established using MTMM). Further details of the debate are in Appendix 7.

However, taking into account the weaknesses of the use of the different criteria, it is fair to say that student ratings of instruction are valid. But the criteria used are validity measures of what? Feldman(1977) looks at the purpose of the ratings - if it is to obtain objective descriptions of teachers, there may be a problem, but not if it is to measure students' subjective responses.

CHAPTER 6 OTHER METHODS OF EVALUATING TEACHING

The majority of work on student evaluation of teaching is based on self-reporting questionnaires. Elliott(1969) tried a variation on the questionnaire by choosing the items after using the repertory grid technique (Kelly 1955). This involves the students generating their own rating scales, through the usual repertory grid procedure - for example, students were asked to think of 3 lecturers, and decide a characteristic that 2 of them possess, but the third is different. This procedure is followed a number of times with different combinations of lecturers.

Elliott then analysed the responses, and extracted 10 variables which appeared in 50% of the students' responses. These were used as the basis of the evaluation form, which was found to be reliable over three years.

6.1 Group Interview Methods

One alternative is interactive discussion. Tiberius, Sackin and Cappe(1987) found that the "common assumption that a free discussion technique would generate more unanticipated issues and suggestions for teaching improvement was not substantiated"(p287). Furthermore, teachers preferred this method.

In another study, Madden and Smart(1983) compared "focus group discussions" with "standardized evaluation questionnaires." They felt that the

"richness of information derived from the focus groups supported the conclusions that valuable information is lost through the 'filter' of standardised evaluation forms"(p167).

With a similar technique, the Small Group Instructional Diagnosis technique,a facilitator works with the instructor and the students. The students are divided into small groups to discuss the strengths of the class, areas for change, and how change can be implemented. The facilitator then takes the comments to the instructor, who addresses the suggestions made. Abbott et al(1990) found this method preferred by the students to individual ratings.

The literature on the use of interviews is limited. Wulff et al(1980) found that students preferred group interviews, while Ory et al(1980) found no difference between close or open-ended questionnaires, and group interviews among students' preferences. Abbott et al(1990) attempted to examine the best way of collecting student ratings, either through group interview or standardised form; at mid-term or end of the course; and the reaction of the instructor to ratings: restricted or extended. The authors conclude that students were more satisfied with group interviews, used at mid-term, followed by extended instructor reaction.

Research suggests that faculty are happier with interviews, but only if the views expressed are representative of the whole class (table 6.1).

TABLE 6.1

CONCLUSIONS	STUDIES ON WHICH CONCLUSIONS BASED
1.Student written comments to open-ended questions are diverse and include comments about both instructor and course.	Braskamp,Ory and Pieper(1981)
2.Students tend to focus their comments on instructor characteristics(enthusiasm/rapport) and what they learned rather than organisation and structure of course.	Braskamp,Ory and Pieper(1981)
3.Students give few detailed suggestions about how to improve a course. They are better critics than course designers.	Braskamp,Ory and Pieper(1981)
4.Faculty regard student written comments as less credible than student responses to global items when information is for personnel decisions, but more credible when purpose was self-improvement.	Ory and Braskamp(1981)
5.Global overall items of instructor and course on scaled items, written comments and student interviews are similar.	Ory,Braskamp and Pieper(1980)

Generalisations about the technical quality of student written appraisals (Braskamp et al 1985 table 4.9;p56).

6.2 Other Written Methods

Doyle(1983) proposes letter-writing as another alternative. Students are encouraged to write letters of compliment or complaint. However, this is not a systematic method.

Table 6.2 shows that unstructured written comments, in whatever form, are limited in the type of characteristics of the lecturer noted. But faculty prefer them for self-improvement purposes, but not for personnel decisions.

TABLE 6.2

CONCLUSIONS	STUDIES ON WHICH CONCLUSIONS BASED
1.Global overall ratings of instructor based on scaled items, written comments, and student interviews are similar.	Ory,Braskamp and Pieper(1980)
2.Rated as more trustworthy by faculty for promotion purposes than student responses to global items and student written comments.	Ory and Braskamp(1981)
3.Representativeness suspect if not all students interviewed and/or individual students dominate the group interview.	Dawson and Caulley(1981)

Generalizations about the technical quality of student interviews (Braskamp et al 1985 table 4.10 p58).

Table 6.3 lists a choice of alternative methods of obtaining student feedback on teaching. Certainly using a variety of methods may be good for both students and staff, particularly to keep interest high in the process of rating teaching, but all the methods have the problem of low reliability. Furthermore, at the moment there is little systematic research reported using these methods (see Partington et al 1993).

TABLE 6.3

METHOD	DESCRIPTION	ADVANTAGES	LIMITATIONS
Open-ended questions		-Identify burning issues. -Student centred.	-Time consuming to process. -Low reliability.
Rounds	Students in group in turn speak, or answer question	-Easy to set up. -Plentiful information. -Current concerns of student.	-How freely will students speak? -Limited number in group. -Low reliability.
Structured Group Discussion(SGD) or pyramiding	First alone, then in twos, then fours etc to discuss issues	-Time to develop ideas. -Small group test of ideas first. -Consensus. -Student centred.	-Low reliability.
Nominal Group Technique(NGT)	Similar to "brainstorming" then ranking of items and importance	-Similar to SGD.	-Similar to SGD. -Aim for consensus.
Endings and beginnings	At end of course, before beginning of next, students' private reflections used in above methods	-Help students gain insight into own learning. -Student centred.	-Too ambitious for some students. -Low reliability.
Diaries and learning logs	Records of what actually happens in class and what they want	-Personal feelings. -Student centred.	-Idiosyncratic. -Low reliability.
"If I were you I would think about.."	Finishing sentence and dialogue follows	-Student centred.	-Low reliability.

Table comparing different methods of obtaining student feedback, and their advantages and limits.

CHAPTER 7 RESEARCH: METHOD AND RESULTS

7.1 Methodology of Project

From the literature surveyed, over 100 different characteristics of the ideal lecturer were amassed. Similar items with slightly different wording were combined - eg "explains criticism of student performance" and "makes helpful and constructive comments on essays"; "encourages to thinking" and "stimulates independent thought".

Items relating directly to course, reading or tests/assignments were removed. This left items that related directly to the lecturer, and the lecture situation: a total of 62 items (which are listed in full in Appendix 2).

7.2 Pilot Study

Because none of the past research has looked solely at overseas students, it was uppermost not to assume what characteristics they would choose as important for the ideal lecturer.

Thus the first section of the pilot study was a simple open question about characteristics of the ideal lecturer. It was used to check if any characteristics important to the students had not been included in the prescribed items. Nothing new was volunteered.

The main section involved the 62 items. The task was to rank in order of importance, the top 20. It was accepted that this was a difficult task. Ranking seemed better than rating every item on a scale of 1 to 6. The variation of rating was tried with a small number of questionnaires, but as expected, it was hard to distinguish between the items.

Furthermore, because of the fear of social desirability, which it was felt could be a major problem with some nationalities, ratings were avoided.

Experience has shown that in some parts of the world, making any comment that could be construed as a criticism of the teacher, is unacceptable. In many cultures, particularly in Asia, the teacher has a very high status (eg Regmi 1987). This also poses a problem for the final rating questionnaire - whether the ratings of a lecturer are a true comment by the student or a socially desirable answer? Certainly further research would need to try to isolate if there is an effect of this potential bias.

It was suggested that the 62 items should have been grouped together in certain ways, in order to ease the ranking. But because of the desire not to influence the results in any way, the 62 items were randomly listed. The sample used for the pilot study were all approached individually, and known to the researcher; so it was hoped this would encourage them to spend the extra effort on ranking the items. They were all students at the University of Birmingham or Aston Business School, and twenty-four questionnaires were returned.

The nationality make-up of returned questionnaires of the pilot study were as follows: 11 Japanese; 3 Brazilian; 2 Indonesian; 2 Taiwanese;1 Dutch; 1 German; 2 Greek; 1 Portuguese; and 1 Thai. This included 16 males and 8 females.

The sample came from the following departments: English for Overseas Students Unit(11); International Studies(2); Development Finance(2); Aston Business School(5); Mechanical Engineering(1); Psychology(1); Economics(1); and Law (1).

From their responses the five most popular items were: well-organised and clearly presented lectures; has good knowledge of own subject; is well-prepared for classes; structures difficult topics in easily understandable ways; and has good general knowledge.

With least important items being: makes good use of the overhead projector; always well-groomed, neat and clean in personal appearance; skilful in blackboard drawing; keeps strictly to textbook in lectures; has good vocabulary; and includes many relevant personal anecdotes.

7.3 Final Study: Choice of Sample

The majority of the @2000 overseas students in Birmingham, are studying at the University of Birmingham.

Table 7.1 shows the number of overseas students by nationality and

department. For convenience, the nationalities have been grouped into regions or homogeneous groups(as far is possible), and the same with departments.

TABLE 7.1

Dept/ Nation	CSS	ACC	ADM	ED	ART	MED	CIV	MEC	MA	SCI	LAW
1.SE ASIA (644)	47	163	16	41	78	15	39	171	19	31	24
2.INDIAN S-C (100)	0	22	17	5	5	5	11	10	7	14	4
3.AFRICA (285)	4	52	34	74	8	6	41	27	7	23	9
4.ARABIC (72)	2	12	2	2	6	10	10	10	7	10	1
5.EUROPE (415)	81	46	4	7	65	24	26	63	31	34	34
6.CARIB/N.AM (74)	14	5	5	3	22	0	6	6	2	5	7
7.C+S.AM (30)	2	5	3	1	4	1	2	4	3	5	0
8.MISC. (23)	0	2	0	1	0	1	4	3	1	1	10
TOTS (1643)	149	307	81	134	188	62	139	294	77	123	89

Table showing numbers of overseas students at the University of Birmingham by national and departmental grouping;(situation on 15/10/92).

Key to table 7.1:

DEPARTMENTS

1. CSS = Commerce and Social Science - Centre for West African Studies; Undergraduate Commerce; Cultural Studies; Economic and social History; Centre for Russian and East European studies; Political and International Studies; Centre for Urban and Regional Studies.

2. Acc = Accounting and Business - including Development Finance.

3. Adm = Administration - Development Administration; Development Management; Health Service Management;Local Government Studies; Social Policy and Social Work.

4. Ed = Education.

5. Art = Arts and Languages - Ancient History and Archaeology; Byzantine Studies; Classics; Drama; French; German; Geography; Hispanic Studies; History; Italian; Music; Philosophy; Russian; Theology; English.

6. Med = Medical and related - Medical Microbiology; Cancer Studies; Nursing Studies; Physiology; Dentistry; Infection; Immunology; Medicine; Obstetrics; Rheumatology; Institute of Occupational Health.

7. Civ = Civil Engineering and Chemical Engineering.

8. Mec = Mechanical, Manufacturing, and Electronic and Electrical Engineering.

9. Ma = Mathematics, Physics, and Computer Science.

10. Sci = Other Science - Biochemistry; Biological Science; Chemistry; Metallurgy and Materials; Earth Sciences; Psychology.

11. Law = Law.

NATIONALITIES

1 - South East Asia: Hong Kong; Indonesia; Japan; Korea(North); Korea(South); Malaysia; China; Phillipines; Singapore; Taiwan; Thailand; Vietnam.

2 - Indian Sub-continent: Bangladesh; Brunei; Burma; India; Madagascar; Maldive Islands; Mauritius; Nepal; Pakistan; Seychelles; Sri Lanka.

3 - South,West and Central Africa: Botswana; Gambia; Ghana; Ivory Coast; Kenya; Lesotho; Malawi; Mozambique; Namibia; Nigeria; Senegal; Sierra Leone; South Africa; Swaziland; Tanzania; Uganda; Zambia; Zimbabwe.

4 - Arabic speaking: Algeria; Bahrain; Egypt; Ethiopia; Iraq; Iran; Jordan; Kuwait; Libya; Oman; Quatar; Saudi Arabia; Sudan; Syria; United Arab Emirates.

5 - Europe: Albania; Austria; Belgium; Bulgaria; Cyprus; Czechoslovak Republic; Denmark; Estonia; France; Germany; Greece; Hungary; Italy; Latvia; Lithuania; Malta; Norway; Netherlands; Poland; Portugal; Romania; Russia; Republic of Ireland; Slovenia; Sweden; Switzerland; Spain; Turkey; USSR; Yugoslavia.

6 - Caribbean and North America: Antigua; Barbados; Bermuda; Cayman Islands; Dominican Republic; Guyana; Jamaica; Mexico; St.Kitts-Nevis; St.Lucia; Trinidad and Tobago; USA; Virgin Islands.

7 - Central and South America:Brazil; Chile; Costa Rica;Ecuador; Peru; Venezuela

8 - Miscellaneous: Australia; Israel; New Zealand.

The aim of the nationality groupings was homogenuity, as much as is possible. For example, South East Asia was classed separately from the rest of Asia (ie Indian sub-continent). Other studies (eg Singh 1970) simply divided

the world by continents. But there are differences between, for example, Indian students and Chinese students; too much to group together. It is accepted that the nationality groupings have limitations. The departmental groupings were for convenience most of all. Though there are some natural groupings - eg medicine and related disciplines.

From table 7.1, it is possible to see that the largest nationality grouping (39.2%) of the total overseas students at the University of Birmingham come from South-east Asia. The largest departmental grouping is Accounting and Business (18.7% of total overseas students).

The final questionnaire was made up of the top twenty items chosen from the pilot study, plus two of the least popular items as controls. The total was 22 items to rank (see copy in Appendix 2).

969 questionnaires were sent out, through the internal post at the University of Birmingham to the departments of the sample. The response was 165 (rate of 17%). But not all the responses were complete, some had information missing. Thus the figures for frequencies vary slightly.

The response was felt to be disappointing. However, it was not possible to send a follow-up letter because it was not clear whether the list of overseas students being used was up-to-date. The list was compiled in October 1992,

but not used until the following March. Thus a number of students may have left the University.

7.4 Analysis of Responses

The aim was for the original sample to mirror the make-up of overseas students at the University of Birmingham. This was so, except for Europe (nationality grouping 5). A large number of the students were short term for 3 months, and had returned home by the time of the questionnaire (March). Otherwise the returns mirror the make-up of the overseas student population fairly closely. This can be seen in table 7.2. For example, students from South-east Asia make up 39.2% of all overseas students at the University of Birmingham, and 43% of the returned questionnaires were from this group.

Table 7.3 shows that once more the returns mirror the make-up of overseas students by departmental groupings in the University. The only exceptions are Accounting/Business (group 2)(higher return rate than University rate), Mechanical and Electrical Engineering (group 8) and Law(group 11) (both lower return rates.

TABLE 7.2

NATIONAL GROUPING	NUMBER OF RETURNS	% OF TOTAL RETURN	NATIONAL GROUPING AS % OF SAMPLE	NATIONAL GROUPING AS % OF ALL OVERSEAS STUDENTS AT B.U
South-East Asia	71	43.0	40.8	39.2
Indian Sub-continent	12	7.2	9.8	6.1
Africa	36	21.8	21.9	17.4
Arabic-speaking countries	8	4.8	7.4	4.4
Europe	17	10.3	10.3	25.3
Caribbean/N America	7	4.2	7.6	4.5
C and S America	9	5.4	2.2	1.8
Missing details	5	3.3	-	-
TOTAL	165	100	100 (969)	100 (1643)

Table showing the returns of questionnaire by nationality.
(See page 82 for full details of nationality groupings).

A large majority of overseas students generally are attending one year courses. Table 7.4 shows that 62.5% of the returns are from students in the first year of their course.

TABLE 7.3

DEPARTMENT GROUPING	NUMBER OF RETURNS	% OF FINAL RETURNS	DEPART GROUPING AS % OF SAMPLE	DEPART GROUPING AS % OF ALL OVERSEAS STUDENTS AT B.U
Commerce/Social Science	11	6.7	7.8	9.1
Accounting/ Business	45	27.8	23.3	18.7
Administration	13	7.9	6.8	5.0
Education	5	2.4	6.6	8.2
Arts/Language	24	14.7	9.0	11.5
Medicine	10	6.0	4.0	3.8
Civil/Chemical Engineering	7	4.3	9.7	8.5
Mechanical Engineering	13	7.9	13.5	18.2
Maths	8	4.9	5.0	4.7
Other Science	11	6.7	8.5	7.5
Law	1	0.8	5.8	4.8
Missing details	17	10.1	-	-
TOTAL	165	100	100 (969)	100 (1643)

Table showing returns of questionnaire by departmental groupings.
(See page 81 for full details of departmental groupings).

A large majority of overseas students generally are attending one year courses. Table 7.4 shows that 62.5% of the returns are from students in the first year of their course.

TABLE 7.4

YEAR OF COURSE	NUMBER OF RETURNS	% OF RETURNS
1	103	62.5
2	29	17.9
3	19	11.6
4	1	0.6
Missing details	13	7.4
TOTAL	165	100

Table showing the returns of the questionnaire by year of course.

The students recorded the length of time in the UK at the point of answering the questionnaire. Because most were here for a 1 year course, then most arrived in October 1992. The distribution of the questionnaire in March/April 1993 meant they would have been in the UK about 6 months. However, some students take an English course beforehand, which can last up to 3 months. Thus these students would have been in the UK around 9months. Table 7.5 shows that 40% of the respondents had been in the United Kingdom for six months or less.

The vast majority of the returns (83.4%) had not studied in the UK before(table 7.6).

TABLE 7.5

LENGTH OF TIME IN UK (MONTHS)	NUMBER OF RESPONSES	% OF RESPONSES
0 - 6	66	40
7 - 10	32	19.6
11 +	62	38.5
Missing details	5	1.9
TOTAL	165	100

Table showing the returns of the questionnaire by length of time in the UK at time of answering questionnaire.

TABLE 7.6

IS THIS YOUR FIRST VISIT TO UK TO STUDY?	NUMBER OF RESPONSES	% OF RESPONSES
YES	136	83.4
NO	24	14.7
Missing details	5	1.9
TOTAL	165	100

Table showing the responses to the questionnaire based on whether the students had studied in the UK before.

Most of the overseas students at the University are male(@70%), and this is shown by nearly three-quarters of the returns are from males (71.4%) (table 7.7). Furthermore, for the most part the age is between 21 to 40 (age groups 2 and 3)(table 7.8).

TABLE 7.7

SEX OF RESPONDENTS	NUMBER OF RESPONSES	% OF RESPONSES
MALE	117	71.4
FEMALE	43	26.7
Details missing	5	1.9
TOTAL	165	100

Table showing the returns of the questionnaire by sex.

TABLE 7.8

AGE GROUPS OF RESPONDENTS	NUMBER OF RESPONSES	% OF RESPONSES
18 - 21	15	9.1
22 - 30	57	35.7
31 - 40	65	40
41 - 50	18	11.4
51 +	5	1.9
Missing details	5	1.9
TOTAL	165	100

Table showing returns to questionnaire by age groups.

Though the overall response rate was below 20%, which was felt to be a little disappointing. It was accepted that this can be the case for many postal questionnaires (Bridge 1974; Monette et al 1990). The responses do give a good cross-section of the overseas students in the University of Birmingham.

7.5 Results of Final Study

The subjects were asked to rank the 22 characteristics in terms of importance for the "ideal lecturer" (1=most important). Then the results were scored as 22points for 1 through to 1 point for the 22nd item. The items chosen for inclusion were the top 20 ranked from the pilot study, plus the two least popular as controls.

It could be argued that the data is ordinal. But it is necessary to class it as interval to allow the use of parametric tests. Thus we assume the difference between each score is equal distance, and there is no absolute zero (Stone and James 1965; see also Coolican 1990; Kenny 1986; Miller 1983; Miller 1984; Monette et al 1990; Runyon and Haber 1977). The analysis was by SPSS PC (version 4.0.1 SPSS Inc 1989).

Table 7.9 shows the final ranking of the characteristics based on mean score. (See Cranton and Smith 1990 for reasoning behind choosing mean score). The top five characteristics chosen were "well-organised and clearly presented lectures"; "is well-prepared for classes"; "explains concepts,principles and abstract theories clearly and precisely"; "has good knowledge of subject"; and "structures difficult topics in easily understandable ways". Not surprisingly "speaks clearly and audibly" is also important. Items in 1st and 2nd position are quite similar, as are those in 3rd and 5th position. Thus the most important characteristics of lecturers for overseas students can be summarised as:

i) Prepared and organised lectures.

ii) Clear explanations of topics.

TABLE 7.9

Characteristics	Mean
1.Organised (G).	18.24
2.Prepared (D).	16.74
3.Explains concepts (H).	16.47
4.Knowledge of subject (F).	15.83
5.Structures difficult topics (E).	15.55
6.Speaks clearly (P).	13.54
7.Objectives (L).	13.00
8.Aware (Q).	12.24
9.Stimulates independent thought (J).	12.29
10.Handouts (A).	12.22
11.Express themselves (M).	11.73
11.Feedback (N).	11.73
13.Developments (B).	11.35
14.Motivates students (R).	10.77
15.Welcomes questions (U).	10.39
16.Interest in subject (S).	10.29
17.Relate to other subjects (I).	10.25
18.Answers questions accurately (K).	10.22
19.General knowledge (C).	10.04
20.Sense of humour (T).	7.13
21.Good vocabulary (V).	4.92
22.Personal anecdotes (O).	4.65

Table showing most popular items of ideal lecturer in rank order. The full characteristics are listed in Appendix 2. The letters in brackets refer to the position on the final questionnaire used.

This compares closely with the earlier British studies, but slightly different to the North American studies. Feldman(1976b) found "respect for students"; "clarity of presentation"; "availability outside class"; "knowledge of subject"; "concern for class progress"; and "interesting to students" to be the most important items, depending on the type of study.

It was hoped at the start of the research that there would be nationality differences in the ranking of characteristics. Table 7.10 shows the top ranked characteristics by nationality groupings. Unfortunately the number of responses from certain nationality groupings are small, thereby questioning the validity of their rankings as representative on their group as a whole. To overcome this weakness, the significance level used was $p<.001$ (Miller 1984).

TABLE 7.10

NAT/ RANK	1.SE ASIA	2.INDIAN S-C	3.AFRICA	4.ARABIC	5.EUROPE	6.CAR/N.AM	7.C+S.AM
1	Organised	Organised	Prepared	Prepared	Organised	Organised	Organised
2	Explains concepts	Explains concepts	Knowledge	Structures difficult topics	Explains concepts	Prepared	Prepared
3	Structures difficult topics	Prepared	Objectives	Objectives	Prepared	Knowledge	Structures difficult topics(=)
4	Prepared	Structures difficult topics	Organised	Organised	Knowledge	Interest	Independent thought(=)
5	Knowledge	Speaks clearly	Explains concepts	Motivates	Structures difficult topics	Independent thought(=)	Developments
6	Speaks clearly	Knowledge	Structures difficult topics	Developments	Aware	Explains concepts(=)	Knowledge
7	Handouts	Independent thought	Speaks clearly	Structures difficult topics	Speaks clearly	Feedback	Feedback
8	Objectives	Express themselves	Aware	Aware	Express themselves	Express themselves	Express themselves
9	General knowledge	Aware	Independent thought	Express themselves(=)	Handouts	Motivates	Interest
10	Independent thought	Handouts	Feedback	Feedback(=)	Feedback	Structures difficult topics	Explains concepts/ Relate to other subjects
N=	53	9	29	6	17	5	8

Table showing top ten ranked characteristics by nationality groupings. Key to characteristics and nationality on pages 81 and 82.

Most of the nationality groupings place "Well-organised and clearly presented lectures (G)" as number 1, closely followed by "Explains concepts,principles and abstract theories clearly and precisely (H)"; "Is well-prepared for classes (D)"; "Good knowledge of subject (F)"; and "Structures difficult topics in easily understandable ways (E)". A Kruskal-Wallis 1 way ANOVA found significant differences on two characteristics:

i) "States objectives clearly at beginning of course (L)" - ($X2=27.6$ $p<.001$)rated higher by African students(nationality grouping 3)($N=29$). But all significantly higher except Arabic-speaking students (nationality group 4)($N=6$).

ii) "Sense of humour (T)" - ($X2=26.0$ $p=<.001$) rated higher by South-east Asian students(nationality group 1)($N=53$) than all others, except Caribbean and North American students(nationality group 6)($N=5$). Not significant, though, with Arabic-speaking and European students (nationality groups 4 and 5)($N=6$ and 17 respectively).

All the nationality groups, except 1 and 3, were combined. A Kruskis-Wallis between nationality groups 1,3 and the rest found one further significant differences. "Good general knowledge (C)" rated higher by South-East Asian students(nationality group 1) ($X2=14.4$ $p=<.001$). Analysis by independent t-test of nationality groups 1 and 3 found no other significant differences.

With nationality group 1 (South-east Asian students) there were 71 responses; this allowed a sub-division into individual nationalities. Table 7.11 shows the number of responses per country.

TABLE 7.11

NATIONALITY OF STUDENT	NUMBER OF RESPONSES	% OF ALL NAT=1 RESPONSES
Japan	19	26.8
Hong Kong	17	24.0
China	13	18.3
Indonesia	11	15.5
Taiwan	4	5.6
Thailand	4	5.6
S.Korea	2	2.8
Malaysia	1	1.4
TOTAL	71	100

Table showing the responses by nationality within the nationality group 1 (South-East Asia).

A comparison of Japanese and Hong Kongee responses produced no significant differences between them, nor between Japanese students and the rest of the South-east Asian group, on all characteristics. Practically though, Japanese students placed "speaks clearly and audibly" (P) higher, and "is well-prepared for classes" (D) lower than Hong Kong students. The numbers of the other individual countries in this group were too small to use. Even the Japanese and Hong Kong samples are relatively small admittedly.

Further analysis was made based on departmental groupings. Only 2 of the groups had responses greater than 20 - ie Accounting/Business (group 2) and Arts/Language (group 5). Table 7.12 shows the rankings of these two departmental groups.

TABLE 7.12

RANKING/DEPT	ACCOUNTING/ BUSINESS	ARTS/ LANGUAGES
1	Prepared	Organised
2	Organised	Knowledge
3	Explains concept	Prepared
4	Structures difficult topics	Aware
5	Knowledge	Express themselves
6	Objectives	Interest
7	Handouts	Explains concepts
8	Speaks clearly	Independent thought(=)
9	Developments	Motivates(=)
10	Aware	Speaks clearly

Table showing top ten characteristics as rated by departmental groups 2 (Accounting/Business) and 5 (Arts/Languages).(See Appendix 2 for full details of characteristics).

An Independent t-test found a significant difference between these two departmental groups on "makes good use of handouts (A)" (departmental group 2 greater rating $t=4.75$ separate variance; 2 tailed $p<.001$). It is also noted that departmental group 5 placed greater importance on "encourages and stimulates students to independent thought(J)" and "shows interest and enthusiasm in own subject(S)" (not significant).

Analysis by year of course, time in the UK, and age group found no significant or real practical differences in characteristics chosen. For sex of respondent, or whether this was their first time of studying in the UK, there

were differences at a lower level of significance. Female students rated "good knowledge of subject (F)", and "well-organised and clearly presented lectures (G)" higher than male students ($p = < .01$ 2 tailed with separate variance; $t = 2.94$ for former and $t = 2.91$ for latter). For "visit", there were differences for "good general knowledge (C)" and "is well-prepared for classes (D)". Those who had not studied in the UK before rated "C" higher ($p = < .01$ 2 tailed; $t = 3.08$ separate variance), but lower on "D" ($p = < .01$ 2 tailed; $t = 3.33$ separate variance).

The lack of differences between sub-groups confirms the strength of agreement on the total rankings.

7.6 Further Analysis

Factor analysis was performed on all the responses. The literature reviewed had found a number of factors. The data from the questionnaire was tested by the Bartlett test of sphericity, producing a significant level of $p < .00001$. This suggested the use of the factor model (Maria and Norusis 1988). But the anti-image correlation matrix produced a large number of high coefficients. This suggested against the use of the factor model (Maria and Norusis 1988). Also, practically from looking at the correlation matrix, few correlations between characteristics were very large (see Appendix 15). Less than five of the correlations were greater than 0.3 (Maria and Norusis 1988). Not surprisingly, factor analysis did not produce any clear-cut results (13 different approaches were tried). Appendix 14 shows an example of the results from factor analysis.

CHAPTER 8 DISCUSSION OF RESULTS/ISSUES

8.1 Introduction

The use of factor analysis on the results of the questionnaire to overseas students at the University of Birmingham produced no clear factors, as detailed in chapter 7. This goes against the other major studies in the literature. For example, Smithers(1970a) starting with a list of fifty characteristics of the ideal lecturer was able to reduce them to twelve rotated factors. Among the rating instruments, the Student Educational Evaluation Questionnaire (SEEQ) proposed by Marsh, has 9 factors: learning/value; instructor enthusiasm; organisation; individual rapport; group interaction; breadth of coverage; exams/grading; assignment/reading; and workload/difficulty. These factors have most recently been confirmed by Marsh and Hocevar(1991).

Feldman(1976b) summarising the research on the characteristics of the ideal lecturer reduces 19 characteristics to 3 groupings:

i)Presentation - stimulation; enthusiasm; knowledge; preparation; clarity; overall evaluation; and course materials; intellectual expansiveness; elocution; sensitivity.

ii)Facilitation - respect; encouragement of questions; intellectual challenge; availability.

iii)Regulation - not so clear.

Full details of Feldman's characteristics in Appendix 4. Figure 8.1 shows how

the characteristics are inter-related.

FIGURE 8.1

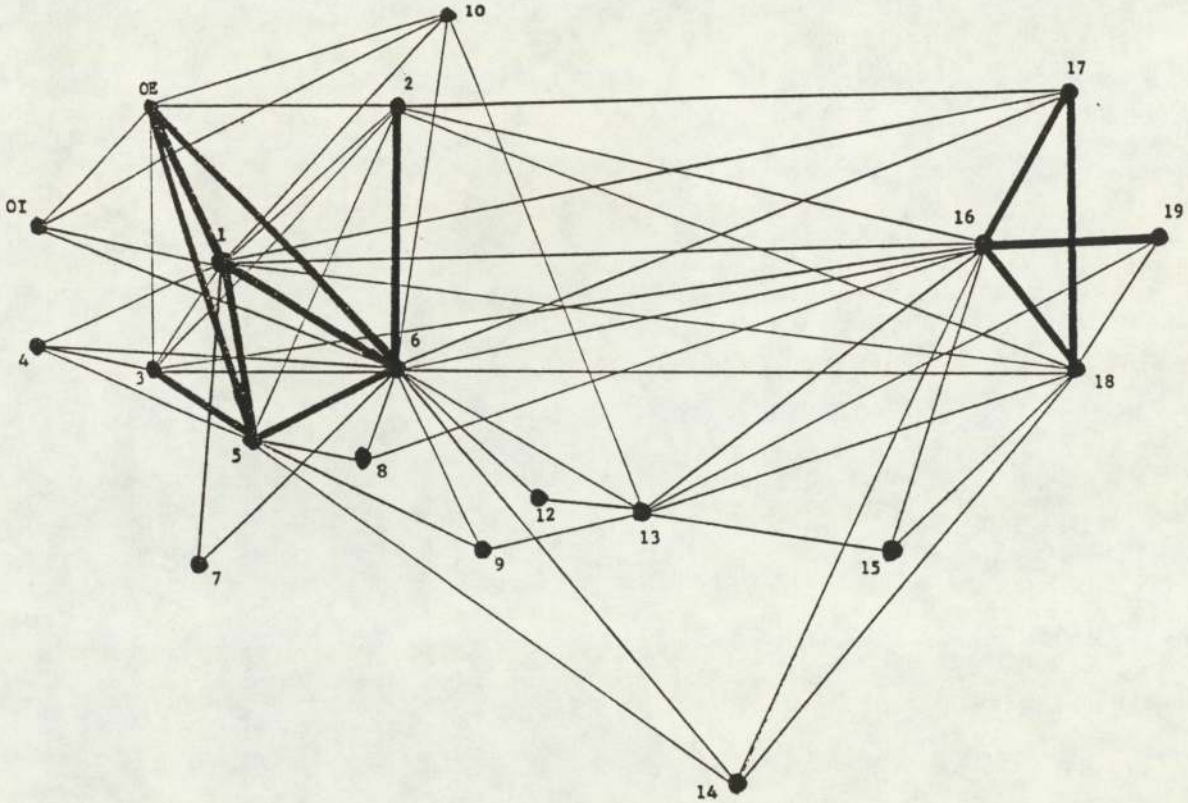


Diagram to show how teaching characteristics can be grouped. Based on Feldman(1976b p261).

However, the fact that the results of the study at the University of Birmingham would not reduce to a smaller number of factors is not seen as a problem. The use of factor analysis generally, is not without difficulties.

Rees(1990) lists the three main objections to the use of factor analysis generally:

i) "Factor" output "wholly dependent upon the data entered into the calculation"(p11).

ii) There are many possible factor analytic solutions.

iii) The number of factors to extract, and the naming of them is subjective.

Furthermore, it highlights the fact that the choice of characteristics of the ideal lecturer by overseas students may be different to previous studies, which probably all used home students. This adds support to the need for the construction of a separate ratings instrument for overseas students.

8.2 Birmingham Overseas Students Teaching Evaluation Questionnaire (BOSTEQ)

A separate rating instrument for the use by overseas students is constructed from the choice of characteristics for the ideal lecturer. But two important points need to be addressed still. Firstly, should there be a general overall evaluation of the lecturer question? Linked to this is the issue of whether teaching is a single or a multi-dimensional behaviour. Secondly, what type of rating scale should be used, and the effect on the response of the type of rating scale used.

8.2.1 Are Student Ratings Single or Multi-Dimensional?

Doyle(1983) originally proposed that all teaching behaviour could be covered by including 3 summary questions -

> i) how would you rate this instructor's overall teaching ability?
> ii) how would you rate the overall effectiveness of this course?
> iii) how much have you learned as a result of this course?
>
> (Doyle 1983 p36).

The issue of whether student ratings are assessing a single or multi-dimensional behaviour in teaching became a hotly debated issue, particularly with the publication of a series of articles in the Journal of Educational Psychology in 1991. Herbert Marsh is the main proponent of a multi-dimensional approach to student ratings, while Abrami disagrees.

Marsh(1984) has no doubt that student ratings "should be unequivocally multi-dimensional (eg a teacher may be quite well organised but lack enthusiasm)"(p709). He is against a selection of items which are then summarised by an average. "If a survey contains a hodgepodge of different items and student ratings are summarised by an average of these items or an overall rating, then there is little basis for knowing what is being measured"(Marsh 1983 p151).

In a recent article, Cashin and Downey(1992) reviewing the whole debate between Marsh and Abrami, point out that a major obstacle to resolving the debate is the "lack of any agreed on criterion measure of instrumental

effectiveness"(p564). Using the Instructional Development and Effectiveness Assessment (IDEA) student rating system (Hoyt 1973), the authors are forced to accept student learning, with controls for possible bias, as the criterion of effective teaching.

"The results of this study have supported that single, global items - as suggested by Abrami(1985) - can account for a great deal of the variance resulting from a weighted composite of many multi-dimensional student rating items"(Cashin and Downey 1992 p569).

Based on this conclusion, a general overall evaluation question will be included in the Birmingham Overseas Students Teaching Evaluation Questionnaire.

8.2.2 Construction of BOSTEQ

The second issue in the construction of the BOSTEQ is the type of rating scale to use.

Doyle(1975) sees ratings composed of scales with a response mode provided (eg agree/disagree); stems that pose a question; and cues or anchors using adjectives or phrases to define the points on a scale. In a later book, Doyle(1983) extends the various ratings that could be used to include graphic, adjectival and numerical scales; Bars (Behaviourally Anchored Rating Scales); forced choice scales; variable-item ratings; and mixed formats.

Flood Page(1974) lists examples of early rating forms, and the different systems they use. The most popular is the list of desirable teacher qualities, and the students must rate their teacher on each of them. The most common used rating is a scale with a numerical score, where 1 = poor to 5 = excellent.

But how many points on the scale? 4 or 5 is most common. Sharpness and reliability is reduced with increasing the number of points. Wherry(1952) constructed a scale with 25 points, while Doyle(1975) recommends avoiding extremes.

An alternative to traditional scales is a double-scale format - eg: Gagne and Allaire(1974) students rate instructor as they are now, and how they would want them to be. The difference between the two scores is used as an index of satisfaction or dissatisfaction.

A variation of this format involves a profile of the student's instructional needs (self ratings), and a description of what the course offers relative to satisfaction of each of those needs (Doyle 1975).

Braskamp et al(1985) summarise research on the instrumentation in student evaluation.

i) Placement of items - specific items placed before global items have a minimal effect on overall ratings. Thus global ratings can be placed at either the beginning or the end of the survey (Ory 1982).

ii) Number of response alternatives - 6 point response scales yield higher item reliabilities than 5 point response scales. Thus global items should use more 5 point response scales (Masters 1974).

iii) Negative wording of items - overall ratings are not significantly affected by number of negatively worded items. Thus both positive and negative worded items can be used (Ory 1982).

iv) Labelling all scale points vs labelling only end points - labelling only end points yields slightly higher means. Thus response format used should be consistent for all items (Frisbie and Brandenburg 1979)

(Braskamp et al 1985 p45).

It is also necessary to ask whether the term "satisfied" and "dissatisfied" should be used. Peterson and Wilson(1992) show how the manner in which the question is asked about satisfaction can influence the response. They asked a question about cars; either as "how satisfied" or "how dissatisfied". The first question produced a 91% response of very or somewhat satisfied, and the second only 82% "Posing a satisfaction question in a positive form appears to lead to greater reported satisfaction than posing it in a negative form"(p65).

Questions asked earlier in the questionnaire influence subsequent answers. Peterson and Wilson(1992) found that "asking a general satisfaction question prior to a specific vehicle satisfaction question slightly increases the tendency for a 'very satisfied' response to the vehicle question"(p66).

Panney(1977) compared two versions of a rating form - one biased towards a lecturer's strengths, the other towards weaknesses. The response to the global items at the end of the rating form were as expected.

McClendon and O'Brien(1988) found that question order had an effect on questions of satisfaction with life. The placing of specific questions before

general questions is important: "respondents must think about specific life domains in order to answer the general questions"(p361). For example, general well-being question will be effected by early specific questions about marriage satisfaction.

This could have consequences for rating instruments asking specific questions, and then finishing with a question about overall evaluation of the teacher.

Schuman and Presser(1981) talk in detail about question construction on attitude surveys, including open vs closed questions; "don't knows" problem; middle position on scales; tone of wording; and order effects of questions.

Taking into account the above comments, and the common practice of student ratings of instruction, the BOSTEQ was constructed. It contains 12 statements about the lecturer (these being the top 12 characteristics chosen by the study), a general overall evaluation question, and an open-ended section for other characteristics. An example appears in Appendix 1.

8.3 Conclusion

The survey of a sample of overseas students about the characteristics of their ideal lecturer did not produce responses that could be reduced by factor analysis. This is contrary to the literature. Furthermore the responses from different nationality groupings did not offer many significant differences ($p<0.001$).

The prime difficulties was the surprisingly small number of responses (17%). This probably accounts for the above two failings, and the spuriously high significance level of the Bartlett test of sphericity (p96).

However, the question of social desirability is always prominent with some nationalities. Thus the items chosen could have been what the subjects felt were the correct answers (ie the ideal lecturer) rather than what they themselves prized in a good lecturer.

Certainly there are many issues that arise in the construction of such an evaluation questionnaire, and further research is without doubt needed. Not least of all, to actually test the BOSTEQ with a lecture class.

BIBLIOGRAPHY

Abbott,R.D,Wulff,D.H,Nyquist,J.D,Ropp,V.A and Hess,C.W(1990) "Satisfaction with processes of collecting student opinions about instruction: the student perspective" Journal of Educational Psychology 82(2), 201 - 206.

Abrami,P.C(1985) "Dimensions of effective college instruction" Review of Higher Education 8, 211 - 228.

Abrami,P.C(1988) "SEEQ and ye shall find: a review of Marsh's 'students' evaluation of university teaching'" Instructional Evaluation 9(2), 19 - 27.

Abrami,P.C(1989a) "SEEQing the truth about student ratings of instruction" Educational Researcher 18(1), 43 - 45.

Abrami,P.C(1989b) "How should we use student ratings to evaluate teaching?" Research in Higher Education 30(2), 221 - 227.

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(3), 411 - 415.

Abrami,P.C,d'Apollonia,S and Cohen,P.A(1990) "Validity of student ratings of instruction: what we know and what we do not" Journal of Educational Psychology 82(2), 219 - 231.

Abrami,P.C,Leventhal,L and Perry,R.P(1982) "Educational seduction" Review of Educational Research 52(3), 446 - 464.

Aleamoni,L.M(1972) The Illinois Course Evaluation Questionnaire Urbana,Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois.

Aleamoni,L.M and Spencer,R.E(1973) "The Illinois Course Evaluation Questionnaire:A description of its development and a report of some of its results" Educational and Psychological Measurement 33, 669 - 684.

Aleamoni,L.M and Thomas,G.S(1977) "Is the instructor's rating of the class related to the class's rating of the instructor?"Research Report No.1, O.I.R.D. Tucson,Arizona:Office of Instructional Research and Development,University of Arizona.

Altbach,P.G and Wang,J(1989) Foreign Students and International Study:bibliography and analysis 1984-8 Boston: United Press of America.

Alwin,D.F(1974) "Approaches to the interpretation of relationships in the MTMM matrix" in H.L.Costner (ed) Sociological methodology 1973-4 San Francisco: Jossey-Bass, 79 - 105.

Amoh,K.O(1984) Newly Arrived Foreign Students at a US University: their adjustment difficulties and coping strategies Unpublished PhD dissertation, University of Minnesota.

Anderson,R.E(1973) "Consumer dissatisfaction:the effect of disconfirmed expectancy on perceived product performance" Journal of Marketing Research 10(Feb), 38 - 44.

Angulo,F,Fernandez,J and Martinez,M.R(1987) La Evaluacion de la Ensenanza Universitaria Madrid: Instituto de Ciencias de la Educacion.

Anikeef,A.M(1953) "Factors affecting student evaluation of college faculty members" Journal of Applied Psychology 37, 458 - 460.

Animasawun,G.K(1963) "African students in Britain" Race V(July), 38 - 47.

Arubayi,E.A(1987) "Improvement of instruction and teacher effectiveness: are student ratings reliable and valid?" Higher Education 16, 267 - 278.

Baldwin,J and Carter,B(1992) "Workshops and direct comment, an alternative to staff questionnaires" in Evaluating the quality of the student experience. Proceedings of conference held on 4/4/91. London:CNAA.

Ballard,M,Reardon,J and Nelson,J(1976)"Student and peer rating of faculty" Teaching of Psychology 3, 88 - 90.

Banjo,G(1975) "Cultural Castration" UKCOSA News 7, 14 - 18.

Barke,C.R,Tollefson,N and Tracy,D.B(1983) "Relationship between course entry attitudes and end-of-course ratings" Journal of Educational Psychology 75(1), 75 - 85.

Barnes,L.L.B and Barnes,M.W(1993) "Academic discipline and generalizability of student evaluations of instruction" Research in Higher Education 34(2), 135 - 149.

Barry,T.E,Gilly,M.C and Schucany,W.R(1982) "Students as consumers: predicting satisfaction" in An Assessment of Marketing Thought and Practice(1982 AMA Educators'Proceedings) B.Walker et al ed. Chicago: American Marketing Association, 109 - 112.

Bausell,R.B,Schwartz,S and Purohit,A(1975) "An examination of the conditions under which various student rating parameters replicate across time" Journal of Educational Measurement 12(4-Winter), 273 - 280.

Bearden,W.O and Teel,J.E(1983) "Selected determinants of consumer satisfaction and complaint reports" Journal of Marketing Research 20(Feb), 21 - 28.

Bendig,A.W(1953)"Student achievement in introductory psychology and student ratings of the competence and empathy of their instructors" Journal of Psychology 36, 427 - 433.

Benington, (1977) "Student evaluation of teaching in management and business studies" SCEDSIP Occasinal Paper No.2 T.Halseshaw ed. Newcastle-upon-Tyne:PETRAS, Newcastle Polytechnic.

Benton,S.E(1992)Rating College Teaching:Criterion validity studies of student evaluation of instructor instruments AAHE Research Report no.1. Washington:American Association for Higher Education.

Betz,E,Klingensmith,J and Menne,J(1970) "The measurement and analysis of college student satisfaction" Measurement and Evaluation in Guidance 3(2 - summer), 110 - 118.

Betz,E,Menne,J.W, Starr,A.M and Klingensmith,J.E(1971) "A dimensional analysis of college student satisfaction" Measurement and Evaluation in Guidance 4(2 - July), 99 - 106.

Biggs,J.B(1987) Student approaches to learning and studying Melbourne: Australian Council for Educational Research.

Bilbow,G.T(1989) "Towards an understanding of overseas students' difficulties in lectures: a phenomenographic approach" Journal of Further and Higher Education 13(3), 85 - 99.

Blackburn,R.T(1974)"The meaning of work in academia" in Assessing faculty effort J.I.Doi ed.(A special issue of New Directions for Institutional Research). San Francisco: Jossey Bass.

Blackburn,R.T and Clark,M.J(1975) "An assessment of faculty performance: some correlates between administrators, colleagues, students and self-ratings" Sociology of Education 48, 242 - 256.

Bock,R.D and Bargmann,R.E(1966) "Analysis of covariance structures" Psychometrika 31, 507 - 534.

Bowers,R(1978) "The background of students from the Indian sub-continent" in Study Modes and Academic Development of Overseas Students K.James ed,104 - 113.

Bradbury,P.S and Ramsden,P(1975) "Student evaluation of teaching at North East London Polytechnic" in Evaluating Teaching in Higher Education Collection of conference papers. London: University Teaching Management Unit, 94 - 101.

Bradley,D and Bradley,M(1984) Problems of Asian Students in Australia: Language, Culture and Education Canberra: Australian Government Publishing Service.

Braskamp,L,Brandenburg,D,Kohen,E,Ory,J and Mayberry,P() Guidebook for Evaluating Teaching Urbana,Illinois:University of Illinois.

Braskamp,L, Brandenburg,D.C and Ory,J.C(1985) Evaluating teaching efficiency: a practical guide Beverley Hills, CA:Sage.

Braskamp,L, Caulley,D.N and Costin,F(1979) "Student ratings and instructor self-ratings and their relationship to student achievement" American Educational Research Journal 16, 295 - 306.

Braskamp,L, Ory,J.C and Pieper,M(1981) "Student written comments: dimensions and instructional quality" Journal of Educational Psychology 73, 65 - 70.

Bridge,R.G(1974) Non-Response Bias in Mail Surveys Defense Advanced Research Projects Agency R-1501. Santa Monica,CA: Rand.

Brown,A.L,Campione,J.C and Day,J.D(1981) "Learning to learn: on training students to learn from texts" Educational Researcher 10, 14 - 21.

Brown,D.L(1976)"Faculty ratings and student grades:a university-wide multiple regression analysis" Journal of Educational Psychology 68(5), 573 - 578.

Bryant,P.T(1967) "By their fruits ye shall know them" Journal of Higher Education 38, 326 - 330.

Burgoon,J.K and Hale,J.L(1988) "Nonverbal expectancy violations:more elaboration and application to immediacy behaviours" Communication Monongraphs 55, 58 - 79.

Burns,D(1965)ed Travelling Scholars:an enquiry into the adjustment and attitudes of overseas students holding Commonwealth bursaries in England and Wales Slough:National Foundation for Educational Research.

Byrne,D and Clore,G.L(1970) "A reinforcement model of evaluative responses" Personality:An International Journal 1, 103 - 128.

Byrne,D and Nelson,D(1965) "Attraction as a linear function of proportion of positive reinforcements" Journal of Personality and Social Psychology 1, 659 - 663.

Cadotte,E.R,Woodruff,R.B and Jenkins,R.L(1987) "Expectations and norms in models of consumer satisfaction" Journal of Marketing Research 24(Aug), 305 - 314.

Cadwell,J and Jenkins,J(1985)"Effects of the semantic similarity of items on SRI" Journal of Educational Psychology 77(4), 383 - 393.

Campbell,A(1981) The Sense of Well-Being in America New York: McGraw-Hill.

Campbell,D.T and Fiske,D.W(1959) "Convergent and discriminant validation by the MTMM matrix" Psychological Bulletin 56, 81 - 105.

Carrier,N.A,Howard,G.S and Miller,W.G(1974)"Course evaluations:When?" Journal of Educational Psychology 66, 609 - 613.

Cashin,W.E(1988) Student ratings of teaching: a summary of the research Center for Faculty Evaluation and Development: Kansas State University.

Cashin,W.E(1990) "Students do rate different academic fields differently" New Directions for Teaching and Learning 43(Fall), 113 - 121.

Cashin,W.E and Downey, R.G(1992) "Using global student rating items for summative evaluation" Journal of Educational Psychology 84(4), 563 - 572.

Cardozo,R.N(1965) "An experimental study of customer effort, expectation and satisfaction" Journal of Marketing Research 2(Aug), 244 - 249.

Centra,J.A(1972)Two studies on the utility of student ratings for improving teaching SIR Report no.2. Princeton, New Jersey:Educational Testing Service.

Centra,J.A(1974)"The relationship between student and alumni ratings of teachers" Educational and Psychological Measurement 34, 321 - 325.

Centra,J.A(1975)"Colleagues as raters of classroom instruction" Journal of Higher Education 46, 327 - 337.

Centra,J.A(1980) Determining faculty performance San Francisco:Jossey-Bass.

Centra,J.A and Creech,F.R(1976) The relationship between student, teacher, and course characteristics and student ratings of teaching effectiveness PR-76-1, Princeton,NJ:Educational Testing Service.

Chambers,E(1992) "Improving foundation level study at OU through evaluation of student experience" in Evaluating the quality of the student experience. Proceedings of conference held on 4/4/91. London:CNAA.

Channell,J(1990) "The student-tutor relationship" in The Learning Experience of Overseas Students 63 - 81.

Chermesh,R(1978) "Instructional evaluation by students: its use for diagnosis and improvement of teaching" Educational Technology 18, 9 - 13.

Churchill jr,G.A and Surprenant,C(1982) "An investigation into the determinants of consumer satisfaction" Journal of Marketing Research 19(Nov), 491 - 504.

Clabaugh,M.G,Mason,J.B and Bearden,W.O(1979) "Consumer alienation and casual attribution as moderators of consumer satisfaction/dissatisfaction and complaint behaviour" in New Dimensions of Consumer Satisfaction and Complaining Behaviour R.L.Day and H.K.Hunt eds. Bloomington,Illinois:Indiana University.

Clark,K.E and Keller,R.J(1954)"Student ratings of college teaching" in Eckert,R et al(eds) A University looks at its programme Minneapolis: University of Minneapolis.

Clarkson,P.C(1984) "Papua New Guinean students' perceptions of mathematics lecturers" Journal of Educational Psychology 76(6), 1386 - 1395.

Clift,J,Hall,C and Turner,I(1989) "Establishing the validity of a set of summative teaching performance scales" Assessment and Evaluation in Higher Education 14(3 - Autumn), 193 - 206.

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 EducationalResearch 51(3), 281 - 309.

Comry,A.L and Lee,H.B(1992) A first course in factor analysis(2nd ed) Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooper,B and Foy,J.M(1967) "Evaluating the effectiveness of lectures" Universities Quarterly 21(2), 182 - 185.

Costin,F(1968)"A graduate course in the teaching of psychology:description and evaluation" Journal of Teacher Education 19, 425 - 432.

Costin,F,Greenough,W.T and Menges,R.J(1971)"Student ratings of college teaching:reliability, validity and usefulness" Review of Educational Research 41, 511 - 535.

Cotsonas,N,J jr and Kaiser,H.F(1962) "A factor analysis of students' and administrators' ratings of clinical teachers in a medical school" Journal of Educational Psychology 53, 219 -223.

Crannell,C.W(1948)"An experiment in the rating of instructors by their students" College and University 24, 5 - 11.

Cranton,P and Smith,R.A(1990) "Recognising the unit of analysis: a model of student ratings of instruction" Journal of Educational Psychology 82(2), 207 - 212.

Crittenden,K.S and Norr,J.L(1973) "Student values and teacher evaluation: a problem in person perception" Sociometry 36(2), 143 - 151.

Cronbach,L.J(1951) "Coefficient alpha and the internal structure of tests" Psychometrika 16, 297 - 334.

Crooks,T.J and Kane,M.J(1981) "The generalizability of student ratings of instructors: item specificity and section effects" Research in Higher Education 15, 305 - 313.

Dawson,J.A and Caulley,D.M(1981) "The group interview as an evaluation technique in higher education" Educational Evaluation and Policy Analysis 3, 61 - 66.

Day,R.L(1977) "Extending the concept on consumer satisfaction" in Advances in Consumer Research Vol IV W.D.Perrault jr ed. Ann Arbor:Association for Consumer Research, 149 - 154.

Dehlavi,N.S(1987) "Iranian students' perception of ideal teacher" Perceptual and Motor Skills 64, 143 - 146.

De Neve,H.M.F and Janssen,P.J(1982) "Validity of student evaluation of instruction" Higher Education 11, 543 - 552.

Downie,N.W(1952)"Student evaluation of faculty" Journal of Higher Education 23, 495 - 496,503.

Doyle,K.O jr(1975)Student Evaluation of Instruction Lexington,Mass:Lexington Books.

Doyle,K.O jr(1983)Evaluating Teaching Lexington,Mass:Lexington Books.

Doyle,K.O jr and Crichton,L.A(1978) "Student, peer and self-evaluation of college instruction" Journal of Educational Psychology 70, 815 - 826.

Doyle,K.O jr and Webber,P.C(1978) "Self-ratings of college instruction" American Educational Research Journal 15, 467 - 476.

Doyle,K.O jr and Whitely,S(1974)"Student ratings as criteria of effective teaching" American Educational Research Journal 11, 259 - 274.

Drucker,A.J and Remmers,H.H(1951) "Do alumni and students differ in their attitudes toward instructors?" Journal of Educational Psychology 42(3),129 -143.

Dunkin,M.J and Barnes,J(1986) "Research on teaching in higher education" in Handbook of Research on Teaching, M.Wittrock ed (3rd edition). London: MacMillan, 754 - 777.

Dunlop,F(1966) Europe's Guests, Students and Trainees: a survey on the welfare of foreign students and trainees in Europe Strasbourg: Council for Cultural Co-Operation of Council of Europe.

Ebel,R.L(1941) "Estimation of the relationship of ratings" Psychometrika 16, 407 - 424.

Echandia,P.P(1964)"A methodological study and factor analytic validation of forced choice performance ratings of college accounting instructors" Dissertation Abstracts 25(4), 2605 - 2606.

Elliott,C.K(1969)"Longitudinal use of a student-constructed teacher evaluation form" British Journal of Educational Psychology 39(3), 309 - 313.

Elsey,B(1990) "Teaching and Learning" in The Learning Experiences of Overseas Students M.Kinnell ed,46 - 62.

Elsey,B and Kinnell,M(1990) "Introduction" in The Learning Experiences of Overseas Students M.Kinnell ed, 1 - 11.

Entwistle,N and Ramsden,P(1983) Understanding Student Learning London:Croom Helm.

Entwistle,N and Tait,H(1990) "Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments" Higher Education 19, 169 - 194.

Erdle,S and Murray,H.G(1986)"Interfaculty differences in classroom teaching behaviours and their relationship to student instructional ratings" Research in Higher Education 24(2), 115 - 127.

Eysenck,H.J and Eysenck,S.B.G(1964) Manual of Eysenck Personality Inventory London:University of London Press.

Feldman,K.A(1976a) "Grades and college students' evaluations of their courses and teachers" Research in Higher Education 4, 69 - 111.

Feldman,K.A(1976b)"The superior college teacher from the students' view" Research in Higher Education 5, 243 - 288.

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 know" 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(2), 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(1), 3 - 124.

Feldman,K.A(1984)"Class size and college students' evaluations of teachers and courses: a closer look" Research in Higher Education 21(1), 45 - 117.

Feldman,K.A(1987)"Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: a review and exploration" Research in Higher Education 26(3), 227 - 298.

Feldman,K.A(1988)"Effective college teaching from the students' and faculty's view: matched or mismatched priorities?" Research in Higher Education 28(4), 291 - 344.

Feldman,K.A(1989a)"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(2), 137 - 194.

Feldman,K.A(1989b)"The association between student ratings of specific instructional dimensions and student achievement: refining and extending the synthesis of data from multisection validity studies" Research in Higher Education 30(6), 583 - 646.

Feldman,K.A(1992)"College students' views of male and female college teachers: Part 1 - evidence from the social laboratory and experiments" Research in Higher Education 33(3), 317 - 375.

Feldman,K.A(1993)"College students' views of male and female college teachers: part 2 - evidence from students' evaluations of their classroom teachers" Research in Higher Education 34(2), 151 - 212.

Feldman,K.A and Newcomb,T.M(1969a)Impact of College on Students Vol I San Francisco:Jossey-Bass.

Feldman,K.A and Newcomb,T.M(1969b)Impact of College on Students Vol II San Francisco:Jossey-Bass.

Feldman,R.S and Theiss,A.J(1982) "The teacher and student as pygmalions: joint effects of teacher and student expectations" Journal of Educational Psychology 74(2), 217 - 223.

Fernandez,J and Mateo,M.A(1992) "Student evaluation of university teaching quality: analysis of a questionnaire for a sample of university students in Spain" Educational and Psychological Measurement 52(3), 675 - 686.

Flood Page,C(1974) Student evaluation of teaching: the American experience London: SRHE.

Follman,J and Phillippy,S(1988) "Orthogonal vs oblique rotation in students' ratings" Psychological Reports 63, 678.

Follman,J,Lucoff,M,Small,L and Power,F(1974)"Kinds of keys of student ratings of faculty teaching effectiveness" Research in Higher Education 2, 173 - 179.

Foundation for Research on Human Behaviour(1956) Training Foreign Nationals in the US Ann Arbor: FRHB.

Foy,J.M(1969) "A note on lecturer evaluation by students" Universities Quarterly 23(3), 345 - 349.

Freehill,M.F(1967) "Authoritarian bias and evaluation of college experiences" Improving College and University Teaching 15, 18 - 19.

Frey,P.W(1973) "Student ratings of teaching: validity of several rating factors" Science 182, 83 - 85.

Frey,P.W(1976) "Validity of student instructional ratings: does timing matter?" Journal of Higher Education 67, 327 - 336.

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 on instruction: validation research" American Educational Research Journal 12, 435 - 447.

Frisbie,D.A and Brandenburg,D.C(1979)"Equivalence of questionnaire items with varying response formats" Journal of Educational Measurement 16, 43 - 48.

Fuller,F.F and Manning,B.A(1973) "Self-confrontation review: a conceptualization for video playback in teacher education" Review of Educational Research 43, 469 - 528.

Gage,N.L(1961)"The appraisal of college teaching" Journal of Higher Education 32, 17 - 22.

Gagne,F and Allaire,D(1974) Summary of research data on the reliability and validity of a measure of dissatisfaction derived from Reality-Desires discrepancies. Quebec:Institute National de la Recherche Scientifique, Universite du Quebec.

Gardner,R.C and MacIntyre,P.D(1992) "A student's contribution to second language learning. Part 1: Cognitive variables" Language Teaching 215(4 - Oct), 211 - 220.

Gardner,R.C and MacIntyre,P.D(1993) "A student's contribution to second language learning. Part II:Affective variables" Language Teaching 26, 1 - 11.

Gaski,J.F(1987)"On 'construct validity of measures of college teaching effectiveness'" Journal of Educational Psychology 79(3), 326 - 330.

Gessner,P.K(1973)"Evaluation of instruction" Science 180, 566 - 570.

Gillie,B(1984) "Being an overseas student" in Student Experience of Higher Education I.Lewis ed. Beckenham, Kent: Croom Helm, .

Glaser,R(1962) "Psychology and instructional technology" in R.Glaser, Training, Research and Education New York: John Wiley.

Glass,G.V(1976) "Primary,secondary and meta-analysis of research" Educational Researcher 5, 3 - 8.

Glass,G.V(1978) "Integrating findings:the meta-analysis of research" in Review of Research in Education (Vol 5), L.S.Shulman ed. Itasca,Illinois: F.E.Peacock, .

Goldsmith,J and Shawcross,V(1985) It Ain't Half Sexist, Mum London: World University Service.

Guilford,J.P(1954) Psychometric methods (2nd ed) New York: McGraw-Hill

Guthrie,E.R(1927) "Measuring student opinion of teachers" School and Society 25, 175 - 176.

Guthrie,E.R(1949)"The evaluation of teaching" Educational Record 30, 109 - 115.

Guthrie,E.R(1954) The evaluation of teaching: a progress report Seattle:University of Washington.

Hall,D.J(1986) "Satisfaction and consumer services" in Advances in Consumer Research Vol XIII R.J.Lutz ed. Ann Arbor: Association for Consumer Research, 311 - 315.

Hampton,G.M(1983) "College student satisfaction:marketing's approach for evaluating higher education" in 1983 AMA Educators'Proceedings P.Murphy et al eds. Chicago: American Marketing Association, 169 - 173.

Harnash-Glazer,M and Meyer,J(1991) "Dimensions of satisfaction with college education" Assessment and Evaluation in Higher Education 16(2 - summer), 95 - 107.

Harvey,L and Green,D(1993) "Defining quality" Assessment and Evaluation in Higher Education 18(1), 9 - 34.

Hatcher,L,Kryter,K,Prus,J.S and Fitzgerald,V(1992) "Predicting college student satisfaction, commitment, and attrition from investment model constructs" Journal of Applied Social Psychology 22(16), 1273 - 1296.

Hawes,J.M and Glisan,G(1983) "A marketing approach to student evaluation of a department of marketing" in 1983 AMA Educators' Proceedings P.Murphy et al eds. Chicago: American Marketing Association, 159 - 163.

Hawkey,R and Nakornchai,C(1978) "Thai students studying" in Study Modes and Academic Development of Overseas Students K.James ed,70 - 78.

Hayton,G.E(1983) An investigation of the applicability in technical and further education of a student evaluation of teaching instrument Unpublished Master's thesis, University of Sydney, Australia.

Hildebrand,M and Wilson,R.C(1970) Effective university teaching and its evaluation Berkeley Centre for Research and Development in Higher Education: University of California.

Hogan,T.P(1973)"Similarity of student ratings across instructors, courses, and time" Research in Higher Education 1, 149 - 154.

Holes,C.D(1972) An Investigation into some aspects of the English Language problems of 2 groups of overseas postgraduate students at Birmingham University Unpublished MA Thesis, University of Birmingham.

Holland,J.L(1973) Making Vocational Choices New Jersey: Prentice-Hall.

Holmes,D.S(1972)"Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor" Educational and Psychological Measurement 31, 951 - 957.

Horst,A.P(1949) "A generalised expression for the reliability of measures" Psychometrika 14, 21 - 32.

Howard,G.S and Maxwell,S.E(1980) "Correlation between student satisfation and grades: a case of mistaken causation?" Journal of Educational Psychology 72, 810 - 820.

Howard,G.S,Conway,C.G and Maxwell,S.E(1985)"Construct validity of measures of college teaching effectiveness" Journal of Educational Psychology 77(2),187 - 196.

Howard,J.A and Sheth,J(1967) The Theory of Buyer Behaviour New York: John Wiley and Sons Inc.

Hoyt,D.P(1973) "Measurement of instructional effectiveness" Research in Higher Education 1, 367 - 378.

Jackson,D.N(1969) "Multimethod factor analysis in the evaluation of convergent and discriminant validity" Psychological Bulletin 72, 30 - 49.

James,K(1978) "Seminar overview" in Study Modes and Academic Development of Overseas Students K.James ed, 7 - 21.

James,K(1978) ed Study Modes and Academic Development of Overseas Students (ELT Documents 109). London:British Council.

Jenkins,J(1987)"Implicit theories and semantic similarities: reply to Marsh and Groves(1987)" Journal of Educational Psychology 79(4), 490 - 493.

Jennings,C.M and Bartling,C.A(1991) "Quality education:what college students think" College Student Journal 25(1 - March), 532 - 536.

Jones,J(1989) "Students' ratings of teacher personality and teaching competence" Higher Education 18, 551 - 558.

Jones,J,Lennie,M and Robinson,A(1985) "Students'expectations of good teachers: primary, secondary and tertiary views" Paper presented at New Zealand Association of Research in Education, Auckland, New Zealand.

Kane,M.T,Crooks,T.J and Gillmore,G.M(1976) Generalizability and the interpretation of student evaluations of teachers Paper presented at Annual Meeting of American Educational Research Association, San Francisco (March).

Kappes,M.M(1988) A comparison of students' ratings of full-time and part-time instructors' teaching effectiveness in a community college Unpublished PHd thesis, University of Pittsburgh.

Kavanagh,M.J, MacKinney,A.C and Wolins,L(1971) "Issues in managerial performance: MTMM analyses of ratings" Psychological Bulletin 75, 34 - 49.

Kelly,G.A(1955) The Psychology of Personal Constructs New York: Norton.

Kelley,H.H(1950)"The warm-cold variable in first impressions of persons" Journal of Personality 18, 431 - 439.

Kember,D and Gow,L(1991) "A challenge to the anecdotal stereotype of the Asian student" Studies in Higher Education 16(2), 117 - 128.

Kempa,R.F and Diaz,M.M(1990a)"Motivational traits and preference for different instructional modes in science. Part 1: students' motivational traits" International Journal of Science Education 12(2), 195 - 203.

Kempa,R.F and Diaz,M.M(1991b) "Students' motivational traits and preferences for different instructional modes in science education - Part 2" International Journal of Science Education 12(2), 205 - 216.

Kendall,M and Hawker,A(1968) Overseas Students in Britain; an annotated bibliography London: United Kingdom Council for Overseas Student Affairs/Research Unit for Student Problems(University of London).

Kenny,D.A(1979) Correlation and causality New York: Wiley.

Kenny,D.A and Kashy,D.A(1992) "Analysis of the MTMM matrix by confirmatory factor analysis" Psychological Bulletin 112(1), 165 - 172.

Kenny,G.K(1986) "The metric properties of rating scales employed in evaluation of research: an empirical examination" Evaluation Review 10(June), 397 - 408.

Kierstead,D,D'Agostino,P and Dill,H(1988)"Sex role stereotyping of college professors: bias in students' ratings of instructors" Journal of Educational Psychology 80(3), 342 - 344.

Kinnell,M(1989) "International marketing in UK higher education: some issues in relation to marketing educational programmes to overseas students" European Journal of Marketing 23(5), 7 - 21.

Kinnell,M(1990) ed The Learning Experience of Overseas Students. Buckingham: Society for Research in Higher Education.

Kirby,S(1992) Working for a Research Degree at Aston:assessing the quality of the experience Unpublished report, University of Aston.

Kirchner,R.P(1967) A central factor in teacher evaluation by students Unpublished research paper, College of Education: Lexington University of Kentucky.

Kirker,M.J(1990)Variance in student ratings of part-time and full-time instructor effectiveness by teaching field and function at a Mid-Western community college Unpublished PHd thesis, Iowa State University.

Klyczek,J.P(1989) A study of faculty characteristics affecting student evaluation of instruction in higher education Unpublished PHd thesis, State University of New York at Buffalo.

Koermer,C.D and Petelle,J.L(1991) "Expectancy violation and student rating of instruction" Communication Quarterly 39(4), 341 - 350.

Kremer,J.F(1990) "Construct validity of multiple measures in teaching, research, and service and relationship of peer ratings" Journal of Educational Psychology 82(2), 213 - 218.

Kuder,G,F and Richardson,M.W(1937) "The theory of the estimation of test reliability" Psychometrika 2, 151 - 160.

Kulik,J.A(1991) Student ratings of instruction Center for Research on Learning and Teaching Occasinal Paper no.4. Ann Arbor:CRLT, University of Michigan.

LaBarbera,P.A and Mazursky,D(1983) "A longitudinal assessment of consumer satisfaction/dissatisfaction: the dynamic aspect of the cognitive process" Journal of Marketing Research 20(Nov), 393 - 404.

Lago,C(1990) Working with Overseas Students London: British Council.

Larson,J.R jr(1979)"The limited utility of factor analytic techniques for the study of implicit theories in student ratings of teacher behaviour" American Educational Research Journal 16(2), 201 - 211.

LaTour,S.A and Peat,N.C(1979) "Conceptual and methodological issues in consumer satisfaction research" in Advances in Consumer Research Vol VI W.L.Wilkie ed. Ann Arbor: Association for Consumer Research, 431 - 437.

Lee,M.Y(1984) "judgments of significant others and self-concept of students from developing nations" Journal of Social Psychology 122, 127 - 134.

Lee-Warner,Sir W(1922) Report of the Committee on Indian Students, 1907. Appendix IV of the Lytton Committee Report, Part 1 London: HMSO for the India Office.

L'Hommedieu,R,Menges,R.J and Brinko,K.T(1990) "Methodological explanations for the modest effects of feedback" Journal of Educational Psychology 82, 232 - 241.

Livingstone,A(1960) The Overseas Student in Britain Manchester: Manchester University Press.

Lundstrom,W.J and Lamont,L.M(1976) "The development of a scale to measure consumer discontent" Journal of Marketing Research 13(Nov), 373 - 381.

McBean,E.A and Al-Nassri,S(1982) "Questionnaire design for student measurement of teaching effectiveness" Higher Education 11, 273 - 288.

McCallum,L.W(1984) "a meta-analysis of course evaluation data and its use in the tenure decision" Research in Higher Education 21(2), 150 - 157.

McClelland,J.N(1970) "The effect of student evaluations of college instruction upon subsequent evaluations" California Journal of Educational Research 21, 88 -95.

McClendon,M.J and O'Brien,D.J(1988) "Question order effects on the determinants of subjective well-being" Public Opinion Quarterly 52, 351 - 364.

McDaniel,E.D and Feldhusen,J.F(1970) "Relationships between faculty ratings and indexes of service and scholarship" Proceedings of the 78th Annual Convention of the American Psychological Association 5, 619 - 620.

McKnight,P(1973) Curriculum and Instruction Survey University of Kansas: Office of Instructional Resources.

Madden,C.S and Smart,D.T(1983) "Alternatives to traditional faculty evaluation measures: a comparative analysis" in 1983 AMA Educators Proceedings P.Murphy et al eds. Chicago:American Marketing Association, 164 - 168.

Maria,J and Norusis,T(1988) SPSS-PC Advanced Statistics Module II New York:SPSS Inc.

Marsh,H.W(1977) "The validity of students' evaluations: classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors" American Educational Research Journal 14(4 - Fall), 441 - 447.

Marsh,H.W(1980) "The influence of student, course and instructor characteristics in evaluations of teaching" American Educational Research Journal 17, 219 - 237.

Marsh,H.W(1981) "Students' evaluations of tertiary instruction:testing the applicability of American surveys in an Australian setting" Australian Journal of Education 25(2), 177 - 193.

Marsh,H.W(1982a) "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(1982b) "Factors affecting students' evaluations of the same course taught by the same instructor on different occasions" American Educational Research Journal 19, 485 - 497.

Marsh,H.W(1982c)"Validity of student evaluation of college teaching: a MTMM analysis" Journal of Educational Psychology 74, 264 - 279.

Marsh,H.W(1983) "Multi-dimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics" Journal of Educational Psychology 75(1), 150 - 166.

Marsh,H.W(1984) "Students' evaluations of university teaching: dimensionality, reliability,validity,potential biases, and utility" Journal of Educational Psychology 76(5), 707 - 754.

Marsh,H.W(1985) "Student ratings of teaching" in International Encyclopedia of Education:Research and Studies (Vol 8) T.Husen and T.N.Postlethwaite eds. Oxford:Pergamon Press, 4867 - 4873.

Marsh,H.W(1986)"Applicability paradigm:students' evaluations of teaching effectiveness in different countries" Journal of Educational Psychology 78(6), 465 - 473.

Marsh,H.W(1987) "Students' evaluations of university teaching:research findings, methodological issues, and directions to future research" International Journal of Educational Research 11, 253 - 388.

Marsh,H.W(1991a) "Multidimensional students' evaluations of teaching effectiveness: a test of alternative higher-order structures" Journal ofEducational Psychology 82(2), 285 - 296.

Marsh,H.W(1991b) "A multidimensional perspective on students' evaluations of teaching effectiveness: reply to Abrami and d'Apollonia(1991)" Journal of Educational Psychology 83(3), 416 -421.

Marsh,H.W and Bailey,M(1993)"Multidimensional students' evaluations of teaching effectiveness: a profile analysis" Journal of Higher Education 64(1), 1 - 18.

Marsh,H.W and Cooper,T.L(1981) "Prior subject interest, students' evaluations, and instructional effectiveness" Multivariate Behavioral Research 16, 83 - 104.

Marsh,H.W and Dunkin,M.J(1992) "Students' evaluations of university teaching: a multidimensional perspective" Higher Education: Handbook of Theory and Research Vol III ed J.C.Smart; New York:Agathon Press.

Marsh,H.W,Fleiner,H, and Thomas,C.S(1975)"Validity and usefulness of student evaluations of instructional quality" Journal of Educational Psychology 67, 833 - 839.

Marsh,H.W and Groves,M.A(1987)"Students' evaluations of teaching effectiveness and implicit theories: a critique of Cadwell and Jenkins(1985)" Journal of Educational Psychology 79(4), 483 - 489.

Marsh,H.W and Hocevar,D(1983) "Confirmatory factor analysis of MTMM matrices" Journal of Educational Measurement 20(3), 231 - 248.

Marsh,H.W and Hocevar,D(1991)"The multidimensionality of students' evaluations of teaching effectiveness: the generality of factor structures across academic disciplines, instructor level, and course level" Teaching and Teacher Education 7(1), 9 - 18.

Marsh,H.W and Overall,J.U(1979a) "Validity of student evaluations of teaching: a comparison o f instructor self-evaluations by teaching assistants,undergraduate faculty and graduate faculty" Paper presented at theAnnual Meeting of the American Educational Research Association San Francisco(ERIC Document Reproduction Service No ED177 205).

Marsh,H.W and Overall,J.U(1979b) "Long term stability of students' evaluations: a note on Feldman's 'consistency and variability among college students in rating their teachers and courses'" Research in Higher Education 10, 139 - 147.

Marsh,H.W and Overall,J.U(1980) "Validity of students' evaluations of teaching effectiveness: cognitive and affective criteria" Journal of Educational Psychology 72(4), 468 - 475.

Marsh,H.W and Overall,J.U(1981) "The relative influence of course level, course type and instructor on students' evaluations of college teaching" American Educational Research Journal 18(1 - Spring), 103 - 112.

Marsh,H.W, Overall,J.U and Kesler,S.P(1979a) "Validity of student evaluations of instructor efficiency: a comparison of faculty self evaluations and evaluations by their students" Journal of Educational Psychology 71, 149 - 160.

Marsh,H.W,Overall,J.U and Kesler,S.P(1979b)"Class size, students' evaluations, and instructional effectiveness" American Educational Research Journal 16(1), 57 - 70.

Marsh,H.W and Roche,L(1992)"the use of student evaluations of university teaching in different settings: the applicability paradigm" Australian Journal of Education 36(3), 278 - 300.

Marsh,H.W and Roche,L(1993)"The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness" American Educational Research Journal 30(1), 217 - 251.

Marsh,H.W,Touron,J and Wheeler,B(1985)"Students' evaluations of university instructors: the applicability of American instruments in a Spanish setting" Teaching and Teacher Education 1(2), 123 - 138.

Martin,E(1984)"Power and authority in the classroom: sexist stereotyping in teaching evaluations" Journal of Women in Culture and Society 9, 482 - 492.

Masters,J.R(1974)"The relationship between number of response categories and reliability of Likert-type questionnaires" Journal of Educational Measurement 11, 49 - 53.

Mazelan,P,Brannigan,C,Green,D and Tormey,P(1992) Report on the 1992 survey of student satisfaction with their educational experience at UCE. Student Satisfaction Research Unit, University of Central England in Birmingham.

Mickle,K(1984) The Cross-cultural Adaptation of Hong Kong Students at two Ontario Universities Unpublished PhD dissertation, University of Toronto.

Mickle,K and Chan,R(1986) The Cross-cultural Adaptation of Hong Kong Chinese Students at Canadian Universities Ottawa: Canadian Bureau for International Education.

Miklich,D.R(1969) "An experimental validation study of the Purdue Rating Scale for Instruction" Educational and Psychological Measurement 29, 963 - 967.

Miller,A(1991) "Personality types, learning styles, and educational goals" Educational Psychology 11(3+4 double issue), 217 - 238.

Miller,D.C(1983) Handbook of Research Design and Social Measurement 4th ed. New York: Longmans.

Miller,D.I and Topping,J.S(1991) "Factors of ability, communication, grievances, and personal optimism as predictors of student satisfaction, involvement, and alienation: an ecological dissonance interpretation" Bulletin of the Psychonomic Society 29(1), 19 - 20.

Miller,J.A(1976) "Exploring some alternative measures of consumer satisfaction" in Marketing:1776-1976 and Beyond(1976 AMA Educators' Proceedings) K.L.Bernhardt ed. Chicago: American Marketing Association, 661 - 664.

Miller,J.A(1977) "Exploring satisfaction, modifying models, eliciting expectations, posing problems and making meaningful measurements" in Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction H.K.Hunt ed. Cambridge,MA: Marketing Science Institute,(May) 72 - 91.

Miller,M.T(1971)"Instructor attitudes towards and their use of, student ratings of teachers" Journal of Educational Psychology 62, 235 - 239.

Miller,R.I(1972) Evaluating Faculty Performance San Francisco: Jossey-Bass.

Miller,S(1984) Experimental Design and Statistics 2nd ed. London: Methuen

Monette,D.R, Sullivan,T.J and DeJong,C.R(1990) Applied Social Research 2nd ed. Orlando,FL: Holt, Rinehart and Winston.

Morris,B.S(1967) International Community? London: National Union of Students.

Morrow jr,J.M(1971) "A test of Holland's theory of vocational choice" Journal of Counseling Psychology 18(5), 422 - 425.

Morstain,B.R(1977) "An analysis of students' satisfaction with thei academic program" Journal of Higher Education XLVIII(1), 1 - 16.

Mueller,F.J(1951)"Trends in student ratings of faculty" American Association of University Professors' Bulletin 37, 319 - 324.

Murphy,K.R and Davidshofer,C.O(1988)Psychological testing:principles and applications Englewood Cliffs, NJ: Prentice Hall.

Murray,H.G(1980) Evaluating University Teaching:A Review of Research Toronto, Canada: Ontario Confederation of University Faculty Associations.

Naftulin,D.H, Ware,J.E and Donnelly, F.A (1973) "The Doctor Fox lecture: a paradigm of educational seduction" Journal of Medical Education 48, 630 - 635.

Nafziger,D.H,Holland,J.L and Gottfredson,G.D(1975) "Student-college congruency as a predictor of satisfaction" Journal of Counseling Psychology 22(2), 132 - 139.

Neumann,Y and Neumann,E(1981) "Determinants of student satisfaction with course work: an international comparison between 2 universities" Research in Higher Education 14, 321 - 333.

Neumann,Y and Finaly-Neumann,E(1989) "An organisational behaviour model of students' evaluation of instruction" Higher Education 18, 227 - 238.

Noesjirwan,J(1970) "Attitudes to learning of the Asian student studying in the West" Journal of Cross-cultural Psychology 1(4), 393 - 397.

Oliver,R.L(1977) "Effect of expectations and disconfirmation on post exposure product evaluations: an alternative interpretation" Journal of Applied Psychology 62(4), 480 - 486.

Oliver,R.L(1980) "A cognitive model of the antecedents and consequences of satisfaction decisions" Journal of Marketing Research 17(Nov), 460 -469.

Oliver,R.L and Winer,R.S(1987)"A framework for the formation and structure of consumer expectations: review and propositions" Journal of Economic Psychology 8, 469 - 499.

O'Neil,M and Pennington,G(1992)"Evaluating teaching and courses from an active learning perspective" Part 1 Effective Learning and Teaching in Higher Education Module 12 University Staff Development and Training Unit: Sheffield.

Ortinau,D.J(1979) "A conceptual model of consumers' post purchase satisfaction/dissatisfaction decision process" in New Dimensions of Consumer Satisfaction and Complaining Behaviour R.L.Day and H.K.Hunt eds. Bloomington,Indiana:University of Indiana, 35 - 40.

Ortinau,D.J and Anderson,R.L(1986) "College students' post-purchase educational satisfaction/dissatisfaction decision process: a conceptual model" in 1986 AMA Educators' Proceedings T.A.Shimp et al eds. Chicago: American Marketing Association, 120 - 125.

Ortinau,D.J, Anderson,R.L, and Klippel,R.E(1987) "The impact of students'involvement and expectancy beliefs on course/faculty evaluations" in 1987 AMA Educators'Proceedings, S.Douglas and M.Solomon eds. Chicago: American Marketing Association, 266 - 271.

Ory,J.C(1982)"Item placement and wording effects on overall ratings" Educational and Psychological Measurement 42, 767 - 775.

Ory,J.C(1990)"Student ratings of instruction: issues for improving practice" New Directions for Teaching and Learning 43(Fall), 63 - 74.

Ory,J.C and Braskamp,L.A(1981)"Faculty perceptions of the quality and usefulness of 3 types of evaluation information" Research in Higher Education 15, 271 - 282.

Ory,J.C, Braskamp,L.A and Pieper,D.M(1980) "Congruency of student evaluative information collected by 3 methods" Journal of Educational Psychology 72, 181 - 185.

O'Shea,J(1992) "Evaluating the educational experience of students" in Evaluating the Quality of the Student Experience Proceedings of conference held on 4/4/91. London: Council for National Academic Awards, 63 - 66.

Overall,J.U and Marsh,H.W(1979) "Midterm feedback from students: its relationship to instructional improvement and students' cognitive and affective outcomes" Journal of Educational Psychology 71, 856 - 865.

Overall,J.U and Marsh,H.W(1980) "Students' evaluation of instruction: a longitudinal study of their stability" Journal of Educational Psychology 72(3), 321 - 325.

Palihawadana,D(1992) "Expectancy-performance paradigm in student evaluations as a tool in quality assurance and marketing" in Evaluating the Quality of the Student Experience, Proceedings of conference held on 4/4/91. London: Council for National Academic Awards, 67 - 74.

Panney,G(1977) An experiment to investigate possible instructor-induced item response bias in the Purdue Cafeteria Instructional Evaluation System Paper presented at the 7th annual South East American Institute for Decision Sciences, Alabama.

Partington,P et al Student feedback - Context, Issues and Practice University Staff Development Unit: Sheffield.

Peterson,R.A and Wilson,W.R(1992) "Measuring customer satisfaction: fact or artifact" Journal of the Academy of Marketing Science 20(1), 61 - 71.

Pike,G.R(1993)"The relationship between perceived learning and satisfaction with college: an alternative view" Research in Higher Education 34(1), 23 - 40.

Pohlmann,J.T(1972) Summary of research on the relationship between student characteristics and student evaluations of instruction at Southern Illinois University, Carbondale Technical Report 1.1 - 72. Carbondale, Illinois: Counselling and Testing Centre, Southern Illinois University.

Political and Economic Planning(1954) "Students from the colonies" Planning 20(374 - 29/11), 281 - 300.

Political and Economic Planning(1965) New Commonwealth Students in Britain (with special reference to students from East Africa) London:George Allen.

Powell,R.W(1977) "Grades, leniency and student evaluation of instruction" Research in Higher Education 7, 193 - 205.

Pratt,M and Pratt,T.A.E.C(1976)"A study of the student- teacher grading interaction process" Improving College and University Teaching 24, 73 - 80.

Prosser,M and Trigwell,K(1990) "Student evaluations of teaching and courses:student study strategies as a criterion of validity" HigherEducation 20, 135 - 142.

Ramsden,P(1975) "Polytechnic students' expectations of their teachers and the use of a student feedback questionnaire: a preliminary report" Higher Education Bulletin 3(2), 73 - 85.

Reed,B,Hutton,J and Bazalgette,J(1978) Freedom to Study: requirements of overseas students in the UK London; Overseas Students Trust.

Rees,J(1990) Foreign language aptitude testing - a place in the 1990s Unpublished MA dissertation, University of BIrmingham.

Rees,R.D(1969) "Dimensions of students' point of view in rating college teachers" Journal of Educational Psychology 60, 476 - 482.

Regmi,M(1987) "Historical development of psychology in Nepal" in G.H.Blowers and A.M.Turtle (eds) Psychology Moving East Colorado: West View Press.

Remmers,H.H(1939) "Appraisal of college teaching through ratings and student opinion" Yearbook of National Society of College Teachers of Education 27 ChXV, 227 - 234.

Remmers,H.H(1959) The appraisal of teaching in large universities Ann Arbor: University of Michigan.

Remmers,H.H(1960) Manual of instruction for the Purdue Rating Scale for Instructors(Rev ed) West Lafayette, Ind: University Book Store.

Remmers,H.H and Brandenburg,G.C(1927) "Experimental data on the Purdue Rating Scale for Instruction" Educational Administration and Supervision 13, 519 - 527.

Remmers,H.H,Martin,F.D and Elliott,D.N(1949) "Are students' ratings of instructors related to their grades?" Purdue University Studies in Higher Education 66, 17 - 26.

Remmers,H.H and Weisbrodt,J.A(1965) Manual of instructions for the Purdue Rating Scale for Instructors(Rev ed) West Lafayette,Ind: University Book store, Purdue University.

Renninger,K.A and Snyder,S.S(1983) "Effects of cognitive style on perceived satisfaction and performance among students and teachers" Journal of Educational Psychology 75(5), 668 - 676.

Rezler,A.G(1965) "The influence of needs upon the student's perception of his instructor" Journal of Educational Research 58, 282 - 286.

Richins,M.L and Bloch,P.H(1991) "Post-purchase product satisfaction:incorporating the effects of involvement and time" Journal of Business Research 23, 145 - 158.

Riley,J.W, Ryan,B.F and Lifschitz,M(1950) The student looks at his teacher New Brunswick, New Jersey: Rutgers University Press.

Roberts,D and Higgins,A(1992) Higher Education:Student Experience London: Polytechnics Central Applications Service.

Rodin,M and Rodin,B(1972) "Student evaluations of teachers" Science 177(29 Sept), 1164 - 1166.

Rokeach,M(1960)The open and closed mind New York: Basic Books.

Root,A.R(1931) "Student ratings of teachers" Journal of Higher Education 2, 311 - 315.

Rosenthal,R and Jacobson,L(1968) Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development New York: Holt,Rinehart and Winston.

Rubinstein,J and Mitchell,H(1970) "Feeling free, student involvement and appreciation" Proceedings of the 78th Annual Convention of the APA, 623 - 624.

Ryan,J.J, Anderson,J.A and Birchler,A.B(1980) "Student evaluation: the faculty responds" Research in Higher Education 12(4), 317 - 333.

Safi,Q and Miller,R.I(1986) "Student evaluations of courses and instructors at Kuwait University" Higher Education Review 18(3 - summer), 17 - 26.

Saracho,O.N(1991) "Students' preference for field dependence -independence teacher characteristics" Educational Psychology 11(3+4 double issue), 323 - 332.

Schuetz,A(1943)"The stranger:an essay in social psychology" American Journal of Sociology 49, 499 - 507.

Schuman,H and Presser,S(1981) Questions and Answers in Attitude Surveys New York: Academic Press.

Scriven,M(1981) "Summative teacher evaluation" in J.Millman (ed) Handbook of teacher evaluation Beverley Hills,CA: Sage, 244 - 271.

Seiler,L.H,Weybright,L.D and Stang,D.J(1977) How useful are published evaluation ratings selecting courses and instructors?" Unpublished manuscript.

Sen,A(1970) Problems of Overseas Students and Nurses Slough:National Foundation for Educational Research.

Sharon,A.T(1970)"Eliminating bias from student rating of college instructors" Journal of Applied Psychology 54, 278 - 281.

Sharon,A.T and Bartlett,C.J(1969) "Effects of instructional conditions in producing leniency on two types of rating scales" Personnel Psychology 22, 251 - 263.

Shatz,M.A and Best,J.B(1986) "Selection of items for course evaluation by faculty and students" Psychological Reports 58, 239 - 242.

Shingles,R.D(1977)"Faculty ratings: procedures for interpreting student evaluations" American Educational Research Journal 14, 459 - 470.

Shotnes,S(1985)ed Overseas Students - Who Learns What? London: United Kingdom Council for Overseas Students Affairs.

Shwalb,B.J and Shwalb,D.W(1990) The design of a college course ratings form by a psychology 'tests and measurement' class Paper presented at annual American Educational Research Association meetings (April 16-20), Boston.

Singh,A.K(1963) Indian Students in Britain London: Asia Publishing House.

Singh,J(1991) "understanding the structure of consumers' satisfaction evaluations of service delivery" Journal of the Academy of Marketing Science 19(3), 223 - 244.

Smith,B.G.N,Fusilier,C.N,Bagramian,R.A and Bottomley,W.K(1969) "A criterion model of the dental school instructor" Journal of Dental Education 33, 523 - 531.

Smith,M.C and Pearn,M.A(1972) "Conventional lecture versus discovery training methods in the preparation of immigrant bus conductors" Journal of Cross-cultural Psychology 3(4), 407 - 412.

Smith,R.A and Cranton,P.A(1992)"Students' perceptions of teaching skills and overall effectiveness across instructional settings" Research in Higher Education 33(6), 747 - 764.

Smithers,A(1970a) "Some factors in lecturing" Educational Review 22(2), 141 - 150.

Smithers,A(1970b)"What do students expect of lectures?" Universities Quarterly 24(summer), 330 - 336.

Snyder,C.R and Clair,M(1976) "Effects of expected and obtained grades on teacher evaluation and attribution of performance" Journal of Educational Psychology 68, 75 - 82.

Startup,R(1972) "Student satisfaction with academic services", Educational Research 14(2), 135 - 140.

Stone,L.A and James,R(1965)"Interval scaling of the prestige of selected secondary education teacher-specialists" Perceptual and Motor Skills 20, 859 - 860.

Sudman,S(1967) Reducing the Cost of Surveys Chicago: Aldine Publishing Co.

Swan,J.E and Jones Combs,L(1976) "Product performance and consumer satisfaction: a new concept" Journal of Marketing 40(April), 25 - 33.

Swan,J.E and Martin,W.S(1981) "Testing comparison level and predictive expectations model of satisfaction" in Advances in Consumer Research Vol VIII K.Monroe ed. Ann Arbor: Association for Consumer Research, 77 - 82.

Tetenbaum,T(1977) "The factor invariance of student ratings of instruction under 3 sets of directions" Research in Higher Education 6, 11 - 23.

Thorndike,R.L and Hagen,E(1977) Measurement and evaluation in psychology and education(4th ed) New York: John Wiley.

Thurstone,L.L and Chave,E.J(1929) The Measurement of Attitudes Chicago: University of Chicago Press.

Tiberius,R.G,Sackin,D and Cappe,L(1987) "A comparison of two methods for evaluating teaching" Studies in Higher Education 12(3), 287 - 298.

Tollefson,N, Chen,J.S and Kleinsasser,A(1989) "The relationship of students' attitudes about effective teaching to students' ratings of effective teaching" Educational and Psychological Measurement 49, 529 - 536.

Treffinger,D.J and Feldhusen,J.F(1970) "Predicting students' ratings of instruction" Proceedings ,78th Annual Convention,APA 621 - 622.

Tse,D.K and Wilton,P.C(1988) "Models of consumer satisfaction formation: an extension" Journal of Marketing Research 25(May), 204 - 212.

Tuckman,B.w and Oliver,W.F(1968) "Effectiveness of feedback to teachers as a function of source" Journal of Educational Psychology 59, 297 - 301.

Van Gelder,L et al(1975) Didactische analyse I Groningen: Pedagogisch Instituut.

Walker,B.D(1969) "An investigation of selected variables relative to the manner in which a population of junior college students evaluate their teachers" Dissertation Abstracts 29 (9-B), 3474.

Watkins,D(1990) "Student ratings of tertiary courses for 'alternative calendar' purposes" Assessment and Evaluation in Higher Education 15(1 - Spring), 12 - 21.

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 teaching: a Filipino investigation" Educational and Psychological Measurement 52(3 - Autumn), 727 - 734.

Watkins,D and Regmi,M(1991) Attributing success and failure in Nepal: a very different story Unpublished manuscript, University of Hong Kong.

Watkins,D and Regmi,M(1992) "Student evaluations of tertiary teaching: a Nepalese investigation" Educational Psychology 12(2), 131 - 142.

Watkins,D,Marsh,H.W and Young,D(1987)"Evaluating tertiary teaching: a New Zealand perspective" Teaching and Teacher Education 3(1), 41 - 53.

Watkins,D and Thomas,B(1991) "Assessing teaching effectiveness: an Indian perspective" Assessment and Evaluation in Higher Education 16(3 - Autumn), 185 - 198.

Watt,J(1978)"Performance of overseas postgraduate students: a management teacher's view" in Study Modes and Academic Development of Overseas Students K.James ed,38 - 43.

Westbrook,R.A and Reilly,M.D(1983) "Value-percept disparity: an alternative to the disconfirmation of expectations theory of consumer satisfaction" in Advances in Consumer Research Vol X R.P.Bagozzi and A.M.Tybout eds. Ann Arbor: Association for Consumer Research, 256 - 261.

Wherry,R.J(1952) Control of bias in ratings Deoartment of the Army, The Adjutant General's Office, Personnel Research and Procedures Division, Personnel Research Branch. PRS Reports 914,915,919,920 and 921.

Whitten,B.J and Umble,M.M(1980) "The relationship of class size, class level, core vs non-core classification for a class to student ratings of faculty: implications for validity" Educational and Psychological Measurement 40, 419 - 423.

Whitely,S.E and Doyle,K.O(1976) "implicit theories in student ratings" American Educational Research Journal 13(4), 241 - 253.

Whitely,S.E, Doyle,K.O and Hopkinson,K(1973) Student ratings and criteria for effective teaching Minneapolis:Measurement Services Centre, University of Minnesota.

Widmeyer,W.N and Loy,J.W(1988)"When you're hot: warm-cold effects in first impressions of persons and teaching effectiveness" Journal of Educational Psychology 80(1), 118 - 121.

Wigington,H,Tollefson,N and Rodriguez,E(1989) "Students' ratings of instructors revisited: interactions among class and instructor variables" Research in Higher Education 30(3), 331 - 344.

Wiley,D.E(1967)"Latent partition analysis" Psychometrika 32, 183 - 193.

Williams,G,Loder,C and Fry,H(1993)"Identifying priorities" Centre for Higher Education Studies Newsletter 3(April);Institute of Education, University of London.

Williams,G, Woodhall,M and O'Brien,U(1986) Overseas students and their place of study: report of a survey. London:OST.

Williams,P(1982) A Policy for Overseas Students London: Overseas Students Trust.

Wilson,R.C(1986)"Improving faculty teaching" Journal of Higher Education 57(2 - March/April), 196 - 211.

Winter,M and Morris,E.W(1979) "Satisfaction as an intervening variable" in New Dimensions of Consumer Satisfaction and Complaining Behaviour R.L.Day and H.K.Hunt eds. Bloomington, Illinois: University of Indiana, 15 - 21.

Woodruff,R.B,Cadotte,E.R and Jenkins,R.L(1983) "Modeling consumer satisfaction processes using experience-based norms" Journal of Marketing Research 20(Aug), 296 - 304.

Wright,J(1985) "Intercultural postgraduate learning the acquisition of study skills: an institutional response to the results of the research" International Journal for the Advancement of Counselling 8, 279 - 296.

Wulff,D.H, Staton-Spicer,A.Q, Hess,C.W and Nyquist,J.D(1985) "The student perspective on evaluating teaching effectiveness" ACA Bulletin 53, 39 - 47.

Yonge,G.D and Sassenrath,J.M(1968) "Student personality correlates of teacher ratings" Journal of Educational Psychology 59, 44 - 52.

Also:

Coolican,H(1990) Research Methods and Statistics in Psychology Hodder and Stoughton,London.

Runyon,R.P and Haber,A (1977) Fundamentals of Behavioural Statistics 3rd ed. Addison-Wesley Publishing Co, Reading, Mass.

APPENDIX 1

BIRMINGHAM OVERSEAS STUDENTS TEACHING EVALUATION QUESTIONNAIRE

NATIONALITY: COURSE:

SEX: LECTURER RATED:

Using the following key, rate the lecturer on each of the statements.

1 = Poor
2 = Below Average
3 = Average
4 = Just Above Average
5 = Good
6 = Excellent

1. Makes good use of handout.

2. Is well-prepared for classes.

3. Structures difficult topics in easily understandable ways.

4. Has good knowledge of own subject.

5. Well-organised and clearly presented lectures.

6. Explains concepts, principles, and abstract theories clearly and precisely.

7. Encourages and stimulates students to independent thought.

8. Clearly states objectives of course at beginning.

9. Encourages students to express themselves freely and openly, and takes their views into account.

10. Provides adequate feedback with helpful/constructive comments on work.

11. Speaks audibly and clearly during lectures.

12. Is aware when students having difficulty in understanding topics.

13. Overall evaluation of lecture.

ANY FURTHER COMMENTS ABOUT THE LECTURE.

APPENDIX 2

ITEMS USED IN PILOT STUDY

A - Makes good use of handouts.
B - Deals fairly and impartially with all students.
C - Discusses recent developments in subject.
D - Has good general knowledge.
E - Is well-prepared for classes.
F - Structures difficult topics in easily understandable ways.
G - Has good knowledge of own subject.
H - Encourages class members to work as a team.
I - Well-organised and clearly presented lectures.
J - Encourages all class members to contribute to discussions.
K - Provides full set of references for studies quoted.
L - Explains concepts, principles, and abstract theories clearly and precisely
M - Relates course material to other subjects, everyday life or future career.
N - Avoids excess of factual material.
O - Encourages and stimulates students to independent thought.
P - Includes material not readily available in textbooks.
Q - Covers course topics in sufficient depth.
R - Answers questions accurately.
S - Clearly states objectives of course at beginning.
T - Finds ways to help students answer their own questions.
U - Clearly defines student responsibility on the course.
V - Welcomes questions and discussion from students.
W - Makes clear how each topic fits in course.
X - Encourages students to express themselves freely and openly, and takes their views into account.
Y - Always well-groomed, neat and clean in personal appearance.
Z - Provides adaquate feedback with helpful/constructive comments on work.
A1 - Concerned about effectiveness of teaching.
B1 - Returns written work promptly.
C1 - Maintains continuity in course.
D1 - Speaks clearly and audibly during lectures.
E1 - Readily available to help outside classes.
F1 - Accomplishes objectives of course,
G1 - Has good vocabulary.
H1 - Awards grades suitably and fairly based on work performed.
I1 - Stresses important material on course.
J1 - Indicates how much attention to give to different parts of reading material.
K1 - Is aware when students having difficulty in understanding topics.

L1 - Speaks with expression, and varies tone of voice during lectures.
M1 - Avoids trivial time-filling material.
N1 - Writes legibly on blackboard.
O1 - Seems genuinely interested in students.
P1 - Skilful in blackboard drawing.
Q1 - Stresses application as well as theory.
R1 - Constrasts varying views objectively.
S1 - Stimulates interest in subject beyond course requirements.
T1 - Invites criticism of own ideas.
U1 - Makes good use of examples and illustrations in lectures.
V1 - Motivates students to do best work.
W1 - Speaks at pace suitable for maximum comprehension, and facilitates note-taking.
X1 - Shows interest and enthusiasm in own subject.
Y1 - Tries to link lectures with other parts of course.
Z1 - Has sense of humour.
A2 - Keeps strictly to textbook in lectures.
B2 - Provides all students need to pass examinations.
C2 - Entertaining to listen to in lectures.
D2 - Punctual and reliable in attendance at classes.
E2 - Changes approach to meet new situations.
F2 - Free from annoying personal pecularities.
G2 - Indicates where relevant information not dealt with in the lecture can be found.
H2 - Demands reasonable amount of work.

Copy of final questionnaire

PART 1: General Questions.

What is your nationality?

What is your course?

What year of your course are you studying now?

How long have you been in the UK?

Is this your first time to study here?

Are you male or female?

Which age group are you? 18-21; 22-30; 31-40; 41-50; 51+

PART 2: Listed below are 22 characteristics that an ideal lecturer could possess. Please rank them in terms of importance, starting with 1 as most important.

A _____ Makes good use of handouts.
B _____ Discusses recent developments in subject.
C _____ Has good general knowledge.
D _____ Is well-prepared for classes.
E _____ Structures difficult topics in easily understandable ways.
F ----- Has good knowledge of own subject.
G _____ Well-organised and clearly presented lectures.
H _____ Explains concepts, principles, and abstract theories clearly and precisely.
I _____ Relates course material to other subjects,everyday life or future career.
J _____ Encourages and stimulates students to independent thought.
K _____ Answers questions accurately.
L _____ Clearly states objectives of course at beginning.
M _____ Encourages students to express themselves freely and openly, and takes their views into account.
N _____ Provides adaquate feedback with helpful/constructive comments on work.
O _____ Includes relevant personal anecdotes in lecture.
P _____ Speaks clearly and audibly during lectures.
Q _____ Is aware when students having difficulty in understanding topics.
R _____ Motivates students to do best work.
S _____ Shows interest and enthusiasm in own subject.
T _____ Has sense of humour.
U _____ Welcomes questions and discussion from students.
V _____ Has good vocabulary.

APPENDIX 3

BREAK-DOWN OF RESPONSES BY TWO CRITERIA

1. Nationality and Department.

South-east Asian students (nationality group 1) are distributed in two main departments. Group 2 includes the Business school(MBA), and department group 5 includes a class of Japanese learning English.
African students(3) are concentrated in department group 2, which is particularly finance and accounting here.

TABLE A3.1

Dept/ Nat	CSS	Acc	Admin	Ed	Arts	Med	Civ	Mech	Ma	Sci	Law
SE Asia	10	24	2	1	15	0	2	7	1	4	0
Indian S-C	1	1	3	1	0	2	0	1	1	1	0
Africa	0	16	5	1	2	2	4	1	1	3	0
Arabic	0	2	0	0	0	3	0	0	0	0	0
Europe	0	1	1	1	0	1	0	4	4	2	1
Car/N.Am	0	1	1	1	4	0	0	0	0	0	0
C+S.Am	0	0	1	0	3	1	1	0	1	1	0

Missing=17.

Table showing the responses to the questionnaire by nationality grouping and departmental grouping. (See pages 81 and 82 for full details of groupings).

2. Nationality and Year of course.

South-east Asian students on the MBA course, and

African students on the finance courses are 1 year postgraduate courses. The South-east Asian students in their 2nd year are mainly undergraduates. A few students have studied for their "A" levels in the UK.

TABLE A3.2

YEAR/ NAT	1	2	3	4
SE Asia	43	16	5	1
Indian S-C	8	1	3	0
Africa	33	1	1	0
Arabic	4	3	1	0
Europe	7	5	5	0
Car/N.Am	4	1	1	0
C+S.Am	4	2	3	0

Missing=12.

Table showing responses to questionnaire based on nationality and year of course. (See page 82 for full details of nationality groupings).

3. Nationality and Visit.

TABLE A3.3

VISIT BEFORE/ NAT	YES	NO
SE Asia	8	63
Indian S-C	1	11
Africa	8	28
Arabic	2	6
Europe	2	15
Car/N.Am	1	5
C+S.Am	2	7

Missing=6.

Table showing responses to questionnaire based on nationality groupings and whether first visit to UK. (See page 82 for full details of nationality groupings).

4. Nationality and Sex.

TABLE A3.4

SEX/ NAT	MALE	FEMALE
SE Asia	48	22
Indian S-C	10	2
Africa	32	4
Arabic	5	3
Europe	11	6
Car/N.Am	2	5
C+S.Am	8	1

Missing=6.

Table showing response to questionnaire based on nationality grouping and sex of respondent. (See page 82 for full details of nationality groupings).

5. Nationality and Age.

TABLE A3.5

AGE/ NAT	18-21	22-30	31-40	41-50	51+
SE Asia	10	22	36	1	2
Indian S-C	1	4	3	3	1
Africa	0	10	17	9	0
Arabic	0	2	4	1	1
Europe	3	13	0	1	0
Car/N.Am	1	2	1	1	1
C+S.Am	0	3	4	2	0

Missing=6.

Table showing the responses to the questionnaire based on nationality grouping and age group. (See page 82 for full details of nationality groupings).

6. Department and Year of course.

TABLE A3.6

YEAR/ DEPT	1	2	3
CSS	9	1	0
Acc	33	8	0
Admin	11	0	0
Ed	4	1	0
Arts	19	3	2
Med	6	0	3
Civ	6	1	0
Mech	5	4	3
Ma	0	3	5
Sci	4	4	3
Law	1	0	0

Missing=26.

Table showing responses to questionnaire based on departmental grouping and year of study. (See page 81 for full details of departmental groupings).

7. Department and Sex.

TABLE A3.7

SEX/ DEPT	MALE	FEMALE
CSS	6	5
Acc	36	9
Admin	9	4
Ed	2	3
Arts	12	11
Med	8	1
Civ	7	0
Mech	12	1
Ma	6	2
Sci	9	2
Law	1	0

Missing=19.

Table showing response to questionnaire based on departmental grouping and sex of respondent. (See page 81 for full details of departmental groupings).

8. Department and Age.

TABLE A3.8

AGE/ DEPT	18-21	22-30	31-40	41-50	51+
CSS	1	2	6	1	1
Acc	4	15	20	4	2
Admin	0	1	7	4	2
Ed	0	2	2	1	0
Arts	3	7	9	4	1
Med	0	3	5	1	0
Civ	1	3	2	1	0
Mech	4	8	1	0	0
Ma	1	5	1	1	0
Sci	0	5	6	0	0
Law	1	0	0	0	0

Missing=19.

Table showing responses to questionnaire based on departmental groupings and age groups. (See page 81 for full details of departmental groupings).

APPENDIX 4

Categories of Instructional Characteristics as used by Feldman.

For each innstructional dimension, examples of evaluative items that would be classified into it are given.

1. Stimulation of interest: the instructor puts material across in an interesting way; the instructor gets students interested in the subject; it was easy to remain attentive; the teacher stimulated intellectual curiosity etc.

2. Instructor's Enthusiasm: the instructor shows interest and enthusiasm in the subject; the instructor seems to enjoy teaching; the teacher communicates a genuine desire to teach students; the instructor never showed boredom for teaching this class; the instructor showed energy and excitement etc.

3. Knowledge of subject matter: the instructor has a good command of the subject material; the teacher has a thorough knowledge, basic and current, of the subject; the instructor has good knowledge about or beyond the textbook; the instructor knows the answer to students' questions; the teacher keeps lecture material updated etc.

4. Intellectual expansiveness: the teacher is well informed in all related fields; the teacher has respect for other subject areas and indicates their relationship to his or her own subject of presentation; the teacher exhibited a high degree of cultural attainment etc.

5. Preparation/organisation: the teacher was well prepared for each day's lecture; the presentation of the material is well organised; the overall development of the course has good continuity; the instructor planned the activities of each class period in detail etc.

6. Clarity: the instructor gave clear explanations; the instructor interprets abstract ideas and theories clearly; the instructor makes good use of examples and illustrations to get across difficult points; the teacher effectively synthesizes and summarises the material; the teacher answers students' questions in

a way that helps students to understand etc.

7. Elocution: the instructor has a good vocal delivery; the teacher speaks distinctly,fluently and without hesitation; the teacher varied the tone of his or her voice; the teacher has the ability to speak distinctly and be clearly heard; the instructor changed pitch,volume and quality of speech etc.

8. Sensitivity: the teacher was skilled in observing students' reaction; the teacher was aware when students failed to keep up in class; the instructor teaches near the class level; the teacher takes an active personal interest in the progess of the class and shows a desire for students to learn etc.

9. Course Objectives: the purposes and policies of the course were made clear to the student; the instructor gave a clear idea of student requirements; the teacher clearly defined student responsibilities in the course; the teacher tells students which topics are most important and what they can expect on tests; the instructor gave clear assignments etc.

10. Course Materials: the teacher has the ability to apply material to real life; the instructor makes the course practical; there is worthwhile and informative material in lectures that doesn't duplicat the text; the course has excellent content; the class considers what we are learning worth learning etc.

11. Supplementary Materials: the homework assignments and supplementary readings were helpful in understanding the course; the teacher made good use of teaching aids such as films and other audio-visual materials; the instructor provided a variety of activities in class and used a variety of media(slides,films,projections,drawings) and outside resource persons etc.

12. Outcome: gaining of new knowledge was facilitated by the instructor; I developed significant skills in the field; I developed increased sensitivity and evaluative judgment; the instructor has given me tools for attacking problems; the course has increased my general knowledge; apart from your personal feelings about the teacher,has he/she been instrumental in increasing knowledge of the course's subject matter etc.

13. Fairness: grading in the course was fair; the instructor has definite standards and is impartial in grading; the exams reflect material emphasized in the course; test questions were clear; coverage of subject matter on exams was comprehensive etc.

14. Personality: the teacher has a good sense of humour; the teacher was sincere and honest; the teacher is highly personable at all times in dress, voice, social grace, and manners; the instructor was free of personal peculiarities; the instructor is not autocratic and does not try to force us to accept his ideas and interpretations; the teacher exhibits a causal, informal attitude; the instructor laughed at his own mistakes etc.

15. Feedback: the teacher gave satisfactory feedback on graded material; criticism of papers was helpful to students; the teacher told students when they had done a good job; the teacher is prompt in returning tests and assignments etc.

16. Encouragement of Questions: students felt free to ask questions or express opinions; the instructor stimulated class discussion; the teacher encouraged the students to express differences of opinions and to evaluate each other's ideas; the instructor invited criticism of his or her own ideas; the teacher appeared receptive to new ideas and the viewpoints of others etc.

17. Intellectual Challenge: this course challenged students intellectually; the teacher encouraged students to think out answers and follow up ideas; the teacher attempted to stimulate creativity; the instructor raised challenging questions and problems etc.

18. Respect: the instructor seems to have a genuine interest in and concern for students; the teacher took students seriously; the instructor established good rapport with students; the teacher was friendly toward all students etc.

19. Availability: the instructor was willing to help students having difficulty; the instructor was willing to give individual attention; the teacher was available for consultation; the teacher was accessible to students outside class etc.

20. Motivated students: instructor motivates students to do their best work; the instructor sets high standards of achievement for students; the teacher raises the aspirational level of students etc.

21. Self-Initiated Learning: students are encouraged to work independently; students assume much responsibility for their learning; the general approach used in the course gives emphasis to learning on the students' own; the teacher does not suppress individual initiative etc.

22. Research Productivity: the talks about his own research; instructor displays high research accomplishments; the instructor publishes material related to his subject field etc.

23. Difficulty (description): the workload and pace of the course was difficult; I spent a great many hours studying for this course; the amount of work required for this course was very heavy; this course required a lot of time; the instructor assigned very difficult reading etc.

24. Difficulty (evaluation): the content of this course is too hard; the teacher's lectures and oral presentations are over my head; the instructor often asked for more than students could get done; the instructor attempted to cover too much material and presented it too rapidly etc.

25. Classroom Management: the instructor controls class discusion to prevent rambling and confusion; the instructor maintained a classroom atmosphere conducive to learning; students are allowed to participate in deciding the course content; the teacher did not rule with an iron hand etc.

26. Classroom Atmosphere: the class does not make me nervous; I felt comfortable in this class; the instructor created an atmosphere in which students in the class seemed friendly; this was not one of those classes where students failed to laugh, joke, smile, or show other signs of humour; the teacher is always criticising and arguing with students etc.

27. Individualization: instead of expecting every student to do the same thing, the instructor provides different activities for different students; my grade

depends primarily on my improvement over my past performance; in this class each student is accepted on his or her own merits; my grade is influenced by what is best for me as a person as well as by how much I have learned; the instructor evaluated each student as an individual etc.

28. Pursuit of Objectives: the instructor accomplished what he or she set out to do; there was close agreement between the announced objectives of the course and what was actually taught; course objectives stated agreed with those actually pursued etc.

Based on Feldman(1989b; Appendix A pp633-636). Original details in Feldman(1976b): only 19 characteristics; 12 was difficulty of course, and 14 classroom management.

Feldman(1993) adds:
29. Overall Rating of Lectures as an Item of a Multi-item Indicator.

30. Overall Rating of Teacher as an Item of a Multi-item Indicator.

31. Overall Rating of Course as an Item of a Multi-item Indicator.

APPENDIX 5

OTHER LESS IMPORTANT POTENTIAL BIASES OF STUDENT RATINGS OF INSTRUCTION

1.STYLE OF LEARNING. Entwistle and Ramsden(1983) proposed 4 styles or approaches to learning:
a) Deep approach - students attempt to understand rather than just accept.Using other approaches.
b) Comprehension learning - building overall description of content and link to previous knowledge.
c) Operation learning - detailed attention to evidence.
d) Surface approach - memorization (p42).

Students were allocated to a style of learning by Lancaster Approaches to Study Inventory, then given the Course Perceptions Questionnaire. The general conclusion, which was replicated 7 years later is "that students who adopt meaning or reproducing orientations also prefer the methods of teaching and assessing which encourage those approaches to learning" (Entwistle and Tait 1990 p188).

Confirmed by Prosser and Trigwell(1990) in Australia:"courses in which students adopted deeper approaches to study were also the courses that had teaching that was rated more highly"(p 141).

2.TEACHER PERSONALITY. Jones(1989) tried to investigate what students actually evaluate about the instructor - is it really the course/teaching, or their personality? After analysis of the results, it was found that the student ratings of teacher personality loaded on the same factor as their rating of teacher competence. Thus teacher personality is seen as part of teaching competence, and that "in fact it would be very surprising if students' perception of a teacher's personality did not affect their rating of her or his teaching competence"(p556).

Jones et al(1985) report that students look at two aspects of the teacher: technical aspects(ability to explain/knowledge of subject) and personological aspects(personality - eg listens to students).

Crittenden and Norr(1983) tried to apply a "person perception" model to teacher evaluation, and sees it as a special case of person perception.

Flood Page(1974) says the relationship is "to say the least, obscure"(p55). It is not always easy "to separate best from worst teachers on personality grounds"(p52). Costin et al(1971) agree after reviewing 12 studies.

Furthermore, mere popularity is not enough. But Guthrie(1954) did find that students rated higher those instructors, who had great interest/enthusiasm for their subject.

Nor is there any relationship between the teacher's

activities outside the classes (ie allocation of time to research/preparation etc), and good/bad teaching (Hildebrand and Wilson 1970).

3.STUDENT'S SENIORITY. This can be looked at as actual age of student or year of course. Studies vary from finding that senior students rate higher (eg Whitten and Umble 1980) to no relationship (eg Marsh and Overall 1981). Smith et al(1969) showed that students' attitudes to what is good teaching on a dental course changes over time, particularly on three items: "is cognizant of student problems"; "encourages student judgment"; and "possesses current knowledge of subject"(p48). But the general ratings did not change (Quoted in Flood Page 1974).

4.SIMILARITY BETWEEN TEACHER/STUDENT. Tollefson et al(1989) looked at the question of whether students would rate higher a teacher who held the same attitudes to themselves about what is effective teaching. Based on the social psychological theory that individuals are attracted to persons who hold similar views (Byrne and Clore 1970; Byrne and Nelson 1965). Earlier studies were unclear. Tollefson et al used the Attitude Toward Effective Teaching Scale(ATET), and the Teacher Rating Scale(TRS)(McKnight 1973). This study was also inconclusive - two separate analyses produced conflicting results.
Feldman(1977)"There are hints that under some circumstances similarity of teacher-student gender is associated with higher ratings"(p245).

5.FORMAL TRAINING.Costin(1968) noted that GTAs in psychology who attended a short teaching course received higher ratings in "feedback" and "group interaction" than those who had not.

6.IMPRESSIONS OF INSTRUCTOR. Overall impression: There is evidence that the overall impression of instruction can influence specific ratings of a lecturer. For example,Pohlmann (1972) found a correlation of approx 0.2 between overall evaluation of instruction at Southern Illinios University, and specific teacher ratings. Other studies find varying correlations.

Initial impression: Feldman(1977) quotes studies suggesting that between one fifth to one third of variance in final ratings is due to the students' early impressions.

Pre-course impressions: Students who have heard a professor is good, rate them higher than those who have not heard about the professor(Miller 1972). But there is a selection effect here - students are more likely to select courses taught by instructors they have heard good comments about or have had good experiences with before, than unfamiliar instructors, or ones who have received poor reports.

However,there is concern over the effects of precourse expectations. Barke et al(1983) compared responses on the Affective Entry Questionnaire and Course Evaluation Questionnaire. Respondents tend to answer "no basis on which to make judgment" in the first questionnaire, suggesting "that, as a rule, students may have fewer expectations or biases that could potentially influence end-of-course ratings than many instructors believe"(p83).

7.STUDENT ABILITY.No relationship of either type. Remmers et al(1949) explain these results by the fact that teaching is aimed at the whole class. For example, it may be too slow for brighter students, leading to a poorer rating of the teacher, but just right for slower students leading to a favourable rating.

8.MISCELLANEOUS FACTORS.

- Arubayi(1987) adds time of day(morning lectures rated higher),and mood of students.Other evidence on time of day inconsistent (see Feldman 1979 p219).

- McClelland(1970) divided students into 3 groups randomly: normal ratings forms given to one group; rating forms that contained alleged previous ratings, but artificially high to another group; and the same to the last group, but ratings artificially low. Significant differences found for groups 2 and 3 ie higher or lower ratings respectively. Student ratings can, thus, be easily influenced it was suggested.

- Students' feeling of control significantly correlated to appreciation of instructor (Rubinstein and Mitchell 1970).

- The fear that students who are hostile to the lecturer my give them poor ratings is not borne out by Crannell(1948). However, there is little other research.

- Kappes(1988) compared ratings of full-time and part-time lecturers. The latter rated significantly higher on "treating students with respect" and "starting/ending class on time". Full-timers rated higher on 8 items. This was confirmed by Kirker(1990).

- Doyle(1982) suggests that based on commom sense, events outside the classroom could influence the evaluation - eg the day before a big event, or the busy last week of term.

Interaction of Biases

Wigington et al(1989) looked at the interaction between class type (ie lecture or seminar etc); class level; class size; instructor reputation, rank and sex. The data were analysed through 15 two-way factorial analyses of variance. The interactions found are detailed in the table A5.1.

There was no significant relationship for reputation by level, reputation by sex, and reputation by size. The authors conclude that " student ratings do reflect differences in instructional effectiveness". But "an interpretation of student ratings needs to reflect an understanding of the variables that interact to produce differences in student ratings of instructors"(p342).

Klyczek(1989) developed a path analysis of professional rank, age, gender, status, communication skills, relationship with students, availability to students, and publishing productivity. All variables had stronger relationships with each other than to student ratings of instruction.

TABLE A5.1

INTERACTION/ VARIABLES	USUAL RELATIONSHIP OF SINGLE VARIABLES	EFFECT OF INTERACTION
Type of course by size of class.	Type:Discussion classes rated higher than lectures etc. Size:Smaller classes rated higher.	Lecture-discussion: small classes rated lower. Lab format: small/large classes rated lower than medium sized.
Level of course by sex of instructor.	Level:Higher courses rated higher. Sex: No difference.	Male teachers rated higher on higher courses.
Type of course by rank of instructor.	Type:As above. Rank:No difference.	No consistent pattern.
Rank of instructor by size of class.	As above.	Teaching assistants have U shaped profile, professors negative correlation.
Level of course by rank of instructor.	As above.	Associate professors have highest rating from highest course.
Sex of instructor by size of class.	As above.	Male teachers higher rating on larger class.
Rank of instructor by sex of instructor.	As above.	Professors higher ratings if females.
Type of course by sex of instructor.	As above.	Female teachers higher rating for lecture,discussion and lab classes, but lower for lecture-discussion format.
Type of course by level of course.	As above.	Postgraduate courses rated lower for lectures and lab.
Reputation of instructor by type of course.	Reputation:Higher rating for greater reputation. Type:As above.	Higher rating for lecture and lab formats by teacher with reputation.
Level of course by size of class.	As above.	Higher courses moderate-sized classes lower rating than large classes.
Reputation of instructor by rank of instructor.	As above.	Professors highest rating when reputation important, and lowest when not.

Interaction of variables producing significant relationships in study by Wigington et al(1989).

APPENDIX 6

FACULTY DIFFERENCES FOUND IN STUDENT RATINGS OF INSTRUCTION

TABLE A6.1

% of students checking as one of the three most important qualities to good teaching.			
Qualities Thought to be Important to Good Teaching	For Teachers of Arts	For Teachers of Biological and Physical Science	For Teachers of Social Science
Ability to encourage thought	47	17	70
Expert knowledge of subject	54	70	42
Systematic organisation of subject matter	32	78	48
Ability to explain clearly	42	89	38
Enthusiastic attitude toward subject	46	16	26
Good speaking ability	31	6	11
Tolerance toward student disagreement	11	2	45
Fairness in marking and grading tests	8	10	7
Pleasing personality	16	4	5
Sympathetic attitude toward students	9	5	3

Table showing qualities chosen as one of the three most important qualities to good teaching based on faculty from Riley et al(1950) (Feldman and Newcomb 1969a p256).

TABLE A6.2

Rank of frequency of mention			
Characteristics of Best Teachers	For Teachers ofHuman-ities	For Teacher of Natural Science	For Teacher of Social Science
Dynamic, stimulating	1	4	1
Interested in and helpful to students	2	2	3
Clear, well organised	5.5	1	6
Skilled technique, well prepared	7	3	4
Holds interest	3	7	5
Broad knowledge, thorough mastery (of field)	5.5	9	2
Great interest in and devotion to field	4	8	8
Fine personality	8	6	7
Makes connection to other subjects	9	5	9

Table showing characteristics of best teacher for each of three subject matters according to students at Amherst, from Birney, Coplin and Grose(1960) (Feldman and Newcomb 1969a p257).

TABLE A6.3

Level of Rating	Course Effect	Instructor Effect
HIGH	Architecture;Art; Communications; Fine and Applied Arts; Health and Technology; Home Economics;Music; Secretarial Studies;Speech	Area Studies; Fine and Applied Art; Foreign Languages; Military Science; Music
MEDIUM HIGH	Agriculture; Education; Foreign Languages; Letters and Humanities; Library Science;Nursing; Physical and Health Education;Political Science and Government; Public Affairs;Religion and Theology;Social Work and Service;Trade and Vocational Technology Education	Art;Communications;English Language and Literature;Health and Technology;History;Home Economics;Law;Letters and Humanities;Library Science;Physical and Health Education;Political Science and Government;Psychology;Public Affairs;Religion and Theology;Secretarial Studies;Speech;Trade and Vocational Technology Education
MEDIUM	Area Studies;Biological Sciences;Health Professions; Interdisciplinary Studies;Military Science;Psychology	Business and Commercial Technology(*);Education (*);Engineering Technology;Nursing(*)
MEDIUM LOW	Accounting;Business and Commercial Technology;English Language and Literature;History;Law; Social Science;Sociology	Agriculture;Biological Sciences;Chemistry;Health Professions;Maths and Statistics;Philosophy; Social Science;Social Work and Service;Sociology
LOW	Business and Management;Chemistry; Computer and Information Science;Data-processing Technology;Economics; Engineering;Engineering Technology;Maths and Statistics;Philosophy; Physical Science;Physics	Accounting;Architecture; Business and Management;Computer and Information Science;Data-processing Technology;Economics; Engineering;Inter-disciplinary Studies;Physical Science;Physics

(*) = received mixed rankings, so average.

Table showing a summary of ratings on course effectiveness and instructor effectiveness for IDEA and SIR academic-field data. (Cashin 1990; table 1 pp116-7).

APPENDIX 7

ESTABLISHING RELIABILITY, GENERALIZABILITY AND VALIDITY OF STUDENT RATINGS OF INSTRUCTION

A7.1. Methods Used to Establish Reliability

1. INTERNAL CONSISTENCY. Using for example, odd-even or split-half, and coefficient alpha(Cronbach 1951) or Kuder-Richardson formulas(Kuder and Richardson 1937).

Aim to correlate various questions within the instrument. Studies have shown good internal consistency, eg Remmers and Weisbrodt(1965) Purdue Rating Scale for Instructors(PRSI) shown to have correlation of between 0.67 to 0.91 using Horst method(Horst 1949).

Costin et al(1971) quote correlations ranging from .77 to .94 for randomly paired students within a class. Feldman(1977) reports an extension of this approach, where two mean scores for a particular item can be obtained by randomly dividing a class in half. The resulting correlation is corrected by the Spearman-Brown Prophecy Formula, and it produces correlations between .70s and .90s (see Guilford 1954 for more details). Most of the commonly used instruments report reliability coefficients over 0.50. Table A7.1 shows a selection of post 1975 studies and the reliability coefficients reported.

But "simply computing the internal consistency of an entire questionnaire would be inappropriate unless the whole instrument were intended to measure a single quality and produce a single summary score across items"(Doyle 1975 p35).

2. TEST-RETEST. Instrument given to same subjects at two different times. Aim to correlate the two scores of each subject. For example, Remmers and Brandenburg(1927) administered the PRSI 3 days after the original use with the same group, and found a correlation between 0.42 to 0.92.

But the instructor may change between administrations of the instrument, and so a small correlation will suggest that the instrument is unstable. This method is also criticised for "being a test of the student's memory instead of being a measure of reliability"(Frey 1978 p85).

3. MEAN RATINGS. It is assumed that mean ratings of instructors should be different, because the instructors display different teaching behaviour. If the means are similar or identical, the ratings are biased. Whitely, Doyle and Hopkinson(1973) used this method in a large multi-section course; finding that the mean ratings varied between instructors.

TABLE A7.1

Name of Study	Instrument / Sample	Method of establishing reliability	Reliability coefficient
Frey(1978)	Endeavour 26 787 students at Northwestern University	Inter rater agreement	0.61 skill 0.32 rapport
Marsh(1982)	SEEQ 250 000 in 4 years at University of Southern California	(1)intraclass correlation (2)coeffic-ient alpha	(1)0.74 - 0.90 (2)0.88 - 0.97
Watkins and Thomas(1991)	SEEQ/Endeav-our combined 111 Indian students	Coefficient alpha	0.54 - 0.93(overall) 0.88(SEEQ median) 0.87(Endeav-our median)
Fernandez and Mateo(1992)	CUTEQ-R 36 589 students at Universidad Comlutense	Coefficient alpha	0.97 - 0.98
Watkins and Akande(1992)	SEEQ/Endeav-our combined 158 undergrads in Nigeria	Coefficient alpha	0.68 - 0.93(overall) 0.92(SEEQ median) 0.91(Endeav-our median)
Watkins and Gerong(1992)	SEEQ/Endeav-our combined 77 undergrads in Phillipines	Coefficient alpha	0.85 - 0.97(overall) 0.93(SEEQ median) 0.94(Endeav-our median)
Watkins and Regmi(1992)	SEEQ/Endeav-our combined with 297 Nepalese students	Coefficient alpha	0.54 - 0.84(overall) 0.79(SEEQ median) 0.73(Endeav-our median)

Table to show the coefficient of reliability found by selected post 1975 studies.

But the assumption that instructors do differ is open to question. However, enough research has shown that the distinction between "good" and "bad" lecturers can be established (eg Marsh 1977).

Frey(1978) used a variation of this method. He chose a sample of the data representing instructors who had taught 3 or more classes (with 10 + students in each), which had filled in ratings. Variance estimates can be calculated for differences among instructors, and differences among classes within instructors - inter rater agreement. A formula used (recommended by Ebel 1951) showed the proportion of observed variance due to differences in instructor.

4. ANOVA. Proposed by Guilford(1954): rather than attempting to remove potential bias, it aims to identify the contribution of
bias to the final rating, and adjust for it. Obviously, this has advantages because some potential biases cannot be easily separated(eg halo effect). For example, Treffinger and Feldhusen(1970), using this method, found that the halo effect only accounted for 10% of the variance in students'ratings(quoted in Doyle 1975 p43).

5.INTER-RATER RELIABILITY. This looks at the consistency of ratings among people. Reliability here is when all raters in a group give the same pattern of responses. Usually estimated by intraclass correlation coefficients, ie comparison of ratings within one class of one lecturer with ratings of different instructors. Because it is sensitive to the number of raters, Centra(1979) suggests intraclass correlations of .70s for 10 raters through to .90s for 20(p27).

Feldman(1977) makes a number of points about interpreting the reliability coefficients.

> i)"reliability coefficients of individual ratings indicate the degree of general or relative consistency among raters; they do not measure exact or absolute agreement"(p229).
>
> ii)inter-rater agreement is only the degree to which independent raters give the same rating for the same lecturer.
>
> iii)inter-rater reliablity is "the degree to which the ratings by different raters are proportional when expressed as deviations from their means"(p229).
>
> iv)reliability coefficients of average college student ratings maybe high, but this does not mean that individual students within the classes are highly consistent in their ratings.
>
> v)consistency in ratings among students may not be a good basis for estimating individual ratings or average ratings reliability, particularly

if the aim is to compare ratings across situations. Guthrie(1927) suggests that student ratings agree at the end of the term because of greater exposure to the lecturer, or student gossip.

A7.2. Generalizability of Student Ratings

This is the question of whether ratings of lecturers can be compared across situations. In a detailed analysis Bausell et al(1975) compared teaching behaviours in five situations:

i)Same course taught by same instructor on two separate occasions (CS-IS).
ii)Same course taught by two different instructors (CS-ID).
iii)Different courses taught by same instructor (CD-IS).
iv)Different course taught by different instructors within the same department (CD-ID).
v)Different courses from different departments taught by different instructors (CD'-ID').

It was found that generalisation was possible with all teaching behaviour in CS-IS; with some behaviour in CS-ID and CD-IS, but not CD-ID and CD'-ID', as expected. The authors conclude that student ratings do replicate as a whole across time, even if individual items are unclear.

Table A7.2 shows a summary of the correlations found by selected studies.

TABLE A7.2

STUDY	SAMPLE	CS-IS	CS-ID	CD-IS	CD-ID
Bausell et al (1975)	Unknown	.69	.33	.17	.07(same dept) .00(diff dept)
Feldman (1978)	2182 students	.66	.16**	.46*	-
Marsh (1982b)	8277 classes	.70	.14	.52	.06(same dept)

* quotes mean of Hogan(1973) and Seiler et al(1977)
** quotes only Hogan(1973).

Table showing the mean correlation coefficient of student ratings in different situations by selected students.

But Smith and Cranton(1992) suggest care. They talk about the "normative

assumptions" of a class, which would restrict generalizability. For example, large classes in certain subjects may see "organisation" as very important, while smaller classes in other subjects, it may be the importance of "interaction" factors. The authors emphasis that student ratings are "rather specific to the instructional setting"(p762).

More recently, Marsh and Bailey(1993) have looked at the generalizability of teaching characteristics - ie is a lecturer who is enthusiastic, but not organised, judged the same in all courses? Using over 1 million SEEQ forms, during a 13 year period, the authors are happy that "instructors appear to have distinct profiles of strengths and weaknesses that are highly generalizable"(p11).

Feldman(1978) takes up the issue of whether the samples of students used in ratings are from populations of comparable raters. It is not always possible to assume that the samples are random, because students self-select themselves for courses. Thus the samples are classed as coming from a population "like those observed". This makes it possible to correlate the average class rating between two classes taught the same course by the same lecturer. The correlations are between .60s and .70s (Feldman 1978 p201).
However, the correlations leave room for other factors - for example, the course context. Feldman's(1978) conclusion is that comparison of lecturer's ratings can only be of similar sized classes, similar subjects, and similar "requiredness".

A7.3 Establishing Validity of Student Ratings

Gaski(1987) produces evidence of studies both supporting (table A7.3) and non-supporting (table A7.4) the validity of student evaluations.

Using the most widely accepted objective measure of validity, the student achievement test, Feldman(1989b) summarises the studies, and finds a correlation with each individual characteristic of teaching (table A7.5). Not surprisingly, significant correlations are found for "teacher's preparation" and "clarity and understandableness", and the student achievement test. Table A7.6 gives the conclusions for use of student achievement tests as an indicator of student learning.

TABLE A7.3

STUDY	SAMPLE	RESULTS
Gessner(1973)	78 students	High correlation between student evaluation and performance
Frey(1973)	13 instructors 354 students	Strong relationship between SR and teaching quality(defined as difference between observed final exam score and score predicted by Scholastic Aptitude Test profile)
Marsh,Fleiner and Thomas(1975)	18 sections 720 students	Student evaluation(across sections) positively correlated with final exam
Marsh(1977)	62 instructors 591 classes 1847 students	Evaluations validated with retrospective reports of most/least outstanding
Marsh,Overall and Kesler(1979)	51 instructors 83 courses	Factor analysis indicated similar student-faculty evaluations dimensions; median r=.49 across evaluative factors; higher SR for courses instructor rated as most effective
Marsh and Overall(1980)	31 sections @960 students	Generally and moderately + relationship between SR and teaching effectiveness criteria,including final exam grade(36 of 60 corrs significant)
Howard and Maxwell(1980)	Two experiments i)8551 courses from 58 schools 200000 students ii)50 students each from 19 classes	Weak + relationship between expected grades and student satisfaction; student motivation and performance explained more of variation in satisfaction
Marsh(1982c)	329 classes	General agreement between student and instructor ratings in MTMM analysis
Howard,Conway and Maxwell(1985)	43 instructors 34students/ classes 30 former students/ instructor	Student and former student ratings reported superior in convergent/discriminant validition to other methods ieself,colleagues and trained observer ratings

Table showing research generally supportive of validity of student evaluations (Gaski 1987 p327).

TABLE A7.4

STUDY	SAMPLE	RESULTS
Rodin and Rodin(1972)	293 students	Inverse partial correlation between objective measure of amount learned and student rating(with initial ability controlled for)
Snyder and Clair(1976)	72 students	Expected grades inversely related to evaluations;perceived obtained grades positively related
Pratt and Pratt(1976)	175 students	Very little correlation between obtained grades and student ratings; strong positive correlation between expected grades and ratings
Brown(1976)	2360 sections 30000 student ratings	In stepwise regression, grades represent more powerful predictor of ratings(r=.353) than any other hypothesized antecedent
Powell(1977)	5 sections 35-45 students per section	Ratings of instructor falls as grading stringency increases; amount learned increases as grading stringency increases

TABLE A7.5

Table showing a summary of the results of studies relating specific evaluations of teaching to student achievement as found by Feldman(1989b).

Instructional Dimension	Number of Studies	Weighted simple average correlation
1.Teacher's stimulation of interest in the course and its subject matter	14	+.35 *
2.Teacher's enthusiasm	10	+.26 *
3.Teacher's knowledge of subject	9	+.24
4.Teacher's intellectual expansiveness	2	No correlation given
5.Teacher's preparation	18	+.44 *
6.Clarity and understandableness	24	+.47 *
7.Teacher's elocutionary skills	6	+.33 *
8.Teacher's sensitivity to, and concern with, class level and progress	11	+.27 *
9.Clarity of course objectives/requirements	6	+.32
10.Nature and value of course material	10	+.17
11.Nature and usefulness of supplementary materials and teaching aids	4	-.10
12.Perceived outcome or impact of instrument	14	+.40 *
13.Instructor's fairness	16	+.25 *
14.Personality characteristics	6	+.23
15.Nature, quality and frequency of feedback from teacher	13	+.22 *
16.Teacher's encouragement of questions and discussion	18	+.34 *
17.Intellectual challenge and encouragement of independent thought	7	+.23
18.Teacher's concern and respect for students	11	+.22 *
19.Teacher's avaialability and helpfulness	13	+.33 *
20.Teacher motivates students to do their best	3	+.33
21.Teacher's encouragement of self-initiated learning	1	-.52
22.Teacher's productivity in research	no cases	
23.Difficulty in course(description)	11	+.07
24.Difficulty in course(evaluation)	9	+.05
25.Classroom management	4	+.25
26.Pleasantness of classroom atmosphere	3	+.23
27.Individualization of teaching	1	+.10
28.Instructor pursued and/or met course objectives	2	+.46 *
29.Overall rating of lecturer as an item of multi-item indicator	1	+.36
30.Overall rating of teacher as an item of multi-item indicator	3	+.38 *
31.Overall rating of course as an item of multi-item indicator	no cases	

(* = significant two-tailed $p<0.001$)

TABLE A7.6

CONCLUSION	STUDIES ON WHICH CONCLUSIONS BASED
1.Different instructors teaching same course can be compared in terms of student performance on common exam, if classes similar in ability, prior knowledge and motivation.	Clark(1980)
2.Pre and post-course test score differences can be used to obtain an index of learning, but obvious weaknesses.	Clark(1980)
3.A pre-established number of students in a course who answer correctly a specified percentage of test items can be used as an indicator of student learning.	Clark(1980)

Table of generalisations on the use of achievement tests as an indicator of student learning (Braskamp et al 1985; table 4.11 p62).

Using construct validity requires the correlation of student ratings of a lecturer with other evaluations. The following tables summarise the conclusions on lecturer self-evaluation (table A7.7), classroom observations by outsiders (table A7.8), and alumni ratings (table A7.9).

TABLE A7.7

CONCLUSIONS	STUDIES ON WHICH CONCLUSIONS BASED
1.Students and self evaluation generally good reliability in agreement on overall ratings.	Blackburn and Clark(1975) Braskamp,Caulley and Costin(1979) Doyle and Crichton(1978) Marsh,Overall and Kesler(1979a)
2.Agreement between students and self evaluation on dimensions of student involvement, teacher support and instructional skill.	Braskamp,Caulley and Costin(1979) Marsh(1980)
3.Self ratings not influenced by age, sex, tenure status, teaching load, or years of teaching experience.	Doyle and Webber(1978)

Table showing a summary of the conclusions on the reliability of lecturer self-evaluation (Braskamp et al 1985; table 4.13 p71).

TABLE A7.8

CONCLUSIONS	STUDIES ON WHICH CONCLUSIONS BASED
1.An observer may affect teaching-learning process.	Fuller and Manning(1973)
2.Not reliable - no agreement with other methods on instructional effectiveness.	Centra1975)
3.The relationship between observed behaviour and student learning is not very strong.	Braskamp et al(1985)
4.Colleagues ratings not highly related to student ratings, if class time was well spent and instructor open to other viewpoints.	Centra(1975)
5.Agreement between colleagues and students on specific instructional practices. They agree on descriptions of activities, but not on their judgments of instructional quality.	Centra(1975)
6.Colleagues are more generous than students in their ratings.	Braskamp et al(1985)

Table showing a summary of the conclusions on the reliability of colleagues ratings of instruction (Braskamp et al 1985; table 4.12 p66).

TABLE A7.9

CONCLUSIONS	STUDIES ON WHICH CONCLUSIONS ARE BASED
1.Same students agree between course and 1 year after graduation.	Overall and Marsh(1979)
2.Alumni of 5 years and current students show good agreement on overall teaching effectiveness.	Centra(1974)
3.Alumni ratings lower than current students.	Overall and Marsh(1979)

Table showing a summary of the conclusions about the reliability of alumni ratings (Braskamp et al 1985; table 4.14 p74).

A7.4 Use of MTMM

Murphy and Davidshofer(1988) summarises three points that a test will possess as established effectively by MTMM.

> "1.Scores on the test will be consistent with scores obtained using other measures of the same construct.
> 2.The test will yield scores that are not correlated with measures that are theoretically unrelated to the construct being measured.
> 3. The method of mesurement employed by the test shows little evidence of bias"(p106).

In their original article, Campbell and Fiske proposed a series of rules to follow for evaluating convergent and discriminant validity.

> 1. The convergent validity coefficients should be statistically significant and sufficiently different from zero to warrant further examination of the validity.
> 2. The convergent validities should be higher than correlations between different traits assessed by different methods.
> 3. The convergent validities should be higher than correlations between different traits assessed by the same method.
> 4. The pattern of correlations between different traits should be similar for each of the different methods.
>
> (quoted in Marsh and Hocevar 1983 p233).

The above rules have been criticised. Firstly, over what constitutes a satisfactory result.
Secondly, the use of correlations based on observed variables to draw conclusions about underlying factors(Kenny and Kashy 1992 p165).
The ANOVA approach(Kavanaugh et al 1971) or the factor analysis approach(Jackson 1969) have been suggested separately to overcome the weaknesses of the MTMM matrix. However, there is not universal agreement, especially over which technique of factor analysis to use. Kenny and Kashy(1992) review a number of techniques with the MTMM matrix:

- equal loading model(Alwin 1974); all traits and methods in the matrix are allowed to correlate.
- correlated uniqueness model(Kenny 1979); no method factors created.
- fixed method model(Bock and Bargmann 1966); reducing one of method factors.

The authors conclude that all approaches using factor analysis have problems, and thus establishing convergent/discriminant validity is difficult.

In an earlier paper, Marsh and Hocevar(1983) compared ANOVA and confirmatory factor analysis(CFA); and recommended the latter as having

specific advantages for use in the MTMM matrix(p246-7).

A7.5. Multi-Section Courses or MTMM Design for Validation?

Attempts have been made to establish validity by using large multi-section courses, where different groups of students are presented the same material by different instructors. Ideally the following should happen - many sections to the course; random assignment of students to the sections; pre-test measures used; each section taught by separate instructors; the final examination graded externally; and common textbooks among the sections (Marsh 1984 p720). Validity is then assessed by correlating the student ratings in each section.
But this does not mean a perfect methodology: each section is usually small; the problem of the influence of presage variables, like initial student motivation; the lack of consistency in measure of course achievement and student ratings. In fact Marsh goes as far as to say that this design is inherently weak(1984 p721).

Abrami,d'Appollonia and Cohen(1990) take up the defence of this methodology by reanalysing 43 multi-section validity studies. They argue that the inconsistencies of past studies were due to lack of proper analysis, and which "lacked the sensitivity necessary to identify characteristics that explain a medium size effect on the relationship between ratings and achievement"(p230).

Abrami et al(1990) point out that over 40 studies have used multisection courses for validation of student ratings. The design is high in internal validity, allows for some control between sections, and a common examination reduces the influence of other factors. The examination score is high in external validity - it is a direct measure of effective teaching.

Marsh(1987) advocates MTMM designs, because of the greater control on threats to internal and external validity. But to show that correlations between student ratings and, for example, instructor self-ratings are adequate measures of instruction is the problem. "In our view, such measures help establish the validity of ratings as measures of teaching processes but not as measures of the products of instruction"(Abrami et al 1990 p221).

APPENDIX 8

SIMILARITIES IN RATINGS BETWEEN STUDENTS AND FACULTY

TABLE A8.1

Characteristic of teaching	Number of studies	Import stated by students (a)	Import stated by faculty (b)	Differ (b-a)
1.Stimulation	10	.28	.50	+.22
2.Enthusiasm	9	.32	.24	-.08
3.Knowledge	17	.28	.23	-.05
4.Expansive	9	.56	.50	-.06
5.Preparation	12	.27	.28	+.01
6.Clarity	8	.33	.39	+.06
7.Elocution	9	.47	.57	+.10
8.Sensitivity	7	.22	.27	+.05
9.Objectives	10	.63	.56	-.07
10.Materials	11	.46	.47	+.01
11.Materials	10	.54	.58	+.04
12.Outcome	4	.43	.51	+.08
13.Fairness	14	.45	.45	-
14.Personality	12	.64	.70	+.06
15.Feedback	9	.49	.53	+.04
16.Questions	15	.51	.48	-.03
17.Challenge	10	.58	.40	-.18
18.Respect	16	.39	.47	+.08
19.Availability	10	.37	.47	+.10
20 Motivates	7	.58	.45	-.13
21 Self-learn	4	.75	.51	-.24
22 Research	5	.91	.88	-.03

Table showing the similar between students and faculty on individual characteristics by standardised ranks found by Feldman (1988). Standardised ranks eg important in 20 of 37 studies = 0.5; smaller = more important. Difference: + = students more important. Characteristics in detail in Appendix 4.

Feldman(1988) finds similarities between faculty and student ratings in the particular teaching characteristics of "enthusiasm"; "knowledge"; "expansive"; and "preparation". But students rate higher "stimulation".

APPENDIX 9

CLASS SIZE AND INDIVIDUAL CHARACTERISTICS OF TEACHING EFFECTIVENESS

TABLE A9.1

CHARACTERISTIC OF TEACHING	POSITIVE CORR.	NEGATIVE CORR.	NO CORR.	OTHER RELATION-SHIP
1.Stimulation	1	8	12	1
2.Enthusiasm	1	3	7	1
3.Knowledge	1	3	5	0
4.Expansive	0	4	4	1
5.Preparation	3	11	12	3
6.Clarity	1	9	14	2
7.Elocution	0	5	3	1
8.Sensitivity	0	6	1	1
9.Objectives	0	5	8	2
10.Materials	0	11	11	3
11.Materials	0	5	3	1
12.Outcome	2	8	3	2
13.Fairness	1	17	8	2
14.Personality	2	1	0	0
15.Feedback	0	7	3	1
16.Questions	1	19	5	3
17.Challenge	1	11	4	2
18.Respect	1	14	2	3
19.Availability	0	15	5	3

Table summarising the number of studies found by Feldman(1984),showing the relationship between class size and different characteristics of teaching.
Full details of the characteristics in Appendix 4.

TABLE A9.2

Individual Characteristics of Ideal Lecturer	Relationship between individual characteristic and class size
1.Stimulation of interest	No corr.
2.Enthusiasm	No corr.
3.Knowledge of subject	No corr.
4.Intelligence	No corr.
5.Preparation/ organisation	Negative corr.
6.Clarity	No corr.
7.Elocutionary skills	Negative corr.
8.Class level	Negative corr.
9.Course objectives	No corr.
10.Practical	Negative corr.
11.Use of aids	Negative corr.
12.Perceived outcome	Negative corr.
13.Fairness	Negative corr.
14.Personality	Positive corr.
15.Feedback	Negative corr.
16.Encourages questions	Negative corr.
17.Encourage independent thought	Negative corr.
18.Respect	Negative corr.
19.Availability	Negative corr.

Table showing the most common relationship between class size and individual characteristics of teaching, as found by Feldman (1984). (Full details on characteristics in Appendix 4).

APPENDIX 10

CORRELATION OF INDIVIDUAL ITEMS WITH OVERALL EVALUATION OF TEACHING EFFECTIVENESS

TABLE A10.1

Individual Characteristics of Ideal Lecturer	Correlation of individual charactieristic with overall evaluation
1.Stimulation of interest	+.20
2.Enthusiasm	+.46
3.Knowledge of subject	+.48
4.Intelligence	+.54
5.Preparation/ organisation	+.41
6.Clarity	+.25
7.Elocutionary skills	+.49
8.Class level	+.40
9.Course objectives	+.45
10.Practical	+.70
11.Use of aids	+.72
12.Perceived outcome	+.28
13.Fairness	+.72
14.Personality	-
15.Feedback	+.87
16.Encourages questions	+.60
17.Encourage independent thought	+.39
18.Respect	+.65
19.Availability	+.74

Table showing the correlation with overall evaluation of individual characteristics of teaching, as found by Feldman (1976b). (Full details of characteristics in Appendix 4).

APPENDIX 11

Which Instrument to Measure Student Ratings?

Here are a selection of the most commonly used instruments for rating teacher effectiveness:

1.PURDUE RATING SCALE FOR INSTRUCTION (PRSI) (Remmers 1960). One of the first student evaluation instruments, originally from the 1920s. Students can rate their lecturers on 10 traits, an overall rating, and 15 aspects of the classroom situation.

2.ILLINOIS COURSE EVALUATION QUESTIONNAIRE (CEQ) (Aleamoni 1972) Based on a comprehensive original 1000 items; reduced to 450 items during the pilot study. Finally 50 items, which with factor analysis produced 6. Aleamoni and Spencer(1973) reduced to 23 items with 5 subscores from factor analysis.

3.COURSE PERCEPTIONS QUESTIONNAIRE (CPQ) (Entwistle and Ramsden 1983).Originally 47 items which were developed to produce 8 dimensions: relationship with students; commitment to teaching; workload; formal teaching methods; vocational relevance; social climate; clear goals and standards; freedom in learning(p124). The authors feel it provides "a useful means of describing certain important and consistent differences in the way students perceive departments"(p129).
Entwistle and Tait(1990) report factor analysis of the 8 dimensions to produce two factors, but "workload" left isolated.

4.ENDEAVOUR INSTRUCTIONAL RATING CARD (Frey,Leonard and Beatty 1975;Frey 1976). Frey(1978) reduced 7 items to two higher order factors: "pedagogical skill" and "rapport".
The former is defined as ".1 x work + .4 x plan - .2 x disc + .3 + .4 x know." Rapport defined as "-.2 x plan + .5 x disc + .5 x help + .2 x grade."(p77).

1.Work = The student had to work hard in this course.
2.Plan = Each class period was carefully planned in advance.
3.Disc = Class discusion was welcome in this course.
4.Help = The student was able to get personal help in this course.
5.Pres = The instructor presented the material clearly and summarised major points.
6.Grade = The grading accurately refelected the student's performance.
7.Know = This course has increased my knowledge and competence in this area.

However, he has to admit that both factors are not equally sound.

5.STUDENTS' EVALUATION OF EDUCATIONAL QUALITY (SEEQ) (Marsh 1982a). Probably the most popular rating used, and a lot of effort has been put in, by Marsh in particular, to establish its reliability and validity etc. Marsh reports that 250 000 SEEQ forms have been completed at the University of Southern California alone(1982a p78).
31 items are reduced to 9 distinct components: learning/value; enthusiasm; organisation; group interaction; individual rapport; breadth of coverage; examination/grading; assignments; workload/difficulty. Further research is reported in Marsh(1984;1987), and most recently, Marsh and Hocevar(1991).

6.COMPLUTENSE UNIVERSITY TEACHERS EVALUATION QUESTIONNAIRE (CUTEQ-R) (Fernandez and Mateo 1992). The most recent development, from Universidad Complutense in Spain; revision of CUTEQ (Angulo et al 1987). 39 items originally, yielding two factors: teaching competence and motivational skills, but accounting for only 71% of total variance. Used on nearly 200 000 students.

7.COLLEGE STUDENT SATISFACTION QUESTIONNAIRE (CSSQ) (Betz et al 1970;1971). Based on job-satisfaction research, it measures 6 dimensions of college satisfaction. Policies and procedures (eg freedom to choose classes); working conditions (physical conditions of college); compensation (amount of study related to grades); quality of education (academic conditions); social life; recognition (attitudes of faculty to students). Originally 139 items reduced to 6 areas.

8.INSTRUCTOR-DESIGNED QUESTIONNAIRE (IDQ) (Kulik 1991)
Produced at the University of Michigan, there are 154 items, but the lecturer chooses only the items they feel are relevant. This solves the problems of ratings, according to Kulik - many items are too general or not relevant, and many teachers feel distanced from the evaluation process.

Most of the instruments produce similar characteristics for teaching effectiveness to be rated on. The table A11.1 shows how the characteristics of Feldman's work, SEEQ, and Endeavour are related.

TABLE A11.1

FELDMAN	SEEQ	ENDEAVOUR
Stimulation of interest	Instructor Enthusiasm	None
Enthusiasm	Instructor Enthusiasm	None
Knowledge	Breadth of coverage	None
Intellectual expansion	Breadth of coverage	None
Preparation	Organisation/clarity	Organisation/planning
Clarity	Organisation/clarity	Presentation clarity
Elocution	None	None
Sensitivity to class	None	None
Clarity	Organisation/clarity	Organisation/planning
Value	Assignments/reading	None
Supplementary materials	Assignments/reading	None
Perceived outcome	Learning/value	Student accomplishments
Fairness	Exams/grading	Grading/exams
Classroom management	None	None
Feedback	Exams/grading	Grading/exams*
Class discussion	Group Interaction	Class discussion
Intellectual challenge	Learning/value	Student accomplishments*
Respect	Individual Rapport	Personal attention
Availability	Individual Rapport	Personal attention
Difficulty	Workload/difficulty	Workload

* = match not particularly close.
Feldman's characteristics are described in more detail in the Appendix 4.

Table showing the categories of effective teaching adapted from Feldman(1976,1983,1984) and the SEEQ and Endeavour factors most closely related to each category (Marsh and Roche 1992; table 1 p283).

APPENDIX 12

Does Feedback Change the Teaching Performance?

It is generally felt that student feedback should improve teaching performance. Tuckman and Oliver(1968) show that high school teachers who received feedback were later rated higher than those teachers who did not receive feedback. But Miller(1971) found no significant difference in the end of term ratings between those teaching assistants who received mid-term feedback and those who did not.

Remmers(1959) makes two important points:

> "1.Knowledge of student opinions and attitudes leads to improvement of the teacher's personality and educational procedures.
> 2.Students are more favourable to student ratings than instructors, but more instructors have noticed improvement in their teaching as a result of student ratings than the studies have"
>
> (Quoted in Flood Page 1974 p68).

Wilson(1986) details a scheme to improve faculty teaching at the University of California, Berkeley. Improvement was found on 9 characteristics of half of the lecturers due to feedback (consultation with lecturer using student comments).

Marsh(1987) reviews the two main type of studies aimed at answering the question of the effect of feedback on teaching.

i) Short term feedback studies - generally it is felt that feedback and consultation can improve teaching. Cohen's(1981) meta-analysis of feedback studies found that instructors receiving mid-term feedback were rated higher than the control group on overall rating. L'Hommedieu et al(1990) point out the problem of the "John Henry effect", ie teachers who know they are being rated tried to improve their teaching.

ii) Long term feedback studies - Marsh feels that there are so few studies, and many problems with such research, that it is difficult to reach a conclusion.

In a recent study, Marsh and Roche(1993) looked at the effect of feedback from students and consultation mid-term and end of term, on the evaluation at the end of the course. The ratings improved for both groups, but only ratings for the end of term group improved significantly more than the control group (no feedback). The authors conclude that "SET (student evaluation of teaching) feedback coupled with consultation is an effective means to improve teaching effectiveness"(p 217).

Ryan et al(1980) looked at SRI from the point of morale of the faculty. Over 90% of the 193 academics felt morale had "greatly or somewhat decreased" through the use of student ratings of their teaching. Furthermore, nearly 45% reported a decrease in job satisfaction, and over 70% a decrease in their confidence in the administration. Many academics admit their behaviour has changed to some extent because of the ratings. But the authors are sceptical of the benefits, suggesting that the most frequently reported change was a reduction in coursework demands on the students. On most behaviours, "no change" was the most often response. The full results to the questionnaire appear in the table A12.1

TABLE A12.1

ACTIVITY	Greatly or somewhat increased(%)	No Change(%)	Greatly or somewhat decreased(%)
Explicit specification of course objectives	39.9	56.0	1.6
Use of group discussion as an instructional method	16.6	74.6	6.2
Use of lecture as an instructional method	6.7	73.6	16.1
Provision of handouts and other course aids	31.6	63.7	2.1
Use of audio or visual aids	22.3	73.1	2.1
Prompt return of exams and papers	20.2	74.6	2.1
Content of instructional input	13.0	68.9	13.0
Amount of material covered	6.7	68.4	22.3
Degree of relationship to student interest	26.9	65.3	3.1
Difficulty level	9.3	50.3	37.8
Attention to organisation of course content	30.1	63.7	3.1
Response to student questions in class	13.0	82.4	1.6
Availability to students outside class	13.5	81.3	2.6
Informal social interaction with students outside class	10.9	79.3	6.7
Use of relative(in contrast to absolute) basis for grading	19.2	69.4	8.3

Table showing responses to the question "How has information obtained from student evaluation of your classes in itself changed or modified your instructional activities?" in Ryan et al(1980) study (Ryan et al 1980; table 3 p323).

APPENDIX 13

RATING ERRORS ON STUDENT RATINGS OF INSTRUCTION

A13.1 Rating Error

All ratings contain an element of measurement error. Forced-choice scales are an attempt to reduce this. The rater must choose, for example, two items from a list of 4 equally desirable. Sharon(1970) found this type of rating did not differ across four conditions, while a usual scale did.
But these scales have been criticised as difficult for raters, among other problems(Doyle 1975).

Research has tried to identify student characteristics that could bias ratings of teachers (reviews by Feldman 1977;1978;1979). When correlations between characteristics and ratings are large, there is seen to be bias, and the ratings lack validity. But Marsh(1987) argues that validity is lost only when "biasing" characteristics influence the ratings, and not the instructional effectiveness criteria at the same time, and vice versa.

A13.2 Bias in Student Ratings

Two areas of bias that have particularly concerned researchers are thr effect of implicit theories ("halo effect"), and the semantic similarity of items.

A13.2.1 Implicit Theories

This is the idea that if the raters notice certain characteristics in the teacher, then they assume the teacher must also have certain others. Whitely and Doyle(1976) feel that students' implicit theories influence their rating of teaching. Using latent partition analysis (Wiley 1967), they identified latent clusters of behaviour, when the unit of analysis was total-class, within-class or between-class ratings. But the authors feel that the implicit theories are based on experience of teaching, and so are quite accurate to which behaviours are associated together. Larson(1979) is critical: "we still have no way of knowing whether a particular set of behaviour ratings reflects the actual behaviour of those being rated or whether they reflect population based normative assumptions about these behaviours"(p210).

More recently, Widmeyer and Loy(1988) replicated Kelley's(1950) "first impressions" experiment finding that subjects who were told that the lecturer had a warm personality rated them as a more effective teacher, than subjects who were told the lecturer had a cold personality. The "warm personality" was also seen as less unpleasant, more sociable, less irritable, less ruthless, more

humourous, less formal and more humane(p119).Full details in table A13.1.

TABLE A13.1

Item:Teaching ability	"Warm" Group	"Cold" Group	Signif-icance
Knows his material - doesn't	1.44	1.65	.05
Considerate of class - self centred	1.90	2.24	.01
Intelligent - unintelligent	1.59	1.84	.05
Organised - not	1.77	1.73	NS
Expresses himself well - difficulty	2.09	2.10	NS
Interesting - boring	2.05	2.32	.05

7 point scale used: 1 = left hand end to 7 = right hand end.
Table showing mean ratings given to a stimulus person designated warm or cold (Widmeyer and Loy 1988 p120).

Marsh(1987) points out that implicit theories can be ruled out by establishing a factor structure similar to that of SRI from another method. For example, the use of lecturer's self evaluation. This method suffers from little or no "halo effect", while colleagues' evaluations may suffer most.

A13.2.2 Semantic Similarity of Items

This is slightly different to the implicit theories, in that the raters score items because they appear to be similar to other items. For example, lecturers who are "friendly towards individual students" will be assumed to also make "students feel welcome in seeking help/advice". Cadwell and Jenkins(1985) found evidence of this process using a hypothetical instructor profile with 28 graduate students. They explain the cognitive processes involved in responding to a SRI, which by its nature must lead to this bias. "Thus, because student ratings are the product of cognitive processes that reconstruct rather than mirror instructor behaviour, these ratings, like all personality assessments, lead us to view behaviour as more organised and more consistent than it actually is"(p392). This study has been criticised at length (Marsh and Groves 1987), particularly because of the use of a "hypothetical profile".

Again, this bias can be eliminated by construct validation.

A13.2.3 Other Bias

1."SIMPLISTIC BIAS HYPOTHESIS". This states that if an instructor gives high grades, demands little work, or teaches only small classes, they will receive a higher rating. Marsh(1987) quotes his own earlier research, which he believes clearly refute this, and show it to be a "strawman"(p310). The use of the Student Educational Evaluation Questionnaire (SEEQ), and multidimensions to the ratings, reduces the possibility of a global item influenced by the above factors. Furthermore, the dimension of Workload/Difficulty was opposite to this "hypothesis".

2.LENIENCY ERRORS. The tendency to rate generously for those people the rater is involved with. Centra(1975) found that colleagues' ratings (mean of 4.47 out of 5) of teaching effectiveness was one standard deviation higher than students'(mean of 3.98). Other studies have found slightly different results. Doyle(1983) feels that "some degree of leniency error can be expected in most evaluation"(p75), but it is higher for colleagues' evaluations.

APPENDIX 14

EXAMPLE OF FACTOR ANALYSIS ON DATA FROM QUESTIONNAIRE AT UNIVERSITY OF BIRMINGHAM

The SPSS default method (principal components with varimax rotation) is typical of the other methods used. Factors are produced, but it very hard to name them. For example, factor I includes "provides adequate feedback with helpful/constructive comments(N)","aware when students having difficulty in understanding(Q)", and "welcomes questions and comments from students(U)" which logically could be called "relationship with students", but also includes a negative "explains concepts,principles and abstract theories clearly and precisely(H)".

Items	I	II	III	IV	V	VI	VII	VIII	IX
Vocabulary	-77								
Feedback	68								
Aware	59								-31
Enthusiasm		85							
Handouts		-66							
Thought			82						
Express			73						
Knowledge				-76					
Prepared			-30	-62			39		
Structures		-33		55					
Relates					75				
Questions					-64				
Anecdotes	35				60	37			
Explains						-78			
Motivates						68			
Developments							65		
General knowledge		-31		-33			-63		
Humour							-51		43
Objectives								-79	
Organised								57	47
Accurately		-39	-30					55	
Clearly									74

Rotated factor matrix on SPSS default factor analysis. The item letters relate to characteristics in final questionnaire, and are described in detail in appendix 2.

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
		*				
A	1.00000	*	1	2.76313	12.6	12.6
B	1.00000	*	2	2.12427	9.7	22.2
C	1.00000	*	3	1.83352	8.3	30.5
D	1.00000	*	4	1.70346	7.7	38.3
E	1.00000	*	5	1.45696	6.6	44.9
F	1.00000	*	6	1.44736	6.6	51.5
G	1.00000	*	7	1.14111	5.2	56.7
H	1.00000	*	8	1.12950	5.1	61.8
I	1.00000	*	9	1.01354	4.6	66.4
J	1.00000	*	10	.95689	4.3	70.8
K	1.00000	*	11	.90496	4.1	74.9
L	1.00000	*	12	.79706	3.6	78.5
M	1.00000	*	13	.73020	3.3	81.8
N	1.00000	*	14	.66445	3.0	84.8
O	1.00000	*	15	.60851	2.8	87.6
P	1.00000	*	16	.57144	2.6	90.2
Q	1.00000	*	17	.52653	2.4	92.6
R	1.00000	*	18	.48254	2.2	94.8
S	1.00000	*	19	.44578	2.0	96.8
T	1.00000	*	20	.36892	1.7	98.5
U	1.00000	*	21	.29702	1.4	99.9
V	1.00000	*	22	.03285	.1	100.0

APPENDIX 15

CORRELATION MATRIX (using Pearsons Product Moment Test)

List of characteristics found on p148.

Correlations:	A	B	C	D	E	F
A	1.0000	.0180	.1154	.0453	.1631	-.1545
B	.0180	1.0000	-.1633	.1581	.0141	-.0101
C	.1154	-.1633	1.0000	-.0691	-.0632	.2095*
D	.0453	.1581	-.0691	1.0000	-.0987	.2871**
E	.1631	.0141	-.0632	-.0987	1.0000	-.2572*
F	-.1545	-.0101	.2095*	.2871**	-.2572*	1.0000
G	-.0813	.0216	-.1388	.1931	-.0816	.0823
H	.0171	.0117	-.0089	.0933	.1526	.0581
I	-.0589	.0836	-.1192	-.0887	.0456	-.1044
J	-.3105**	.0060	-.0636	-.2365*	-.1607	.0806
K	.2008	.0315	.0425	-.0861	.1040	-.0042
L	.1101	-.0307	-.0324	.0062	.0951	-.0898
M	-.2133*	-.0511	-.2041	-.2119*	-.1542	-.2035
N	-.0817	-.0186	-.2642*	-.0088	-.0206	-.1813
O	-.0097	-.1378	.0231	-.0953	.0035	-.1883
P	-.0016	-.1912	-.0014	-.0131	-.1259	-.1174
Q	-.1572	-.0932	-.1486	-.0789	-.0160	-.2299*
R	-.1133	-.1688	-.1184	-.1988	-.0926	-.2227*
S	-.3926**	-.0977	-.1431	-.0783	-.3385**	.0937
T	-.1275	-.2277*	.0737	-.2911**	-.1695	-.0922
U	-.0758	-.0790	-.2113*	-.1688	-.0398	-.1686
V	.0267	-.0817	.2352*	-.0800	-.0275	.1387

	G	H	I	J	K	L
A	-.0813	.0171	-.0589	-.3105**	.2008	.1101
B	.0216	.0117	.0836	.0060	.0315	-.0307
C	-.1388	-.0089	-.1192	-.0636	.0425	-.0324
D	.1931	.0933	-.0887	-.2365*	-.0861	.0062
E	-.0816	.1526	.0456	-.1607	.1040	.0951
F	.0823	.0581	-.1044	.0806	-.0042	-.0898
G	1.0000	.0736	.0024	-.1440	.1318	-.2549*
H	.0736	1.0000	.0023	-.0950	.0123	-.0279
I	.0024	.0023	1.0000	.1419	-.1462	.0892
J	-.1440	-.0950	.1419	1.0000	-.1705	-.1547
K	.1318	.0123	-.1462	-.1705	1.0000	-.2151*
L	-.2549*	-.0279	.0892	-.1547	-.2151*	1.0000
M	-.1378	-.1625	-.0554	.4135**	-.2776**	-.0707
N	-.0318	-.0016	-.0264	.0246	.0449	-.1285
O	-.0056	-.2436*	.2696*	-.0619	-.1177	-.0413
P	.0935	-.1593	-.2117*	-.1664	.0552	.0383
Q	-.0622	-.0316	-.1518	-.0622	-.0873	-.0149
R	-.0311	-.2969**	-.0709	.1923	-.0745	-.1682
S	-.1391	-.1184	.0225	.0231	-.2628*	-.1012
T	-.0125	-.1117	-.1477	-.0092	-.1761	-.1254
U	-.0927	-.1492	-.2807**	.0775	-.0586	-.1290
V	-.1134	-.0021	-.0462	-.1129	.0150	.0223

	M	N	O	P	Q	R
A	-.2133*	-.0817	-.0097	-.0016	-.1572	-.1133
B	-.0511	-.0186	-.1378	-.1912	-.0932	-.1688
C	-.2041	-.2642*	.0231	-.0014	-.1486	-.1184
D	-.2119*	-.0088	-.0953	-.0131	-.0789	-.1988
E	-.1542	-.0206	.0035	-.1259	-.0160	-.0926
F	-.2035	-.1813	-.1883	-.1174	-.2299*	-.2227*
G	-.1378	-.0318	-.0056	.0935	-.0622	-.0311
H	-.1625	-.0016	-.2436*	-.1593	-.0316	-.2969**
I	-.0554	-.0264	.2696*	-.2117*	-.1518	-.0709
J	.4135**	.0246	-.0619	-.1664	-.0622	.1923
K	-.2776**	.0449	-.1177	.0552	-.0873	-.0745
L	-.0707	-.1285	-.0413	.0383	-.0149	-.1682
M	1.0000	.2249*	.0297	-.1733	.0649	.1935
N	.2249*	1.0000	.1036	-.1713	.1687	.1157
O	.0297	.1036	1.0000	-.0544	.0418	.1236
P	-.1733	-.1713	-.0544	1.0000	.0634	-.0468
Q	.0649	.1687	.0418	.0634	1.0000	-.1405
R	.1935	.1157	.1236	-.0468	-.1405	1.0000
S	.0736	-.0569	-.0542	-.0709	.0992	.1433
T	.1482	-.1911	.0864	.1172	-.0074	.0560
U	.2351*	.0906	-.1362	-.0255	.1731	.1124
V	-.2040	-.4050**	-.2017	.0852	-.2917**	-.0111

	S	T	U	V
A	-.3926**	-.1275	-.0758	.0267
B	-.0977	-.2277*	-.0790	-.0817
C	-.1431	.0737	-.2113*	.2352*
D	-.0783	-.2911**	-.1688	-.0800
E	-.3385**	-.1695	-.0398	-.0275
F	.0937	-.0922	-.1686	.1387
G	-.1391	-.0125	-.0927	-.1134
H	-.1184	-.1117	-.1492	-.0021
I	.0225	-.1477	-.2807**	-.0462
J	.0231	-.0092	.0775	-.1129
K	-.2628*	-.1761	-.0586	.0150
L	-.1012	-.1254	-.1290	.0223
M	.0736	.1482	.2351*	-.2040
N	-.0569	-.1911	.0906	-.4050**
O	-.0542	.0864	-.1362	-.2017
P	-.0709	.1172	-.0255	.0852
Q	.0992	-.0074	.1731	-.2917**
R	.1433	.0560	.1124	-.0111
S	1.0000	.1874	.0930	.0939
T	.1874	1.0000	-.0381	.0511
U	.0930	-.0381	1.0000	-.0754
V	.0939	.0511	-.0754	1.0000

1-tailed Signif: * - .01 ** - .001