

Durham E-Theses

*How university students rate their teachers : a study
of the attitudes and rating behaviours of university
students in teaching evaluations.*

Kam-Por Kwan

How to cite:

Kwan, Kam-Por (2000) How university students rate their teachers : a study of the attitudes and rating behaviours of university students in teaching evaluations. Doctoral thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a <https://etheses.durham.ac.uk/id/eprint/1606/> is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

**HOW UNIVERSITY STUDENTS RATE THEIR TEACHERS:
A STUDY OF THE ATTITUDES AND RATING BEHAVIOURS OF
UNIVERSITY STUDENTS IN TEACHING EVALUATIONS**

The copyright of this thesis rests with the author. No quotation from it should be published in any form, including Electronic and the Internet, without the author's prior written consent. All information derived from this thesis must be acknowledged appropriately.

KAM-POR KWAN

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF EDUCATION

SCHOOL OF EDUCATION

UNIVERSITY OF DURHAM

2000



17 AUG 2001

STATEMENT OF COPYRIGHT

The copyright of this thesis rests with the author. No quotation from it should be published without their prior consent and information derived from it should be acknowledged.

How University Students Rate Their Teachers: A Study of the Attitudes and Rating Behaviours of University Students in Teaching Evaluations

Thesis submitted for the Degree of Doctor of Education, University of Durham, 2000.

Kam-Por KWAN

ABSTRACT

Despite the voluminous research into student evaluations of university teaching, little is known about what goes on in students' minds when they fill out the evaluation forms. This study aimed to examine how university students go about rating their teachers. Seventy-one students from six different departments of a university in Hong Kong were interviewed about their perceptions of the purposes and impact of the evaluation, as well as their attitudes about the rating process. Students were also asked to identify a 'good' and a 'poor' teacher they had encountered, rate them separately and anonymously on six items drawn from the University's student feedback form, and report verbally how they determined the specific ratings. Results indicated that: (a) most students demanded to have a chance to evaluate their teachers, but for different reasons; (b) students had different perceptions of the purposes and impact of teacher evaluations, which were related to their overall reactions to the exercise; (c) students varied in their attitudes towards the rating process, which also changed according to the circumstances and contexts under which the evaluations were conducted; (d) students used different strategies to determine their ratings for teachers; (e) students generally based their ratings on identifiable and sensible criteria, but there were variations in the evidence and standards that students used for making their ratings; and (f) students differed widely in their interpretations of the rating values they gave for their teachers. The implications of the findings were discussed, and directions for future research were also suggested.

TABLE OF CONTENTS

STATEMENT OF COPYRIGHT	i
ABSTRACT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
ACKNOWLEDGMENTS	ix
DECLARATION	x
CHAPTER 1 INTRODUCTION	1
The research questions	2
CHAPTER 2 REVIEW OF LITERATURE	4
2.1 The controversy over student evaluations of university teaching	5
Arguments for the use of student evaluations	6
Arguments against the use of student evaluations	7
2.2 The reliability, validity, and utility of student evaluations	9
Reliability of student ratings	9
Validity of student ratings and the question of bias	10
Effect of student evaluation feedback on teaching improvements	19
Problems with prior validity studies of student evaluations	20
The role of the students in the evaluation process	22
CHAPTER 3 METHODS	28
3.1 The University and the student feedback system	28
The student feedback system	28
The Student Feedback Questionnaire (SFQ)	30
3.2 Methodology	30
The naturalistic inquiry	31
The interview as a research technique	33
Participants	36
The interview procedures	37

The interview protocol	37
The verbal report method	38
Analysis of the interview data	40
3.3 Limitations and delimitation	40
CHAPTER 4 STUDENTS' PERCEPTIONS OF TEACHER EVALUATION	44
4.1 Students' perceptions of the purposes of teacher evaluation	44
A channel for collecting feedback	45
A quality control measure	45
A marketing strategy	46
A bureaucratic formality	46
4.2 Students' beliefs about their roles in teacher evaluation	46
4.3 Students' perceptions of the impact of teaching evaluation	49
4.4 Students' overall reactions to teacher evaluation	50
CHAPTER 5 STUDENTS' ATTITUDES ABOUT THE EVALUATION PROCESS	53
5.1 Seriousness of the students in the rating process	53
The meaning of being 'serious' in evaluation	54
Contextual influences on students' attitudes	57
5.2 Factors affecting students' attitudes in the evaluation process	59
Evaluation arrangements	59
Students' perceptions of the subject and the teaching	60
Knowledge of the purpose, procedure, and use of the evaluation	61
Perceptions of the benefits of evaluations	61
Differential perceptions among students	62
5.3 Students' anxieties about teacher evaluation	64
CHAPTER 6 STUDENTS' STRATEGIES FOR MAKING RATINGS	67
6.1 Student ratings for 'good' and 'poor' teachers	67
6.2 General strategies for deriving the ratings	73
Global impression versus item-specific considerations	73
Other lecturers' performance versus students' own standards	75
Self experience versus other students' reactions	76
Using versus avoiding extreme scores	77

6.3	Why students avoid giving the minimum rating value	78
6.4	Peer influence on student ratings	80
CHAPTER 7 CRITERIA AND EVIDENCE THAT STUDENTS USED FOR MAKING RATINGS FOR INDIVIDUAL ITEMS		84
Item 1:	The staff member's method of teaching has helped my understanding	84
Item 2:	The staff member encouraged active participation in class	88
Item 3:	The staff member provided appropriate help for students with learning difficulties	92
Item 4:	The staff member presented the subject materials clearly	96
Item 5:	The staff member's teaching stimulated my interest in the subject	98
Item 6:	The staff member gave me regular feedback on my progress	103
CHAPTER 8 MEANINGS UNDERLYING THE NUMERICAL VALUES OF STUDENT RATINGS		107
8.1	Students' interpretations of the rating values	107
8.2	Standards for making ratings	110
8.3	Student ratings and conceptions of 'good teaching'	112
8.4	Halo effects on student ratings	114
8.5	Subjectivity of student ratings	116
CHAPTER 9 SUMMARY AND DISCUSSION		118
9.1	Students' perceptions of teaching evaluation	118
9.2	Attitudes of the students about the rating process	120
9.3	Students' general strategies for making ratings	121
9.4	The criteria and standards that students use for making ratings	122
9.5	Implications for developing student evaluation systems	126
9.6	Implications for using student ratings for formative evaluation of teaching	127
9.7	Implications for using student ratings for summative evaluation of teaching	130
9.8	Appropriate interpretation and use of student ratings	132
9.9	Conclusions	134

Suggestions for further research	135
APPENDICES	137
Appendix A The Student Feedback Questionnaire Form	137
Appendix B Letter to Chairpersons of Departmental Learning & Teaching Committees	139
Appendix C The Interview Protocol	140
Appendix D Questionnaire for Rating of 'Good' and 'Poor' Teachers	142
REFERENCES	143

LIST OF TABLES

Table 1:	Contributions of each of the project team members to the study	xi
Table 2:	Relationship between student ratings and administrative procedures	15
Table 3:	Relationship between student ratings and course characteristics	15
Table 4:	Relationship between student ratings and teacher characteristics	16
Table 5:	Relationship between student ratings and student characteristics	17
Table 6:	Descriptives of the ratings for the ‘good’ and the ‘poor’ teachers	68
Table 7:	Distribution of the paired differences in ratings	70
Table 8:	Results of the paired <i>t</i> -tests of differences in ratings	71
Table 9:	Criteria/evidence that students used in making ratings for Item 1	85
Table 10:	Criteria/evidence that students used in making ratings for Item 2	89
Table 11:	Criteria/evidence that students used in making ratings for Item 3	93
Table 12:	Criteria/evidence that students used in making ratings for Item 4	96
Table 13:	Criteria/evidence that students used in making ratings for Item 5	99
Table 14:	Criteria/evidence that students used in making ratings for Item 6	103

LIST OF FIGURES

- Figure 1: Axiomatic differences between the Rationalistic and Naturalistic Paradigms 32
- Figure 2: Distribution of the ratings for the 'good' and the 'poor' teachers 69

ACKNOWLEDGEMENTS

I would like to express my most sincere thanks to Professor Carol T. Fitz-Gibbon, my thesis supervisor, for her consistent reassurance, unfailing support, critical comments, and expert advice throughout the process of the study.

I wish to extend my heartfelt gratitude to Dr. John Jones, Director of Educational Development, Educational Development Centre of the Hong Kong Polytechnic University, for his invaluable support and encouragement both for my work and for my study. I am also indebted to Ms. Louisa Yan, a dear colleague of mine, for her excellent support as temporary Research Fellow of the Project on which this thesis is based. Her insights and hard work have been invaluable to the successful completion of the study. I owe a special debt to Dr. Jeff James for his many helpful editorial suggestions on an earlier draft of the manuscript.

I wish to thank the support of the colleagues from the six departments of the Hong Kong Polytechnic University, for their help in identifying and encouraging students to participate in the study. I am also grateful to the Learning & Teaching Committee of the Hong Kong Polytechnic University, for funding this study.

Special thanks are also due to the 71 students who took part in the study, for their generosity in sharing their views and patience in explaining to us how they determined ratings for their teachers. Without their support, the research could never have been completed.

Finally, my gratitude goes to my family – my wife, Helen, and our daughter, Bonnie, and our son, Billy. Their total support, understanding, and love have made it possible for me to progress through this long, sometimes lonely, and often challenging process.

DECLARATION

The work submitted in this thesis is part of the results of a funded teaching development project that I carried out in the period March 1999 to May 2000, in the university where I am currently working. I am the Leader of the Project Team, which consisted of one other full-time member of the University and a temporary Project Research Fellow. While I have been greatly benefited by the invaluable contributions and inputs of my colleagues in carrying out the Project, I honestly declare that the work submitted is primarily the results of my own work, as I have taken the lead and played a primary role in every stage of the Project, from the initial stage of formulating the ideas for the study to the final stage of writing up the thesis. The individual contributions of each of the members of the Project Team were briefly delineated in Table 1 on the next page to support my claim. I further declare that the material in this thesis has not been previously submitted for a degree in the University of Durham or any other university.

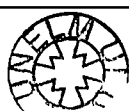
Table 1: Contributions of each of the project team members to the study

Tasks	The Author (Project Leader)	Collaborator	Temporary Research Fellow
Formulating research questions and design	First proposed the study and sought collaborator	--	--
Writing up the research proposal	Drafted, discussed, and modified proposal based on feedback	Commented on the drafts	--
Devising plan for implementation	Drafted implementation plan and set target dates	Commented on the plan	--
Selecting students to participate	Devised sampling method, criteria, and size	Sent letters to departments to invite students to participate	--
Developing the interview schedule	Drafted the interview protocol, and modified according to comments	Commented on the interview protocol	--
Conducting the interviews	Conducted 39 student interviews	--	Conducted 32 student interviews
Translating & transcribing interview records	Translated and transcribed 45 interview records	--	Translated & transcribed 26 interview records
Analysis of the interview data	Took primary responsibility for analysing and interpreting the results	--	Contributed to initial identification of themes and categories
Verifying the interpretation and analysis	--	--	Commented on the credibility of the interpretations, and conducted a confirmatory audit of the interpretations
Writing up of the project report	Solely responsible for writing up the project reports	--	--
Writing up of this thesis	Solely responsible for writing up the thesis	--	--

CHAPTER 1 INTRODUCTION

Student ratings have become an increasingly important source of information for evaluating teaching quality in higher education. In many universities, they are being used as one, and often the only, evidence for judging the teaching quality of lecturers, courses and departments. Given the momentous influence of student evaluations on personnel decisions and thus the careers of the academics, it is not surprising to find that there has been a revived concern over the reliability and validity of student ratings in the last few years (e.g., Greenwald, 1997; Kerridge & Mathews, 1998; Newport, 1996; Williams & Ceci, 1997).

Numerous studies have been carried out to examine the reliability and validity of student evaluations of university teaching (SETs). The general conclusion is that SETs are reasonably reliable and valid, and relatively free from bias (see, for example, the reviews by Cashin, 1995; d'Apollonia & Abrami, 1997; Marsh, 1987; Marsh & Roche, 1997). However, most of the studies to date have been quantitative in nature and focused exclusively on the procedures and instrumentation of student evaluations. Little research has been done to examine the attitudes and behaviours of the students who are directly involved in the rating process. Furthermore, while some research has been carried out to investigate the relationship between student characteristics and ratings, few studies have examined how students' ratings are mediated by the role they play in, or the meanings they ascribe to the evaluation exercise. Reflecting upon over six decades of research in student evaluations of teaching, Feldman (1998) points out that the role of the students in the rating process is an important area where much more research is needed.



On the practical side, all student evaluation systems operate under the assumption that students will make their ratings in a serious and conscientious manner, and that they make ratings on the basis of meaningful criteria. However, relatively little research has been done to check the extent to which these assumptions are met in reality. Despite the voluminous research on student ratings, we still do not know much about how students perceive teaching evaluations, how they approach the evaluation tasks required of them, or what criteria they actually use for evaluating their teachers. These questions are important for the appropriate interpretation and use of the student evaluation data for both formative and summative purposes. Theall & Franklin (1990) argue that the important question about student evaluations of university teaching concerns the interpretation and use of the data by teachers and administrators. Unless we have a better understanding of the nature and meaning of student ratings of university teachers from the students' perspectives, it is unlikely that we will be able to interpret and use the data in a sensible and valid manner.

The research questions

The present study attempted to bridge this gap in the literature on student evaluations of university teaching by probing into how university students go about making ratings for their teachers in the process of teaching evaluations. More specifically, the study aimed to examine the following four inter-related sets of questions:

1. What do students perceive as the main purpose of teacher evaluations in the university, and how can their perceptions be categorised? To what extent do they believe that they should have the opportunity to evaluate their teachers, and what are the justifications for their beliefs? To what extent do they believe that student evaluations will lead to improvements in teaching and learning? To

what extent are their various perceptions related to their overall reactions to the evaluation process?

2. What are the perceptions and attitudes of the students towards student evaluations of teachers? How serious are they in making ratings? Why are some students more serious than the others in the process? What factors may contribute to the differential attitudes?
3. How do students arrive at a particular rating for a particular lecturer? What strategies or approaches do they use to derive their ratings? To what extent are their ratings influenced by the views of their peers?
4. What determines the particular ratings that students choose to give for their teachers on a particular evaluation item? What criteria do they use to arrive at the ratings they give? What meanings do they attach to the various rating values that they give for their teachers?

Given the paucity of prior research in this area, the study was exploratory in nature. The emphasis was on describing, categorising, exploring and hypotheses-generating rather than confirming relationships or testing theories or models. It was conceived as the essential first step in researching the role of students in the process of teacher evaluation. It was hoped that the findings of the study would lead to a better understanding of the nature and meaning of student evaluations, and stimulate further research in this area.

CHAPTER 2 REVIEW OF THE LITERATURE

Student evaluation, most commonly in the form of ratings, is the predominant form of teaching evaluation at universities (Magner, 1994). In the United States, Seldin (1998) found that 88% of deans reported to have always used systematic student ratings in evaluating staff performance. In the UK, numerous universities have formalised student evaluations of teaching and made use of the results in appraising and making personnel decisions about their staff (Husbands, 1998). In Australia, student evaluations have been used as an indicator of teaching effectiveness in all higher education institutions (Ramsden, 1991). In Hong Kong, all universities, in anticipation of the Teaching and Learning Quality Process Review, have set policies for course or teaching evaluations in which student evaluations are the most prevalent tool for evaluating instructional effectiveness (Lo, Wong, Barrett, & Wong, 1999).

Student ratings are particularly attractive to university administrators because they are relatively easy to collect, simple to score, and inexpensive to analyse and report. Furthermore, they are supported by a large body of research on the validity and reliability of the forms, and provide a seemingly objective assessment of teaching that does not require justifications (Stone, 1995). They are also the most visible way to demonstrate to external bodies that a system is in place to monitor and assure the teaching quality of the institution. Thus, student evaluation has become "... a central part of the fabric of university life and the average score on a question asking whether an instructor is an 'effective teacher' is often taken as a barometer of teaching competence and an indicator over time of the instructor's expertise." (Kulich & Dean, 1999: 27)

2.1 The controversy over student evaluations of university teaching

Despite its widespread use, student evaluation of teaching is perhaps the most controversial technique used to measure teaching effectiveness. Boyer (1990: 37) remarks that teaching evaluation “remains a mare’s nest of controversy.” Similarly, Braskamp & Ory (1994: 5) suggest that the “credibility of faculty assessment remains one of the most precarious and sensitive issues on campus... Because evaluation of teaching effectiveness is often based solely on student ratings, it is seen as a mere popularity contest.”

Academics have contrasting views concerning the usefulness of student evaluation of teaching. Supporters argue that student evaluation is reasonably reliable and valid, relatively free from biases, and is useful for formative and summative evaluation purposes (Cashin, 1995; Marsh, 1987). They also maintain that student ratings, *if properly collected*, are the single, and often the only, valid source of data on teaching evaluations (Scriven, 1988; McKeachie, 1997a). All advocates, however, agree that for summative evaluation purposes, student ratings must be interpreted and used in a cautious and appropriate manner, and should only be *one* source of information for evaluating teaching.

Sceptics take the view that student evaluation has poor validity and reliability, and is detrimental to teaching quality in the long run. Trout (1997) argues that student evaluation reflects the views of the many students who are hostile to the necessary routine and rigours of higher education. Similarly, Stone (1995) alleges that student evaluation encourages teachers’ accommodations to students and leads to a decline in academic standard. Haskell (1997) argues that student evaluation of teachers is a serious infringement on academic freedom, because it provides a control mechanism over curricula, course content, grading, and teaching methodology. Sacks (1997: 29)

concur, and opines that in this age of hyper-consumerism, student evaluations “are a pernicious influence on academic quality.”

Arguments for the use of student evaluations

Several arguments have been put forward to make the case for using student evaluations. Some researchers maintain that students can provide useful feedback for both diagnostic and summative evaluations of teaching. Scriven (1995) suggests that by virtue of their unique position and qualifications, students are best placed to rate (a) their own increased knowledge and comprehension, (b) changed motivation towards the subject, careers, and further learning, (c) observable matters of fact relevant to competent teaching, (d) the regular teaching style indicators, and (e) coverage of facts. Ramsden & Dodds (1989) point out that although students are not experts in the subject matter, they have experienced a lot of teaching and only they, as non-experts in a subject, can accurately describe how effective a course is in helping non-experts to understand.

Furthermore, proponents claim that student evaluation is reasonably valid as a measure of teaching effectiveness, as indicated by the general agreement between teachers and students on components of good teaching (Feldman, 1988), and the positive and significant correlation of student ratings with student learning and other measures of teaching effectiveness (Marsh, 1987). Some researchers even contend that student evaluation is the single most valid source of information for evaluating teaching (Scriven, 1988; McKeachie, 1997a).

Other authors argue that by incorporating student evaluations into personnel decisions, it provides the impetus and incentive for teachers to engage in professional

development aimed at improving teaching (Murray, 1987), and increases the chances that good teaching is recognised and rewarded (Aleamoni, 1981).

Arguments against the use of student evaluations

There are four major arguments against the use of student evaluations, particularly for making summative judgements about teachers or courses. The first and foremost criticism is that student evaluations are inappropriate measures of teaching quality because students lack the maturity or expertise to make judgements about teaching. For example, Bauer (1997: 26) argues that anonymous student evaluation of their teachers' performance is reprehensive because "it presumes that the not-yet-competent are qualified to evaluate the already-competent." Similarly, Newport (1996: 19) suggests that there is "little or no reason to believe that university students have acquired the knowledge to rate their instructors on... [the] high inference items."

Student ratings have also been criticised for being biased in the sense that they may be affected by factors unrelated to the teaching of the instructor. Chandler (1978: 151) writes:

[We] should not maintain the illusion that student ratings of teachers are measuring teaching effectiveness, intellectual achievement or even basic understanding. Rather, these student ratings may be reflecting something about the raters – their psychological needs, feelings of satisfaction and attitudes, or they reflect personality characteristics, popularity, and speaking characteristics of the teachers...

His view is shared by Cahn (1987), who argues that students know if instructors are likeable or the lectures are enjoyable, but not if they are useful to their learning. Other researchers have also expressed concerns over the 'biasing' effects of other background variables on student ratings such as course and student characteristics (Husbands, 1998; Kwan, 1999), grading policies (Greenwald & Gillmore, 1997), and

the instructor's teaching style (Williams & Ceci, 1997). Husbands & Fosh (1993) point out that the possible existence of biases must necessarily entail some problems in the interpretation and use of the ratings for decision-making by universities:

Some invalidities one way or the other may not be too important when student evaluations of teaching are intended only for formative purposes... On the other hand, even a small amount of invalidity may be important when results are used by educational managers for complex decisions about promotion, tenure, salary progression or the appropriate recipients of rewards (often financial ones) for 'excellence in teaching' (p. 109-10).

Critics also believe that student evaluations are detrimental to academic quality and standards in the long run, for a number of reasons. Trout (1997) alleges that students want only easier courses, lighter workloads, and higher grades. He thus concludes:

It is hard to imagine a practice more harmful to higher education than one that encourages instructors to satisfy the demands and pleas of students who resent the appropriate rigors of college instruction. These [student evaluation] forms are not just invalid and unreliable; they are pernicious (p. 30).

Cholakian (1994: 26) argues that by asking students to evaluate their teachers, "we are implicitly teaching students to deflect responsibilities [for their own learning from themselves to the teacher]... Are we improving the profession or appeasing the Philistines?" Stake (1997) further submits that student evaluations undermine the trust and faith of the students in their teachers, and have strong adverse effects on teachers' behaviours.

Finally, there have been criticisms that many of the items or instruments for collecting student feedback are vague, ambiguous, subjectively stated, or irrelevant (Kolitch & Dean, 1999; Tagomori & Bishop, 1999). It follows that data collected by such instruments are neither credible nor valid for evaluating teaching.

It is clear from the discussion above that the controversy over student evaluations centres around three main issues: (a) To what extent are students able to give reliable and valid information on the quality of teaching they receive at university? (b) To what extent are their ratings 'biased' by other factors? (c) To what extent can student evaluations lead to teaching improvements?

In the following sections, the main findings of past studies pertinent to the three questions above are examined, and the limitations of existing research are discussed.

2.2 The reliability, validity, and utility of student evaluations

To date, more than 2000 books and journal articles have been published on student evaluations of teaching (McKeachie & Kaplan, 1996). Given the huge amount of literature on this topic, a comprehensive review is beyond the scope of this chapter. Hence, only a brief summary of the pertinent findings is reported below, drawing heavily upon the results of the meta-analyses and reviews previously conducted by other researchers wherever appropriate.

Reliability of student ratings

There has been relatively little dispute over the reliability of student ratings. In educational measurement literature, reliability refers to the consistency, stability, and generalisability of the ratings (Cashin, 1995). Regarding consistency, past studies have clearly demonstrated that *well-designed* student rating instruments can be highly reliable, with inter-rater reliabilities of about .70 for 10 raters, .80 for 20 raters, and .90 for 30 raters or more (Cashin, 1995; Marsh & Roche, 1997). Marsh & Roche (1997: 1188) thus conclude:

Given a sufficient number of students in any one class (or, perhaps, averaged across different classes), the reliability of class-average SETs compares favorably with that of the best objective tests.

Research has also shown that student ratings are relatively stable over time. In a longitudinal study, Overall & Marsh (1980) found an average correlation of .83 between end-of-course ratings with retrospective ratings by the same students years later. Marsh & Hocevar (1991) further showed that there were no systematic changes in ratings for a diverse group of 195 teachers over a 13-year period. In addition, Hativa (1996) found evidence of stability in both the levels of ratings and the shape of strength/weakness profiles over four sets of student evaluation data. Cross-sectional studies have also shown agreement between ratings by current students and alumni (Centra, 1993).

As regards the generalisability of the ratings, Gilmore, Kane, & Naccarato (1978) found that the influence of the instructor who teaches the course is much higher than that of the course that is being taught. Marsh (1981) compared the effect of the instructor and course on student ratings, and found that the correlation between overall ratings of different instructors teaching the same course was $-.05$, whereas correlations for the same instructor in different courses and in two different offerings of the same course were $.61$ and $.72$ respectively. These results support that student ratings reflect the general teaching effectiveness of the teacher, not just his/her performance in a particular course.

However, many universities use student rating forms that are locally developed, often in an *ad hoc* manner without any systematic attempt to establish their psychometric properties. Thus, the reliability of such instruments is questionable.

Validity of student ratings and the question of bias

The basic question concerning validity of student ratings is: To what extent do the ratings measure what they purport to measure? Abrami, d'Apollonia, & Cohen (1990) suggest that there are two views of validity of student ratings – their validity in accurately reflecting students' opinions about the quality of instruction (i.e., the satisfaction of the students with the instruction as consumers of the education process), and their validity as a measure of instructional effectiveness. While the first view of validity is seldom contested, the latter has been a subject of continuous heated debate (Marsh & Roche, 1997).

The fundamental problem in validating student evaluation of teaching is the lack of a single criterion of effective teaching (Goodwin & Stevens, 1993; Elton, 1984). Four major validation designs and research strategies have been used for studying the validity of ratings in assessing teaching effectiveness – the multisection validation design, the multitrait-multimethod (MTMM) design, correlational studies, and laboratory studies (Abrami, d'Apollonia, & Cohen, 1990).

(a) Multisection validity design

In the multisection validity design, the section average score on student ratings is correlated with the section average score on a common achievement test across multiple sections of the same course taught by different teachers. In the strongest multisection designs, a common textbook and a common syllabus is used for all sections. Furthermore, in many of the studies, section differences in student features were controlled either experimentally via random assignment, or statistically through ability pretests. Abrami, d'Apollonia, & Rosenfield (1996) argue that although the

design is not perfect, it is the strongest design for addressing the degree to which student ratings predict teacher-produced learning.

Over 40 multisection studies have been carried out to date. The literature has been extensively reviewed both qualitatively and quantitatively (e.g., Cohen, 1981, 1987; d'Apollonia & Abrami, 1996; Dowell & Neal, 1982; Feldman, 1989a; Marsh, 1987). The findings of the individual studies seem to vary widely – the reported validity coefficients range from -0.75 to $+0.95$. However, meta-analyses of the studies reveal a moderate positive correlation between ratings and student learning. For example, Cohen (1987) reported mean correlations of $.45$ and $.34$ between students' achievements and their global ratings on the instructor and course, and between achievement and students' ratings on specific dimensions respectively. In a more recent meta-analysis, d'Apollonia & Abrami (1996) found that the mean validity coefficient of general instructional skills across the 43 studies was $.47$ after correcting for attenuation, with a 95% confidence interval extending from $.43$ to $.51$. They conclude that student ratings are “moderately correlated with student learning in multisection college courses” and that, on average, “there exists a reasonable, but far from perfect, relationship between some student ratings and learning” (p. 238).

It should be noted, however, that most multisection studies have been carried out in large introductory classes. In many of the studies, the common examinations are in the form of standardised objectively-scored tests focusing on low-level cognitive outcomes (Feldman, 1998). Thus, the generalisability of the results to higher-level courses and to students' learning beyond what is measurable by standardised objective tests is questionable. They are also criticised by other researchers for having an unduly limited criterion of effective teaching, and being methodologically problematic (Marsh, 1987; McKeachie, 1997a).

(b) Multitrait-multimethod design

Marsh (1987) argues that there is no adequate single indicator of effective teaching and thus, advocates a construct-validation approach. Under this approach, student evaluations are posited to be positively related to a wide variety of other indicators of effective teaching (e.g., instructor self-ratings, alumni ratings, peer ratings, ratings of trained observers, etc.), and ratings on specific dimensions are posited to be most highly correlated with variables to which they are most logically and theoretically related.

Empirical evidence tends to support that student evaluations are related to other measures of teaching effectiveness. Feldman's (1989b) meta-analysis of 19 studies of the correlation between student ratings and instructor's self-ratings reported a mean correlation of .29 between the two. In larger studies involving ratings of more than 50 instructors, the correlations between student and instructor self-ratings were even higher, ranging from .45 to .62 (Marsh, 1987: 293). Student ratings have also been found to correlate with alumni ratings – the correlations range from .40 (Overall & March, 1980) to .75 (Braskamp & Ory, 1994). Feldman's (1989b) meta-analysis reported a mean correlation of .69, suggesting a strong relationship between student and alumni ratings. Positive correlations have also been found between student ratings and ratings by trained external observers (Feldman, 1989b; Marsh & Dunkin, 1992). With regard to the correlation between student and administrators' ratings, Kulik & McKeachie (1975) found mean correlations between .47 and .62 for specific rating dimensions. Feldman (1989b), using the global items, was only able to find a lower average correlation of .39. As for colleagues' ratings, Kulik & McKeachie (1975) reported mean correlations between .48 and .69, while Feldman (1989b) found a mean correlation of .55. The usefulness of administrators' and colleagues' ratings based on

classroom observations, however, has been questioned by Marsh & Dunkin (1992) because of low reliability. Reviewing the available evidence, Marsh & Roche (1997: 1190) conclude:

[Student evaluations are] significantly and consistently related to ratings by former students, students' achievement in multisection validity studies, teachers' self-evaluations, and extensive observations of trained observers on specific processes such as teachers' clarity. This pattern of results supports their construct validity.

However, it must be acknowledged that while student ratings have been found to correlate with other measures of teaching effectiveness, the estimates of validity of the ratings reported are not impressive and at times inconsistent. It is also debatable whether the alternative measures such as instructor self-ratings, former student ratings, etc., represent adequate criteria of effective instruction (Gaski, 1987). Furthermore, these ratings can only help establish the validity of ratings as measures of teaching process, but not as measures of the product of the teaching (Abrami, d'Apollonia & Rosenfield, 1996).

(c) Correlational and laboratory studies of student ratings

A common concern over the use of student evaluations is the belief that they are often 'biased' by factors unrelated to teaching effectiveness. Numerous correlational and laboratory studies have been carried out to investigate the effect of the various background variables such as course characteristics, teacher characteristics, student characteristics, and administrative procedures on student ratings. They have also been extensively reviewed both quantitatively and qualitatively in literature (for example, Cashin, 1988, 1995; Centra, 1993; Feldman, 1978, 1979, 1983, 1984, 1986, 1993; Kulik & McKeachie, 1975; Marsh, 1984, 1987; Marsh & Roche, 1997; Murray, 1991; Wachtel, 1998). The relationships between student ratings and various background variables as identified in past research are summarised in Tables 2 to 5 below.

Table 2: Relationship between student ratings and administrative procedures

Variables	General findings
Administrative procedures	
a. Timing of evaluation	No or little effect (lower ratings for evaluations carried out during final exam)
b. Anonymity of raters	Higher ratings when students need to identify themselves
c. Teacher's presence	Higher ratings when teacher being evaluated is present when evaluation is conducted
d. Stated purpose of evaluation	Higher ratings if stated purpose is for promotion and tenure

Table 3: Relationship between student ratings and course characteristics

Variables	General findings
Course characteristics	
a. Electivity	Higher ratings for elective or non-required courses
b. Time of class	No consistent effect (some found lower ratings for courses in very early mornings, late afternoons or immediately before after lunch)
c. Level of course	Higher ratings for higher-level courses
d. Class size	Higher ratings for smaller classes (some found a curvilinear U-shaped relationship)
e. Discipline area	Higher ratings for humanities & arts, lower ratings for social sciences, still lower for mathematics and sciences
f. Difficulty/workload	Higher ratings for more challenging difficult courses

Table 4: Relationship between student ratings and teacher characteristics

Variables	General findings
Teacher characteristics	
a. Rank	No or a weak positive relationship for regular teachers
b. Teaching experience	No or a weak negative relationship
c. Research productivity	No or a weak positive relationship
d. Personality	Few personality traits correlate with ratings
e. Enthusiasm and expressiveness	Higher ratings for more enthusiastic and/or expressive teaching ¹
f. Gender	No consistent significant relationship but a possible interactive effect between teacher and student gender
g. Physical appearance	No or a weak positive relationship, some found a significant interactive effect between gender, physical appearance and style

-
1. This is often known as the 'Dr. Fox' effect (Naftulin, Ware, & Donnelly, 1973). In the original study, a professional actor gave a lecture which was devoid of meaningful content to a group of educators and graduate students in an enthusiastic and expressive manner, and received favourable ratings from the audience. The results were interpreted to mean that enthusiastic and expressive teachers could 'seduce' favourable student evaluations. However, the study was severely criticised for its methodological flaws. The issue was re-examined by a number of later studies (Ware & Williams, 1975, 1977; Williams & Ware, 1976, 1977). Marsh & Ware (1982) re-analysed the data from the Ware & Williams studies and found that the 'Dr. Fox' effect was not supported in the condition most similar to classroom teaching, i.e., when students were given the incentive before viewing the lecture to learn the materials. Abrami, Leventhal, & Perry (1982) reviewed the Dr. Fox studies and concluded that while expressiveness did interact with the content manipulation and a host of variables examined in Dr. Fox studies, none of the interactions accounted for more than 5 percent of the variance in student ratings. Marsh (1987: 331-6) provides a critical review of the literature on the topic.
-

Table 5: Relationship between student ratings and student characteristics

Variables	General findings
Student characteristics	
a. Prior interest in subject	Higher ratings from students with a greater prior interest in the subject
b. Expected grade	Higher ratings from students expecting to get higher grades ²
c. Gender of students	Inconsistent findings (some found that female students tended to give higher ratings than male students)
d. Major or minor	Higher ratings from majors than from non-majors
e. Personality	No consistent or meaningful relationship

The methodological problems with correlational and laboratory studies of bias have been pointed out by Abrami, d'Apollonia, & Cohen (1990) and Marsh (1987). Yet, the results of the studies clearly suggest that student ratings do correlate with some of the background variables, albeit to varying degrees. Writers disagree, however, on whether this constitutes a bias in ratings. In fact, there is no clear consensus on the conceptualisation of bias in student ratings (Feldman, 1998). Marsh (1987) argues that the mere existence of a significant correlation between students' evaluations and some background characteristics should not be interpreted as a bias in student ratings. He argues that to establish the existence of bias, it must be demonstrated that the variable is not correlated with effective teaching. Other writers disagree, and suggest that biases

2. The literature clearly shows that there is a positive relationship between student ratings and expected grades. The interpretation of the results, however, is controversial. Three alternative explanations have been proposed: (a) the grading leniency hypothesis – that instructors giving more lenient grades receive higher ratings, (b) the validity hypothesis – that more effective instructors help student learn better and earn higher grades, thus supporting the validity of student ratings, (c) the student characteristics hypothesis – that students' entry characteristics such as prior subject interest and motivation contribute to the covariation between grades and ratings. Evidence suggests that all three effects may operate simultaneously (see Marsh, 1987: 317-321). Greenwald & Gillmore (1997) suggest that student ratings should be adjusted for grading leniency effect to make them more useful for evaluating teaching. Others researchers disagree (Feldman, 1998; Marsh & Roche, 1997) on the ground that it may eliminate valid effects of good teaching reflected in superior learning and higher grades and thus, introduce another type of unfairness in the evaluation.

in student ratings occur insofar as the ratings are affected by variables that are not directly under the control of the lecturer. For example, Cashin (1988) writes:

When using ratings for personnel decisions or the instructors' improvement, I would suggest an even narrower definition, restricting bias to variables *not* a function of the *instructor's* teaching effectiveness... Student motivation and class size may impact teaching effectiveness, but instructors should not be faulted if they are less effective teaching large classes of unmotivated students than their colleagues are with small classes of motivated students. In this case, student motivation and class size, although they are related to teaching effectiveness, are not a function of the instructor's characteristics, but of student and course characteristics, and so they should be considered sources of bias (p. 3, emphasis in original).

Feldman (1998) stresses that a distinction should be made between bias and unfairness in ratings. He points out that some variables such as class size do have an impact on teaching effectiveness although it may be unfair to compare teachers in classes of widely different sizes. In this case, the unfairness lies in the differences in teaching conditions rather than biases in ratings. But the problem is: even if we accept such a distinction, the fact remains that student ratings are influenced by background variables not related to the performance of the lecturer. No matter how such effects are to be labelled, the validity of ratings as a measure of the teaching effectiveness *of the teacher* must be subject to question.

Writers also disagree with respect to the significance of the effects of the background variables on ratings. Summarising the large number of studies on the effects of background variables on student ratings, Marsh (1987) concludes that the effects are small, as the variables were only "able to explain a small portion of the variance in student ratings" (p. 310). Kwan (1999) argues, however, that in actual practice judgements are being made mainly on the basis of the absolute ratings rather than the explanatory power of the model. Insofar as the background variables lead to large differences in the absolute ratings, they should be considered as having a large

practical effect. He found that some of the background variables such as disciplinary differences and class sizes did have a large effect on the absolute ratings, with effect sizes ranging from 0.43 to 1.01. Husbands & Fosh (1993) further point out that the significance of the effects cannot be separated from how the ratings are to be used. Even a small amount of invalidities may be important when the ratings are used for complex decisions about promotion, salary progression, or recipients of rewards. In other words, what is a 'small' effect on student ratings to researchers may be a 'substantial' one to the teachers being evaluated, especially when their performance is to be judged by the ratings they receive.

Effect of student evaluation feedback on teaching improvements

A number of studies have examined the effect of student evaluation feedback on teaching. They have been systematically reviewed by Brinko (1993), Cohen (1980), and L'Hommedieu, Menges, & Brinko (1990). Results show that there is a modest but significant positive effect of feedback on improvement of teaching. However, most of the studies were short-term studies and measured the effect in terms of changes in ratings rather than student learning. Longer-term studies have been rare, but studies in which feedback is coupled with consultation have shown stronger and longer-term effects (e.g., Marsh & Roche, 1997; Piccinin, Christi, & McCoy, 1999).

It is also clear that the effect of feedback on performance is mediated by a number of variables (Coe, 1998). Brinko (1990) suggests that to encourage teaching improvements, feedback should contain accurate data and irrefutable evidence, as the credibility of the feedback to the recipients will influence how they are affected by it. In this connection, the credibility of student ratings as perceived by the teachers will have a paramount effect on how they react to them. If teachers believe that student

ratings are basically erratic or flawed in nature, it is highly unlikely that they will treat them seriously, or make use of them for improving their teaching.

Problems with prior validity studies of student evaluations

Any use of student evaluations must be based on the assumption that students are willing and able to provide valid judgements about the teaching they have received.

Dunkin (1986: 769) comments that:

Presumably, student evaluations are based, at least partially, on their perceptions of the teaching they receive and, presumably, those perceptions are accurate in that they reflect the actual processes engaged in by the teachers. If these presumptions are not justified, then it is difficult to see how student evaluations can be legitimately useful except in terms of intrinsic benefits for those making them, such as in venting their spleens or expressing their pleasure.

Ryan, Anderson, & Birchler (1980) argue that the diversity in staff opinions about the use of student evaluations for personnel decisions may be based more on a difference in belief about student qualifications to provide such information than from their own ratings or observations concerning possible abuses. They suggest that for faculty to accept the use of SETs for personnel decisions, it must be demonstrated that students can be a valid source of such information. Yet, Landy & Farr (1980) argue that the traditional psychometric attempts have had limited success in addressing the validity issues of ratings, and Cooper (1981) contends that studies of how the ratings are actually formed in the students' mind may be more revealing. Researchers also point out that psychometrically sound instruments alone are unlikely to provide valid evaluations (Kishor, 1995; Williams & Ceci, 1997). Thus, Kishor (1995: 193) suggests

[The] validity of student ratings may be improved further by focusing on actually how the students form their ratings... More effort needs to be devoted to students' rating strategies rather than on instrument development... understanding raters' cognition may at least be as important as improving instrumentation. If one can identify how judgements are caused, those judgements could be improved for their reliability and validity.

In other words, it is of vital importance for us to know more about the role of the students in the process of evaluating their teachers or courses. For example, we need to understand more about how students make sense of the evaluations, what influences their attitudes towards evaluations, how they react to the rating process, and what goes on in their minds while making their ratings, etc.

Given the important part played by students in the evaluation process, it is surprising to note that relatively few studies have been carried out to investigate the role of students in the process of rating their teachers. In fact, most research to date has been driven by the quantitative research paradigm (Benz & Blatt, 1996), focusing primarily on the psychometric properties of the instruments. Comparatively little research attention has been given to the perspectives of the different groups involved in the evaluation process (Schmelkin, Spencer, & Gellman, 1997), particularly the students who give the ratings.

Crittenden & Norr (1973) point out that we know too little about the process by which students arrive at their ratings of courses and teachers. Yet, little progress has been made in our understanding of the issue over the years. Feldman (1998) makes the same observation in a recent review of the literature: “We do not, in fact, know very much about what does go on in students’ minds when they fill out rating forms...” (p. 51). He restates that “[we] need to know more than we do about how students arrive at their opinions about any particular teacher...” (p. 62). This clearly shows that there is a significant gap in the literature on student evaluations of teaching where more research is needed, particularly from the qualitative paradigm (McKeachie, 1997a). The present study is an attempt to fill this gap by looking into the attitudes and mental processes of the students in evaluating their teachers.

Another problem with the existing literature on student evaluations is that most research to date is based on the experience of North American universities and, to a lesser extent, Australian, and UK universities. The generalisability of the findings to other countries with different cultural backgrounds is still uncertain. In a study of the applicability of instructional evaluation instruments widely used in Western universities to the Asian settings, factor analysis failed to replicate the same underlying constructs expected from these instruments (Lin, Watkins, & Meng, 1995). Thus, it seems reasonable to question if the findings from the West are representative of all contexts and conceptions of teaching, particularly those in different cultures. As evaluation policies, procedures, and instruments are inevitably rooted in specific cultural values and norms (Pratt, Kelly, & Wong, 1999), there is a definite need to extend the existing body of research by collecting evidences from different cultural settings. This study will contribute to the growing literature on student evaluation of teaching by examining how students in a university in Hong Kong go about rating their teachers.

The role of the students in the evaluation process

Research into the role of students in the process of evaluating their teachers or courses is scanty. Prior studies tended to focus on three main themes: students' attitudes towards the evaluations, students' implicit theories of teaching, and the meanings underlying students' evaluations of their teachers and courses.

(a) Students' reactions to teaching evaluations

A few studies have examined the reactions of students to teaching evaluations. Taylor & Ricketts (1982) found that the majority of students in their study reported that they answered the evaluation questions as honestly and fairly as possible, and favoured the continuation of student evaluations. However, the students surveyed showed some

degree of concern about the extent to which their teachers would actually make use of the feedback to improve the courses and instructions. Furthermore, the more senior students were found to have more pessimistic views about the value of the evaluations than the more junior ones. According to Jacobs (1987), the majority of students believed that they should have the opportunity to rate their instructors, and that they were competent to do so. Most students also stated that they gave serious considerations to the ratings they made. Marlin (1987) also found that the majority of the students reported to have treated the evaluations seriously, and tried their best to be fair and accurate in their ratings. Students in the survey also complained that their opinions were often neglected by the teachers and administrators, and that most teachers did not alter their teaching on the basis of their feedback. Similar results have been obtained in other studies (e.g., Smith & Carney, 1990; Spencer & Schmelkin, 1995; Wulff, *et al.*, 1985).

The affective reactions of the students in the process of evaluating their teachers was examined by Ingrim (1979), who found that some students had anxieties in evaluating their teachers, as a result of their uncertainty about their own ability to evaluate their teachers fairly, their fear of retribution from their teachers in the case of negative evaluations, their powerlessness over the use of their opinions by their teachers, and dread from saturation. In those cases, students might either ritualistically endorse all teachers or assign all to a middle-ranged average score.

As regards students' attitudes towards different evaluation methods, Abbott *et al.* (1990) found that students were more satisfied with interview methods at midterm followed by extended instructor reactions than with the conventional end-of-course evaluations using standardised rating forms. Wulff *et al.* (1985) also found that students predominantly recommended methods in which they played a role. They favoured

methods that focused on both oral and written procedures in generating the data, and preferred mid-term or a combination of mid-term and end-of-course evaluations.

Most of the studies so far have relied mainly on quantitative questionnaire surveys and thus, reflected the researchers' perspectives rather than those of the students. They describe students' attitudes and reactions without probing more deeply into why and how students develop such attitudes. Furthermore, as most of them were conducted in the United States, the generalisability of the findings to places with different social and cultural settings is unclear. Thus, more studies are needed to understand how students react to teaching evaluations, especially those from the students' perspectives (Wachtel, 1998).

(b) Students' implicit theories of teaching

Research has shown that students hold implicit theories of teaching which influence both the rating process and the dimensions of teaching which are identified from factor analyses of rating data (Whitely & Doyle, 1976; Cadwell and Jenkins, 1985; Lodzinski, 1991). The findings suggest that the dimensions used in various rating scales developed from factor analytic studies may reflect dimensions that are inherent in the beliefs of the student raters, rather than dimensions that characterise instruction and the behaviour of instructors (Harrison, Ryan & Moore, 1996). Thus, researchers argue that student ratings are a reflection of the rater's schema rather than a measure of the teachers' performance (Tang & Tang, 1987). In a more recent study, Kishor (1995) found that student raters tended to infer traits and behaviours and provided ratings for corresponding items even when the instructor behaviour was limited to a subject of performance data. He concludes that students' implicit theory of instruction is a source of invalidity in student ratings. Cadwell & Jenkins (1985) illustrate this with an example:

... if students feel good about their instructor, they might rate the instructor as accessible outside of class, even if they never attempted to contact the instructor outside of the classroom, or they might respond on the rating form that the instructor provided different points of view without any attempt to recall specific instances of this type of behavior (pp. 391-392).

Other researchers, however, argue that students' reliance on general impressions is consistent with the information-processing model of student ratings (d'Apollonia & Abrami, 1997), which may increase rather than distort the accuracy of the overall ratings (Becker & Cardy, 1986; Cooper, 1981). Furthermore, Harrison, Ryan, & Moore (1996) found that students had self-insights and a reasonably high level of consensus in making their overall evaluations, indicating that students' overall ratings are not based on spurious or purely situational factors. They conclude that students' overall ratings are credible, are based on identifiable factors, and are the result of their previous experiences with the instructors rather than groundless conjectures.

(c) Meanings underlying student evaluations of teaching

Adelman (1997) points out that the validity question with student feedback is whether the inquirer or data-collection method uses constructs which the students understand and to which they can provide experiential answers. However, very few studies examine how students actually make sense of and give ratings on the evaluation items.

In an informal study, Talley & Timmer (1992) found that while faculty assumed without question that students would read and understand the instructions, students admitted not reading the instructions, and did not understand the comparative intent of the Likert categories and substituted them with their own interpretations. Moreover, many students commented that they did not understand the items clearly, or could not see the differences between items. They also observed that students often interpreted the items in ways different from those of the faculty. They noted that many students

regarded their input on the evaluations as a potent and justifiable punishment for 'bad' teachers and were quite upset if a concrete administrative response to their anger at a particular faculty member was not forthcoming at once. They conclude that the data they obtain

... exemplify the distinctly different social worlds of faculty members and students. The undergraduates' lived world is one of pragmatic realities, in which many abstract conceptual categories are meaningless... Students know whether or not they like a teacher and whether or not they are satisfied with a course because likability and satisfaction can be experienced concretely in an instructor's teaching style and personality (p.78).

Benz & Blatt (1996) examine the meanings that students attach to various evaluation items by gathering open-ended comments from students about their reasoning in giving their ratings on an evaluation form. They found that (a) students used a variety of evidence in making their ratings; (b) some students attributed their ratings to evidence of their own behaviour; (c) students held some interesting, perhaps naïve and odd, understandings about teaching; and (d) students often felt ambiguous about items that were presented as if they were concrete and quantifiable.

While both studies provide useful information on how students make sense of the evaluation items, neither examines the strategies that students actually use in making their ratings, nor do they relate students' personal interpretations and meanings of the items to the ratings they give. Thus, it is still unclear as to how students actually arrive at their ratings. Feldman (1998) suggests that a useful approach would be to interview students immediately after they have filled out rating forms, making use of the verbal report methodology.

This study contributes significantly to the existing body of research in student ratings by examining the strategies and criteria that students use to decide on their ratings of their teachers. It employs a qualitative approach that probes into the thinking

process of the students and enables the students' perspectives to emerge. The research methodology of the study is explained in Chapter 3.

CHAPTER 3 METHODS

In this chapter, the contextual background and the methodology of the study are explained. In the first section, the University and the student evaluation system are described to provide the reader with an understanding of the context under which the study has been carried out. The second section describes the general approach of the study, the methods of data collection, and the procedures adopted in analysing and interpreting the data. The last section discusses the limitations and delimitation of the study.

3.1 The University and the student feedback system

The study was carried out in a university in Hong Kong with a student population of over 20,000 (14,000 Full-Time Equivalents) and a full-time teaching staff of about 1,000. There are six faculties in the University, in the areas of applied technology, business and information systems, communication, construction and land use, engineering, and health & social sciences. An overwhelming majority of the students are Chinese, speaking the Cantonese dialect. However, the official language of instruction of the University is English, except for a few subjects or courses where the use of other languages are deemed necessary. As for the teaching staff, a considerable proportion of them are expatriates and cannot speak Cantonese, nor read and write Chinese. Most of the programmes and courses are oriented towards specific professions, often leading to professional qualifications.

The student feedback system

The student feedback system of the University has the following features:

1. All teaching staff members were to collect feedback from at least two classes of students they taught in an academic year, using the University-wide standardised Student Feedback Questionnaire (SFQ).
2. Staff members were allowed to choose the two classes of students to be surveyed, in consultation with the respective heads of department.
3. All Student Feedback Questionnaire (SFQ) forms were to be administered near the end of the teaching of the staff member, usually towards the end of the semester.
4. The SFQ forms were to be distributed and collected by departmental support staff, in the absence of the lecturers concerned.
5. The SFQ results would be used for both developmental (formative) and judgmental (summative) purposes, and would contribute as one source of evidence of the effectiveness of teaching of the staff member in the staff appraisal exercise.

However, in the actual implementation, there were variations in practice across departments, particularly in terms of the frequency of the evaluations. Some departments requested their staff to collect feedback on all of the classes they taught in an academic year. Coupled with the fact that some of the subjects were team-taught and that some subject lecturers might decide to collect feedback from students on the subject itself, students in some departments might have to complete more than one feedback form for a single subject for a semester.

The student feedback system was co-ordinated by an academic support unit, which was also responsible for analysing the data, generating the reports, and advising staff on interpreting the results. The author has been responsible for co-ordinating the University's student feedback exercise since its first inception in 1995.

The Student Feedback Questionnaire (SFQ)

The Student Feedback Questionnaire (SFQ) was the University-wide standardised questionnaire designed to collect feedback from students on the teaching of the individual lecturers. It had twelve closed-response type questions focusing on six aspects of the staff member's teaching, namely, Learning Outcome, Interaction, Individual Help, Presentation & Organisation, Motivation, and Feedback, with two items for each dimension. It also had two open-ended questions which asked students to comment on the aspects of the lecturer's teaching that were most useful to their learning, and to suggest how the teaching of the staff member could be changed to help them learn better (see Appendix A). Departments and individual lecturers could, if they wished, include up to 20 extra closed-response type items on the form.

Again, there were differences in practice across departments in the use of extra questions. Some departments developed additional departmental core items for evaluating their staff and requested every lecturer to collect feedback on those items as well. Most departments did not, and only used the twelve University-wide standardised items. Similarly, some lecturers chose to include their own extra questions on the SFQ form, but the majority did not.

3.2 Methodology

Given the scanty amount of prior research in this area, the present study was exploratory in nature and adopted a naturalistic approach. Data were collected mainly through in-depth individual interviews with students.

The naturalistic inquiry

Naturalistic inquiry is a model for conducting inquiries which is characterised by the following three features: (a) it is carried out in the natural setting, (b) it normally utilises a case study format, and (c) it relies on qualitative rather than quantitative methods (Guba & Lincoln, 1984: 311). It is based on the naturalistic paradigm, which is often conceived as an alternative to the rationalistic (also known as ‘positivist’ or ‘scientific’) paradigm which, until quite recently, has been the dominant methodology in educational research³ (Rist, 1977: 42).

Owens (1982) points out that there are basic differences between the naturalistic and rationalistic paradigms in terms of their assumptions and beliefs about the nature of the real world, the relationship between the inquirer and the participant under study, the generalisability of observations, the methods for collecting and analysing data, the role of theory in the inquiry, and the feasibility of a preordinate design of the study. Guba & Lincoln (1984) also point out that there are axiomatic differences between the two paradigms. Their views are summarised in Figure 1.

3. Some writers argue that there is a third paradigm, known as the ‘critical’ or ‘emancipative’ approach (see for example, Candy, 1989; Carr & Kemmis, 1983).

Figure 1: Axiomatic differences between the Rationalistic and Naturalistic Paradigms⁶

Axioms about	Rationalistic Paradigm	Naturalistic Paradigm
Reality	Single, tangible, convergent, fragmentable	Multiple, intangible, divergent, holistic
Inquirer / respondent relationship	Independent	Inter-related
Nature of truth statement	Context-free generalizations – nomothetic statements – focus on similarities	Context-bound working hypotheses – idiographic statements – focus on differences
Attribution / explanation of action	Real causes; temporally precedent or simultaneous; manipulable; probabilistic	Attributional shapers; interactive (feed-forward and feed-back); non-manipulable; plausible
Relation to values to inquiry	Value-free	Value-bound

Source: Guba & Lincoln (1984: 316)

There have been heated debates over the relative merits of the two different research paradigms. However, it has been increasingly recognised that the distinction is overly divisive — there are commonalities as well as differences between paradigms such that distinct paradigms do not exist⁴ (McCutcheon, 1981). Researchers argue that “no one methodology can answer all questions and provide insights on all issues” (Rist, 1977: 42) and therefore, tend to think of the two approaches as “complementary to each other in the search for knowledge which involves both an explanation in terms of causes and an interpretation or understanding in terms of motives and intentions” (Keeves, 1988: 4). Thus, the choice of paradigm for a disciplined inquiry should

4. There are three major views of the relationships between paradigms: (a) the oppositional diversity thesis which sees the paradigms as oppositional and competitive to each other; (b) the complementary diversity thesis which views them as having differential merits and limitations and thus, complementary to each other; and (c) the unity thesis which argues that the different paradigms are not epistemologically diverse – they share at least a common language. They are competing approaches or theories of educational research methodology, but not competing paradigms (For a summary, see Walker & Evers, 1999: 40-56). Furthermore, while many researchers impute their choices of methods to differences in philosophies or theories, it is undeniable that practical considerations such as resource constraints or ease of the method used may account for the widespread use of some methods over the others.

ultimately be “based on its ‘goodness of fit’ or appropriateness to the subject of the inquiry” (Candy, 1989: 10).

The choice of the naturalistic approach for the present study is justified by its appropriateness to the objective of the study, which is to investigate what is in the minds of the students when they are rating their teachers. Guba & Lincoln (1984: 318) argue that naturalistic inquiry is particularly apt for investigating the meanings and interpretations people ascribe to human behaviours because these constructions exist only in the minds of people and thus, are substantially inaccessible and must be dealt with in a holistic manner. Their view is supported by Hitchcock & Hughes (1995: 26), who argue that what goes on in the classrooms “is made up of complex layers of meanings, interpretations, values and attitudes.” They believe that a qualitative or naturalistic inquiry is needed to understand the actions, ideas, values and meaning “*through the eyes of the participants rather than quantification through the eyes of an outside observer*” (Hitchcock & Hughes, 1995: 26, emphasis in original). To summarise, the naturalistic approach has the clear advantages that it allows the respondents’ personal meanings and constructs to emerge, enables the researcher to collect rich information about the social phenomenon under study, and offers a contextual relevance and richness that is unmatched by other approaches.

The interview as a research technique

The interview is one of the most common methods for collecting data for qualitative or naturalistic inquiries. Cannel & Kahn (1968) [cited in Cohen & Manion, 1994: 271] define the interview as “a two-person conversation initiated by the interviewer for the specific purpose of obtaining research-relevant information, and focused by him [sic] on content specified by research objectives of systematic

description, prediction, or explanation.” The key feature of this method is the direct verbal interaction between the interviewer and the interviewee.

There are merits as well as limitations of using the interview for data collection. Rudduck & Hopkins (1985: 233) point out that a major advantage of the interview is “its usefulness in collecting personal information, attitudes, perceptions, or beliefs by probing for additional information.” Borg (1987: 110) also suggests that the interview has the advantage of being adaptable, which enables the interviewer to “follow up leads that show up during the interview, and thus obtain more data and greater clarity.” Gay (1976: 134) contends that the interview will “result in more accurate and honest responses since the interviewer can explain and clarify both the purpose of the research and individual questions.”

There are, however, certain drawbacks of the interview. First, there is a risk that the respondents may conceal their true feelings or beliefs in the interview. Miller & Cannell (1988: 457) have noted:

Reporting embarrassing events or properties presents considerable difficulty for many respondents, and not being able to report normatively valued characteristics can also cause some psychological discomfort. Therefore, respondents are often apt to censor response intentions which do not meet their perceived standards of social propriety.

Other limitations of the interview include the issues of subjectivity and bias in the data collection process, which are often referred to as the “Hawthorne” and the “Halo” effects. The first arises when the inclusion of the participants in the study distorts their feeling and behaviour. The second refers to the tendency for the participants to react positively to persons they like. Consequently, subject expectancy (i.e., the participants’ desire to help the researcher achieve his/her goals) and researcher

expectancy (i.e., that the expectations of the researcher alter the results of the study) may occur, and influence the trustworthiness of the results.

The aim of the present study is to investigate university students' perceptions, beliefs, attitudes, and rating decisions in the process of evaluating their teachers. To achieve this, it is important that the method of data collection is flexible and adaptive enough for the respondents' constructs to emerge, and for the researcher to probe for clarifications in the cases of inconsistent or vague replies. The interview is thus chosen as the method of data collection for the study.

The risk that participants might not reveal their true feelings and beliefs in the interviews was lessened because the interviewers were from an academic support unit of the university, not regular teaching staff of the department from which the students were drawn. It therefore reduced the needs for the students to 'please' the interviewers. Furthermore, by clearly explaining the purpose of the study to the participants before the interviews started and emphasising the importance of their feedback to improving the student feedback system of the University, it encouraged participants to give their honest opinions for the various questions asked. Finally, the potential biases due to researcher expectancy were reduced by employing two researchers to conduct the interviews and analyse the interview data in a collaborative but independent manner.

Hitchcock & Hughes (1995) suggest two ways to check the validity of interviews – by triangulation of data from various sources, and by re-interviewing. Both options had been carefully considered but were finally rejected on practical grounds, due to time and resource constraints. However, attempts had been made to collect quantitative ratings that participants made for the two anonymous teachers in a simulated rating exercise, and to triangulate the findings with the interview data. In addition, the trustworthiness of the data was improved by examining the consistency of

the responses of the individual students in different parts of their interviews rather than focusing on the response to the specific questions asked. Finally, the use of two researchers in the interviewing and interpretation processes has made possible investigator triangulation for enhancing the validity of the study (Smith, 1975).

Participants

Six departments broadly representing the major discipline areas of the University were identified for inclusion in the study. The departments selected were: Building & Real Estate (BRE), Business (BUSS), Design (SD), English (ENGL), Textiles & Clothing (ITC), and Rehabilitation Sciences (RS)⁵. The departments represented subject areas varying from the relative 'soft' disciplines (e.g., Design, English) to the relatively 'hard' ones (Rehabilitation Sciences) according to Biglan's classifications (1973). All of them, however, were more oriented to the 'applied' rather than the 'pure' side because of the nature of the programmes offered in the University.

Four students from each year group of a full-time undergraduate programme in the six departments (total = 72) were selected at random for participating in the interviews. A letter was sent to the chairpersons of the Learning & Teaching Committees of the respective departments asking for their assistance in selecting and inviting their students to take part in the interviews, according to the criteria specified by the project team (see Appendix B).

5. The six departments were selected from five faculties of the University. Two departments were chosen from the Faculty of Communication, because it was believed that design as a discipline was quite different from languages. None of the departments from the Faculty of Engineering were included, because in the year when the study was conducted, that Faculty piloted its own student feedback system which was distinctly different from the University-wide system. The decision was made in order to bypass the problems of interpretation that might arise because of the vast differences in the systems, policies, and instruments.

When the lists of students with their contact telephone numbers were received from the departments, a member of the clerical staff of the academic support unit would make contact with the students to set up the time and place for the interviews. The interviews were conducted in April and May, 1999. One student failed to turn up despite the various attempts to re-schedule the interview. As a result, only 71 students were finally included in the study.

The interview procedures

The interviews were conducted individually, in the offices of the respective interviewers located in the academic support unit. Thirty-nine interviews were conducted by the author and the other 32 interviews were conducted by a temporary Research Fellow. Each interview lasted for about 60 minutes, and all except one were conducted in Cantonese (the other one was conducted in English). The interviews were audio-recorded and subsequently translated and transcribed before data analysis was carried out.

The interview protocol

The interviews for the study were semi-structured. Borg & Gall (1979: 312) argue that semi-structured interviews have the advantage of “being reasonably objective while still permitting a more thorough understanding of the respondents’ opinions and the reasons behind them than would be possible using the mailed questionnaire... It provides a desirable combination of objectivity and depth and often permits gathering valuable data that could not be successfully obtained by any other approach.”

To ensure that the key themes would be covered in each of the interviews, an interview protocol was developed by the author in consultation with the other members of the project team to guide the focus of the interviews (see Appendix C).

Before the interviews started, the purpose of the interview was explained to the participants. Their consent to being interviewed and recorded was also sought. The interviews focused on three broad areas. The first focused on the implementation of the teaching evaluation system in their department, and how they and their classmates behaved and reacted to the system. The second part inquired into how they arrived at the ratings for their teachers for the individual items on the rating form. The last part asked students about their various perceptions of the student feedback system, and their beliefs concerning the usefulness of the system for improving teaching.

The verbal report method

Kelly (1955) proposes a ‘personal construct theory’ which suggests that events are only meaningful in relation to how they are construed by the participants. Personal constructs are “the dimensions that we use to conceptualize aspects of our day-to-day world” (Cohen & Manion, 1994: 299). Kelly argues that people create and use their own personal constructs to make sense of their environment and forecast events, which ultimately influence how they behave. He develops a ‘repertory grid’ to elicit constructs and identify their relationships with the phenomena under study (Kelly, 1969). The technique basically involves asking participants to evaluate phenomena (known as ‘elements’) in terms of the constructs they employ.

A similar strategy was adopted for this study. To investigate students’ thinking processes in making ratings for their teachers, students were asked first to think of a ‘good’ teacher and a ‘poor’ teacher they had actually encountered in their study at the

University, and then evaluate them by filling out a rating form consisting of six items drawn from the University-wide standard form (Appendix D). The teachers being rated were anonymous to the interviewers. To maintain consistency with the original SFQ form, ratings for each of the items were given on a five-point scale, ranging from '5' (strongly agree) to '1' (strongly disagree). The students were then asked, immediately after the rating simulation, to report verbally how they came to those decisions, and the criteria and evidence that they had employed in making the various ratings.

This method of data collection made use of the retrospective verbal recall method commonly used in psychological research on thought processes. Essentially, the verbal report method (also known as 'protocol analysis') involves asking participants to give verbal reports of their actual thinking processes while performing the task of interest (Payne, 1994). Ericsson & Simon (1993) argue that asking people to verbalise their thoughts as they enter and are attended to in their short-term memory is a useful method for obtaining information about people's thinking. They further contend that participants can report their thought processes retrospectively, provided that the tasks are of relatively short durations and the verbal reports are gathered immediately after the tasks are completed. Their claims have been supported by considerable empirical evidence (Crutcher, 1994). Some researchers have contested the validity of verbal reports in reflecting thinking accurately, and had serious concerns over the reactive effects of verbal reports on respondents' behaviours (see, for example, Nisbett & Wilson, 1977; Wilson, 1994). However, even to the critics, verbal reports are accepted as an excellent source of inspiration for studies aiming more at hypothesis generation than testing causal hypotheses (Wilson, 1994: 251).

Analysis of the interview data

The interview data were analysed in a stepwise replication manner (Guba & Lincoln, 1984). First, the author and the Research Fellow each read half of the interview transcripts to search broadly for significant themes and categories. The initial themes and categories were then compared and refined through discussion. The author then re-read all of the transcripts to categorise the responses in the interviews. The allocated categories were recorded, and quotations from the transcripts were identified to support the categorisation. Where appropriate, categories assigned to students on different themes were cross-tabulated to identify possible relationships between constructs. Meetings were then held in which the themes and categories identified by the author were presented to the Research Fellow who also read all the transcripts, to verify the credibility of the interpretations. Peer debriefing sessions (Pidgeon, 1996) were then organised to present the preliminary results in two open forums for teaching staff in the University. Finally, a confirmability audit (Guba & Lincoln, 1984.: 329) was conducted by the Research Fellow to verify that each finding could be traced back to the original data and that interpretations of the data were reasonable and meaningful. Any discrepancies identified in the process were resolved by discussions, with reference to the original interview records.

3.3 Limitations and delimitation

By choosing to follow a naturalistic qualitative approach, the study has suffered from a reduction in the representativeness and precision of the results obtained as the sample was smaller than what could be managed with a quantitative study. But it was believed that the losses were more than compensated by the ability of the study to make use of the intelligent human-as-instrument (Guba & Lincoln, 1984), and to enable the

perspectives of the students to emerge from the data rather than imposing on them the views and values of the researchers. Given the lack of prior research in this area, the method was deemed appropriate.

Another problem inherent in a naturalistic inquiry is the subjectivity and potential researcher biases in the analysis and interpretations of the data. While the study could not claim that the problem was completely eliminated, the trustworthiness of the study had been enhanced by the stepwise replication and the confirmability audit adopted in the data analysis process.

Given the small number of respondents interviewed, the nonprobabilistic sampling, and that all of the students came from a single university in Hong Kong, the study did not make any claim of generalisability of the findings to other contexts. However, by purposive selection of students from the widest possible variety of disciplinary backgrounds available and from different year groups, the study had maximised the possible range of information that was collected, and increased the dependability of the study. Furthermore, with the 'thick descriptions' provided in the study, it is hoped that the readers could make their own judgements on the applicability of the findings to their own situations.

Another limitation of the study was its sole reliance on students' own descriptions and reports of their attitudes and behaviours in evaluating their teachers. It has been recognised that what people say may not correspond to what they really think or the ways they behave (Dean & Whyte, 1958), especially when the respondents have a need to please the interviewer. However, as LeCompte, Preissle, & Tesch (1993: 162) point out, "self reports are useful for assessing how individuals make judgements about people and events, and they do register what people think they do or what they think is socially acceptable to do." In the study, the interviewers were not part of the teaching

staff of the departments where the students came from. To some extent, this reduced the need for the students to please the interviewers and encouraged them to state their honest opinions. It seemed that respondents were not afraid to express their views frankly, as evident in the critical and uninhibited remarks students made about the University and their teachers in the interviews. Yet, the lack of other sources of information made it impossible to triangulate the data collected to verify its credibility, although some attempts have been made to at least triangulate the responses of the students in various parts of the interviews.

The present study used a method of stimulated recall to facilitate students to recount their thinking processes in making ratings for their teachers, and the criteria they employed to arrive at the exact ratings to give on each item. It should be noted that the rating simulation was not truly 'natural' because students were not making their ratings as part of the normal teacher evaluation exercise. Furthermore, requesting students to think of a 'good' and a 'poor' lecturer and rate them before reporting verbally how they arrived at the ratings for each of them on each item might influence the impressions and reasoning of the students in their responses. However, the present design had three advantages. First, it avoided the sensitiveness of asking the students about the ratings he or she gave for a particular teacher and thus, made it easier for students to give their honest opinions since anonymity was guaranteed. Second, it overcame some of the practical problems in setting up the interviews because students might have other obligations after the evaluation and were therefore not available for interviews immediately afterwards. It also helped to spread out the interviews over a longer period of time to enable the researchers to complete all interviews before the semester ended. Third, by asking students to rate 'good' and 'poor' teachers, we were able to maximise the range of information to be collected and contrast students'

descriptions of the extreme cases to ensure more credible interpretations of their views. It should be noted that as students were asked to make their ratings first and then recount how they made their decisions, their descriptions may represent their after-the-matter justifications for the ratings rather than their true mental reasoning while making the ratings. The alternative was to employ the method of think-aloud protocol, which is even more artificial in terms of their normal practice in making the ratings. The existing arrangement was not a perfect solution but it was believed that it balanced the needs for a good design with the various constraints faced by the researcher. However, all these limitations should be taken into account in interpreting the findings of the study.

CHAPTER 4 STUDENTS' PERCEPTIONS OF TEACHER EVALUATION

This chapter explores the various perceptions by the students of the teacher evaluation system. The first section analyses the students' perceptions of the purposes of teacher evaluations in the University. The second section examines their views concerning whether or not they should have the opportunity to evaluate their teachers, and why. The third section discusses their perceptions of the impact of the evaluation on teaching improvements. The fourth section examines the overall reactions of the students toward the teacher evaluation system, and how they might relate to the various perceptions of the students.

4.1 Students' perceptions of the purposes of teacher evaluation

Most students commented that the purposes of the teacher evaluation system had not been explicitly explained to them by the University⁶. Nevertheless, they had their personal views of the underlying intentions of the University for setting up such a system.

Students' perceptions of the purposes of the teacher evaluation system can be grouped into four main categories: to improve the quality of teaching and learning, to monitor and appraise the lecturers for personnel decisions, to enhance the reputation and competitiveness of the university, and to meet the external demands for accountability. While it is possible for a student to include more than one view of the

6. A statement did appear at the top of the rating form stating the purposes of the evaluation. It is not clear, however, whether students felt that the purposes were not clear to them because they tended to neglect this in the exercise, or because they felt that the statement was not sufficient in helping them understand the purposes of the system.

purposes of evaluation in his/her perception, most students in the study showed a clear inclination to emphasise either one of the four views. The foci of the four perceptions were described below, and illustrated by relevant quotations from the interview transcripts.

A channel for collecting feedback

About 40 percent of the students perceived that the main purpose of the University's teacher evaluation system was to collect feedback from students to improve the quality of teaching and learning. They opined that with feedback from students, lecturers would be able to identify the weaknesses in their teaching or understand better the problems that students had encountered in their learning. Typical comments are:

I think that the purpose is to collect the views of the students on teaching, and provide some feedback to the lecturers so that they can improve. (K36 RS)

I think the main purpose is to enable us to learn better, and to find out how to help us learn better. (L28 RS)

A quality control measure

Another 40 percent of the students perceived teacher evaluation as a quality control measure by the University to monitor and assess the performance of the lecturers to ensure that they met required standards. They believed that student feedback was used by the University to appraise the lecturers for personnel decisions such as promotions or contract renewals, as in the following examples:

To evaluate the lecturers. I heard that their salary increments depend on the [student feedback] results. (L32 RS3)

I think that the main purpose is to help the University to determine whether or not the contract of a certain lecturer should be renewed. (K25 ENGL)

A marketing strategy

A small proportion (about 10 percent) of students believed that the University did this primarily for its own benefits rather than the students'. They argued that the ultimate aim of the University was to enhance its status and reputation, and to increase its competitiveness with other universities. For example, one student stated:

I think what the University wants is to maintain the quality of the lecturers of the University, to ensure that students can learn something, so as to uphold the reputation of this University and its graduates, and to compete favourably with other universities. (K08 SD)

A bureaucratic formality

Some other students (about 10 percent) were sceptical about the exercise, and believed that the system was a mere bureaucratic routine set up solely to demonstrate to outside agencies that the University had done something to assure its teaching quality. They alleged that the whole system was a 'show', and that the University was not genuinely interested in improving teaching:

I believe that the main purpose is to set up a system so that if other people ask about it, the University can claim that she has done something to collect students' feedback. I don't believe that it has any genuine influence on the lecturers. (L25 ENGL)

To the extent that people's perceptions often shape their behaviours, it is logical to speculate that the different perceptions of the students might influence the way they approached and reacted to the evaluation task. We would explore the relationship between the two in the later sections.

4.2 Students' beliefs about their roles in teacher evaluation

All students interviewed except four believed that university students should have the opportunity to evaluate their teachers, albeit with different justifications. The

other four students had some reservations about the system, not because they thought that students should not be given the opportunity to do so, but rather, because they thought that the existing system was ineffective, that a lot of time was wasted on it, or that they preferred a more direct dialogue with their teachers.

Students who believed that they should have the opportunity to evaluate their teachers tended to justify their positions on different grounds. A large proportion of students (about 40 percent) argued that students could provide useful information for the lecturers to improve their teaching and thus, should not be denied the opportunity. Some believed that the existence of student evaluations itself would put pressure on teachers to pay more attention to their teaching:

Definitely. In teaching, it is important to see if the student can really learn. If the lecturer was not able to make the students understand, he or she should be considered deficient. Furthermore, if a lecturer's teaching is poor, student feedback may enable him or her to improve. (K14 BUSS)

Yes... It at least ensures that the lecturers would not be too lazy or unprepared in their teaching. (K27 ENGL)

Many students (about 20 percent) opined that as recipients of the teaching, they were in the best position to evaluate the teaching of their teachers because they had the most direct, first-hand experience on a day-to-day basis. They argued that no valid judgement could be made on the teaching of a lecturer without considering inputs from students. One student said:

Certainly. I think that only students have the direct experience to assess the teaching of the lecturers. University administrators usually don't know anything about the teaching of the lecturers. If students were not given the chance, no one else will have the direct experience of the teaching to give a valid assessment. (K34 ENGL)

Some other students (about 20 percent) took the hard stance of a customer, and demanded the right to evaluate the providers of the service (i.e., the teachers) to ensure

that they got what they paid for. Some went further to suggest that their evaluation would enable the University to identify the incompetent lecturers and do something about them, including dismissal from employment. Two typical comments are:

Certainly, as we have to pay a high tuition fee for our education. It is like buying things in the market. As a consumer, I should have the right to choose what I like. Why shouldn't I have the right to influence which teacher we can get for our studies? In the university, we need a channel for making complaints, similar to the Consumer Council. (L25 ENGL)

Yes, to fire those who are not competent [in teaching]. (L10 SD)

A small but significant proportion of students (about 10 percent) believed that even if student evaluations had no influences on the lecturers, they should have the opportunity to evaluate their teachers because it at least provided a channel for them to air their grievances, express their dissatisfactions, or even retaliate by giving a low rating:

Definitely. It is better than not having the chance. If a teacher's teaching is really poor, the student can give him a '1'. Although it may be useless, students at least have a chance to express their dissatisfaction. Having a chance to complain is better than not having one. (K18 ITC)

I think that it... raises the status of the students: because now we have influences over the lecturers. Furthermore, we have the opportunity to retaliate and express our dissatisfaction, if we don't like the lecturer. At least, my views can be made known to other people. (K04 BRE)

While all these justifications are perfectly legitimate reasons *to the students* for taking part in the evaluation, they may have different implications in terms of students' expectations of how their feedback should be considered and acted on, which, in turn, may influence their perceptions of how useful the exercise was in achieving its purpose, their attitudes towards the rating process, and the stance they take when giving their feedback.

4.3 Students' perceptions of the impact of teaching evaluation

Student feedback is often collected with the stated aim of improving teaching. But to what extent did the students perceive that the teaching of the lecturers had actually improved as a result of their feedback?

When asked about this, most of the students either maintained that they had witnessed no change at all in the performance of their teachers (about 55 percent), or replied that they simply did not know (about 25 percent). A lot of students observed that their teacher were teaching essentially in the same way, and had made no attempt to change. Some students complained that there was no way for them to see if there was any improvements in the performance of the teachers because the evaluations were usually conducted at the end of the semester, and they were seldom taught by the same lecturers in the following year. Students were frustrated about the lack of effect of their efforts, and began to question the value of the exercise, as shown in the comments below:

I feel that the performance of the lecturers is more or less the same. I can see no differences at all. I think that students are quite frustrated because they cannot see any effect. (K35 RS)

I don't know if [the teachers] have improved or not, because most of them will not teach me in the following semester, after we fill out the forms... I have the feeling that some lecturers will continue with their own methods even after receiving our feedback... I feel that a lot of teachers are not concerned about teaching evaluations. (K15 BUSS)

Some students seemed to have a slightly more positive view, either because they had come across a few examples of observable improvement, or because they still had hope. For example:

A little bit, but not much... Frankly, I have witnessed some changes among the teachers, but how much of the changes were attributable to student feedback? I would say 20%. (L13 BUSS)

I believe that it would [lead to teaching improvements], because I trust that most teachers have the intention of teaching their students well. But the problem is: it is difficult to improve in a short period of time because it takes years for teachers to change... (K02 BRE)

However, this group of students was a minority (about 15 percent) among the respondents. Furthermore, it was quite obvious from their responses that the changes were quite small. Taking all the responses of the students together, it is safe to conclude that most students did not perceive that their feedback had a strong impact on the quality of teaching they received.

4.4 Students' overall reactions to teacher evaluation

As expected, students had widely different views and feelings concerning the teacher evaluation system. About 30 percent of the students believed that it is necessary and certainly has its value. Another 17 percent felt that having the system is better than none.

I think that it is necessary to have a system for collecting student feedback. It provides useful information for the teachers and for formulating future policies. (K02 BRE)

I believe that it is quite good to have a system for students to give feedback because some students think that we are stuck with the teachers and we can't do anything about it because we are only students. But when we have questionnaires, students can at least voice their opinions... (L24 RS)

I think that having such a system is better than having none. If the questionnaire forms are not administered, I might not even think about it... (K08 SD2)

Other students (about 30 percent) were more ambivalent in their responses. They were indifferent to the exercise, and filled out the forms as they were given without any strong feelings or expectations. One such student stated:

I don't have any particularly strong feeling towards the feedback exercise. I'll do it when I am given a form to fill out... I don't particularly like it, but I don't particularly hate it, either... I just do it as it is available. (K21 ITC)

A significant minority of students (about 18 percent) found the exercise dreadful and useless, and had very strong negative feelings about it:

We do it every year. I feel that it is just a cosmetic thing. Everything is the same [after completing the forms], no improvements seen. We don't know what will be done after we complete them. [The evaluation forms] may turn into rubbish. We don't know if there's any influence on the teacher. The poor teachers are still teaching in the same way. (L04 BRE)

Students also made a lot of critical comments on the format of the questionnaires and the implementation of the system. Their major comments or criticisms are summarised below:

- The questions on the questionnaire forms were not specific enough, and sometimes irrelevant to the teaching and learning contexts of their respective departments.
- The items focused exclusively on the teachers, but in some cases, the problems were with the subject content and workload, etc., which were not included in the existing questionnaire.
- Questionnaire surveys of this sort were very indirect and slow in reflecting their feedback, and they found it difficult to fully express their views through ratings. Some students preferred interviews in which they could have a direct dialogue with their teachers. However, some other students felt that they could not give their honest opinions when talked to their teachers in a face-to-face situation, and preferred the anonymity provided in the existing system.
- A lot of students reported that they found it tiresome to complete the many questionnaire forms at the end of the semester, which made them less and less interested in completing the forms in a serious and conscientious manner.
- Many students were uncertain about the usefulness of their feedback. They hoped that the system could be made more transparent, and their feedback taken up more seriously by the University or the lecturers concerned.

It is also interesting to note that students' overall reactions to the teacher evaluation system were related to their perceptions of the purpose and utility of the system. While there was not a clear-cut one-to-one correspondence in the relationship between the two, there were indications that:

- Students perceiving the teacher evaluation system as merely a formality to meet the demand for external accountancy tended to have more negative overall reactions to the system,
- Students who thought that the main purpose of the system was to monitor the performance of the lecturers tended to have a more positive view about the system, and
- Students who believed that teaching evaluation would have a positive impact on teaching improvements tended to view the system more favourably.

However, the conclusions are tentative and much more research is needed to establish relationships. Furthermore, association is not causation. The observed relationships alone do not prove that the perceptions of the students *caused* the differences in their overall reactions, nor do they imply that we can change students' overall feelings towards the system by changing their perceptions. Most likely, they are both the results of the students' prior evaluation experiences, which are dependent on the design and implementation of the system, as well as the reactions of the various parties involved. Furthermore, other factors such as students' levels of cognitive achievements, academic motivations, personalities, etc., might also come into play. Much more research is needed before we can come close to understanding fully why students might develop different perceptions of teaching evaluations.

CHAPTER 5 STUDENTS' ATTITUDES ABOUT THE EVALUATION PROCESS

This chapter focuses on students' attitudes about the evaluation process. The first section looks into the extent to which students were serious in the process of evaluating their teachers by examining both their own claims as well as their depictions of their classmates' behaviours in the rating process. It then explores the meaning of 'being serious' in the evaluation to the students, followed by a discussion on the context-dependent nature of students' attitudes in the evaluation process. The second section attempts to identify the various contextual factors that influenced students' attitudes, and examines how they might affect students' seriousness in making their ratings. The third section investigates the sources of anxieties of the students in the evaluation process.

5.1 Seriousness of the students in the rating process

The extent to which students were serious in giving their ratings is a question of concern because it directly affects the quality of data that will be generated. When students were asked about the manner in which they normally filled out the student feedback forms, a minority of students (less than 10 percent) admitted explicitly that they completed the forms in a hasty and less-than-serious manner:

I was not serious at all. Very often, I gave my ratings without reading the items.
(L25 ENGL)

Over 50 percent of the respondents, however, claimed that they filled out the forms in a very serious manner. Another 17 percent of the respondents reported that they were quite serious when making the ratings. The remaining proportion said that their attitudes towards the rating process varied, depending upon the specific context under which the evaluation was conducted.

However, it is possible that what students claimed may not be the same as what they actually did in practice. Also, different students may have different interpretations of being 'serious' in the evaluation process. It is therefore important to examine further what students really meant when they claimed that they were serious or otherwise in evaluating their teachers.

The meaning of being 'serious' in evaluations

From the students' responses, it appeared that students differed in their meanings of being 'serious' in evaluation. To some students, being serious just meant that they were not too mindless and erratic in their ratings, or that they had at least given some thoughts to the items when making their ratings. It did not, however, mean that they would consider each item carefully before deciding on their ratings, nor did it imply that they would try their best to put down their views of the teaching fully on the evaluation form. As two students explained:

I would not be too casual in completing the forms... (K33 ENGL)

I would give some thoughts to the items before making my ratings. But I seldom answer the open-ended questions at the back of the form. (K36 RS)

For others, they considered the performance of their teachers carefully against what the items asked when making the ratings, and might even go to some length to put down their comments and suggestions for the open-ended questions.

I did the evaluation quite seriously. I would try to recall how the teacher performed in class, or whether the comments they gave me were useful or not. (L10 SD)

I did it very seriously, and would make the effort to write down my comments on the two open-ended questions at the back of the form... I would think about each of the items and see if the performance of the lecturer met the requirements. (K34 ENGL)

Some students underscored their seriousness by stating that they gave ratings that honestly reflected their own views of the quality of teaching of the teachers. They believed that the whole exercise would be meaningless unless every student did the same:

I would give ratings that reflect my true feelings about the teaching of the lecturer. If he teaches well, I would give him a high rating; if not, I would give a low rating... If everybody gives every teacher a neutral rating, it is meaningless. (K08 SD)

A significant proportion of students admitted that they completed the forms in haste, but they had different explanations for doing so. While some students described their lack of seriousness by their haste in making ratings, others argued that they were quite serious in the evaluation although they were very quick in completing the forms. Familiarity with the items and the format of the evaluation questionnaire, they said, made it possible for them to do it conscientiously yet quickly:

I was quite serious in completing the forms, and would examine the items carefully. But as I am quite familiar with the items on the form now, I don't need a lot of time for studying the items [before completing it]. (K39 BUSS)

I would have already formed my views [about the teacher] before completing the form, because I am now very familiar with the items. I knew what ratings I would give as soon as I glanced through the form. Although I made my ratings very quickly, it didn't mean that I was not serious with the evaluation. I just know the questionnaire so well that I can give a rating almost immediately as I go through the items. (K04 BRE)

Analysis of the students' responses in different parts of their interviews further revealed that there were contradictions in some of the students' accounts of their attitudes in the process of evaluating their teachers. A student who claimed that she was 'quite serious' in the evaluation said in a later part of the interview that:

I feel rather disillusioned every time I fill out the questionnaire form. I don't have any incentive. I complete it very hastily. If I was in a good mood, I would choose '3' for all items. Otherwise, I will throw the form away. Losing one

form should not have any influence. I have a very strong negative feeling about the evaluation exercise... (L02 SD)

Another student who described himself as 'very serious' when making ratings for his teachers later admitted that:

Personally I do not like completing the evaluation forms. We get sloppier and sloppier about giving ratings, and the whole exercise has been reduced to a ritual or habit. (L05 BRE)

On the other hand, some students who contended that they were not very serious in doing the evaluations indicated that they did give some thoughts to the items in the process of rating their teachers, and believed that the system had some positive values:

I would glance through the items, although I would not examine them in great detail... I would take a cursory look... Although I sometimes find it very tiresome to complete so many questionnaire forms, I think that it is good to have such a system. (K12 BRE)

There is also a clear discrepancy in students' report of their own rating attitudes and their depiction of what their classmates did when evaluating their teachers. When students were asked to comment on the manner in which their classmates completed the evaluation forms, over 55 percent of the students mentioned that some if not most of their classmates did it in a casual or even erratic manner. For example:

Most of my classmates were not serious in completing the forms, because all the evaluations were conducted at the end of a class session. Many students would urge their friends to hurry up: "Be quick! Be quick! We have to go now!" Thus, everybody was anxious to leave the classroom and they would just rush through it as quickly as possible. (L24 RS)

I had the feeling that most students were not very serious, they would not examine the items and consider the performance of the teachers carefully before they decided on the appropriate ratings to give. They just did it in an arbitrary manner. (K22 ITC)

The results showed that while the majority of the respondents claimed that they themselves were reasonably serious in evaluating their teachers, they at the same time reported that some students in their own classes were not. While it is possible that this

discrepancy is legitimate and defensible, given the contradictory descriptions by the students of their own rating attitudes mentioned above, we suspect that some of it may be attributable to the conscious or subconscious attempts of the respondents to overstate their own seriousness in the evaluation process, for reasons of social desirability.

It therefore seems appropriate to conclude that being 'serious' in evaluation may have different meanings to different students, and may entail a wide range of evaluation practice or behaviours actually adopted by the students. In other words, when two students said that they were serious in the evaluation process, they might be referring to two totally different mind-sets and patterns of rating behaviours. It is also evident that students' attitudes towards evaluating their teachers are complex, and cannot be adequately understood by taking the students' self descriptions of their own attitudes at face value. It follows that to understand how students go about evaluating their teachers, it is necessary to look deeper into what they are thinking about when they are making the ratings, the pattern of actual ratings they give for their teachers, and the strategies and criteria they use to arrive at their ratings. These will be discussed in the following chapters.

Contextual influences on students' attitudes

There is evidence to suggest that students' attitudes in the evaluation process are not fixed but rather, influenced by a number of contextual factors. For example, a number of students maintained that their attitudes towards evaluating their teachers varied on different occasions, depending upon the circumstances under which the evaluations were conducted:

It depends. If I was free and had nothing to do afterwards, I would be more serious. On the other hand, if I was in a hurry to leave after the evaluation, I would do it hastily to get it over with. (L20 ITC)

I sometimes filled in the forms mindlessly when I felt tired and bored. But if I was serious, I would read each item carefully before making a rating. Yet, if the lecturer was good in teaching, I would give her a high rating no matter whether I was serious in completing the form or not. (K25 ENGL)

There is also indication that students' attitudes may change as a result of their changing perceptions of the reactions of their lectures towards student evaluations, and the perceived utility of the evaluation exercise in bringing about teaching improvements. A student explained that he was not interested in the evaluation because he felt that the lecturers would not take his feedback seriously:

I strongly feel that some lecturers will totally disregard our feedback because it has no effect on them... [They] will not pay any serious attention to it. (K24 ENGL)

Another student described how her attitudes in the rating process changed over time as a result of the lack of visible impacts on her teachers:

In Year 1, I tended to complete the evaluation forms seriously and conscientiously because I thought that my comments would be fed back to the department head. If the teaching of a certain lecturer was not good, I expected that our views would be reflected in our evaluations and the department would take some actions to improve the situation. But now I feel that my comments have not been followed up, and so I am less inclined to do the evaluation seriously this year. (K37 ENGL)

The quotations above clearly illustrate that students' attitudes in the evaluation process are not static. Instead, they are the results of a complex interplay between the students, the lecturers, and the contexts in which the evaluations are conducted. They further suggest that students' perceptions play an important part in determining their attitudes towards teacher evaluations, and how they would go about making ratings for their teachers.

5.2 Factors affecting students' attitudes in the evaluation process

To the extent that students' attitudes and behaviours are affected by the contexts under which the evaluations take place, it is imperative to identify the various contextual factors, and examine their impact on students' attitudes and behaviours in making ratings.

Analysis of the interview transcripts reveals that there are a number of factors influencing students' attitudes, including

- the specific arrangements for the evaluation, including: timing of the evaluation, time allowed for completing the forms, the length and format of the evaluation questionnaire, the frequency of evaluation, etc.;
- students' perceptions of the subject and/or the teaching;
- students' knowledge about the purpose, procedure, and use of the evaluation; and
- students' beliefs and perceptions of the benefits of the evaluation.

Evaluation arrangements

Students often mentioned that their attitudes in completing the evaluation forms were affected by the timing of and time allowed for the evaluation. As most of the evaluations were conducted at the end of a class session, students said that they were usually quite tired after the long lecture and were anxious to leave. If they had other commitments or obligations after the lecture such as lunch dates, tests, or practical sessions, they would tend to do the evaluation in a hurry and rush out of the lecture room as soon as they could. Some students alleged that they were usually given very little time to fill out the evaluation form. They blamed the insufficient time given for their hasty manner in evaluating their teachers. Some students maintained that they

would sometimes feel impatient with completing the evaluation forms because they found it monotonous and boring to fill out the same standardised questionnaire form over and over again, especially when there were a lot of items on the form and when they thought that some of the items were not appropriate for their departments or subjects.

Another aspect of evaluation arrangements that affected students' attitudes was the frequency of evaluation. Students complained that they had to complete a large number of questionnaire forms in a short period of time, because all the evaluations took place near the end of the semester. They said that they were fed up with it, and would tend to complete the forms in a routine and mechanical manner:

I am fed up with [doing the evaluation], especially when we are asked to complete so many forms in the short period of time before the end of the semester. It is hard... and boring. It may be better for the first few forms. For the later ones, we will not spend too much time thinking about them, unless we have a strong feeling – positive or negative – about the lecturer. For the rest, we just complete them casually to get it over with. (K09 BRE)

Students' perceptions of the subject and the teaching

Students' satisfaction with the teaching and perceived importance of the subject also influenced their seriousness in teacher evaluations. Normally, students would be more eager to fill out the forms when there were things that they felt were worth commenting on, for example, when they found the teaching of the lecturer particularly good or disappointing, or when they thought that the subject was of importance to them. As two of the students stated:

If students feel that the subject is important to them or that the lecturer really cares for them, they will treat the evaluation more seriously and give objective ratings on the lecturer according to what the items ask. If the students feel that the subject is not relevant to them, or that the lecturer doesn't care for their well-being, they tend to give their ratings in a less serious manner. (K38 BUSS)

I would be serious when rating the extreme cases. For the good teachers, I want to praise them. For the poor ones, I want to state my views so that they will reflect on their own teaching... (L08 BRE)

Knowledge of the purpose, procedure, and use of the evaluation

The transparency of the evaluation system to the students was also important. Many students said that they knew very little about the purpose of the exercise, how the evaluation data would be processed, when and to whom the results would be disclosed, and to what use the results would be put. For example:

I don't know what will happen to the forms after I have completed them. I don't even know when the results will be disclosed. Students simply know nothing about the evaluation after completing the forms... (K08 SD)

We don't know the overall evaluation results of the lecturer... We don't know if the results reflected what we felt, if the results were fed back to the lecturers, or if they were in fact utilised... We don't know why we should do it. (K09 BRE)

As a result, students often failed to see the significance of the effort they had to put in completing the forms, and became sceptical about the purpose of the exercise. Consequently, they were less inclined to do the evaluation in a serious manner:

The whole system is just a cajolery, to coax us that our views are valued and there is a channel for us to give our feedback to the University. The whole thing is just a 'show'. (L25 ENGL)

Perceptions of the benefits of evaluations

To a large extent, students' attitudes towards teacher evaluations were dependent on their beliefs and perceptions of the effects of the evaluations on teaching improvements. Some students maintained that they lost interest and confidence in the evaluation exercise because they perceived that their evaluations would not have any impact on the lecturers' performance, and believed that their evaluation would not bring about any significant improvements in the quality of teaching and learning:

I have doubts about its usefulness [for improving the teaching of the lecturers]. I am not sure that it really has any real influences. Furthermore, I don't know how many teachers will make use of the feedback... We do not have strong confidence in the system: we just do the evaluations like another piece of homework. (K01 SD)

I feel that the performances of the lecturers are more or less the same [after the evaluation]. I can see no differences at all. I think that students are quite frustrated because they could not see any effects of their evaluation, and that is probably why a lot of them are not interested in completing the evaluation forms. (K35 RS)

Other students found the exercise of low personal importance to them because even if it led to any improvements in teaching, it would only benefit the following cohort of students. They could not derive any direct personal benefits from the process themselves. As one student explained:

I think the reason why students are not interested in the evaluation is that it is no longer their concern. The course would be near the end when the evaluation was carried out. Even though we give our feedback, the tutor concerned will not teach us anymore... (L11 BUSS)

Differential perceptions among students

While many of the factors identified above are legitimate reasons for students to adopt a less-than-serious manner in the evaluation process, it seems that the fundamental factor is the extent to which they think that the evaluations are meaningful and valuable to them *personally*. Take the evaluation arrangements as an example: while some students complained about the timing, the length and format of the questionnaire, the frequency of evaluations, and the effort they had to spend on completing evaluation forms, other students found the same arrangements quite acceptable because they could see the potential value of the exercise. For example:

I do not find it tiring for having to complete the many feedback forms because it only takes us a few minutes to complete one. But the influence can be profound, and it will affect our own benefits – for example, when the same teacher teaches us later. (K02 BRE)

Similarly, while some students thought that the evaluations were meaningless because they could not benefit directly from them, other students thought otherwise. They believed that student feedback is meaningful as long as it can benefit the following cohort of students, as illustrated by the following quotation:

I think that student feedback is useful. It is helpful to the next cohort of students, even though it is not directly helpful to us. (L07 BRE)

Almost all students perceived that their feedback had no noticeable impact on the teaching of their lecturers or in personnel decisions. Yet, some students still held that student evaluation had its values:

Although we express our dissatisfaction on the evaluation form, we do not expect our feedback will lead to any substantial change. We know that it may be useless, but it is important for us to let the lecturer know that we are dissatisfied with her teaching... I think that the student feedback system is useful in that it allows students to express their views and dissatisfaction. We hope, though, that it has a stronger effect on the lecturers, for example, in their contract renewals or career advancements. (K24 ENGL)

We don't know whether the teacher has improved. We really don't know. Maybe, the next group of students can tell. But I believe that there must be some improvements... It must have some positive influence. It is better to have such a system than not having one. Perhaps, we just sowed the seeds but have not been able to reap the fruits yet. (K19 ITC)

Of course, this is not to argue that students' criticisms and dissatisfactions with the student feedback system can be ignored. Instead, it points to the fact that students' values and beliefs may influence their perceptions of the evaluation experience, which may in turn influence the way that they treat the evaluation process. If that is the case, then any changes in the evaluation arrangements or questionnaire format alone will not significantly alter the attitudes of the students in the process of rating their teachers, unless there are appropriate changes in the beliefs and perceptions of the students of the value and meaning of the evaluation at the same time.

5.3 Students' anxieties about teacher evaluations

Evaluation of teachers is a relatively novel experience for many students at university. It involves a temporary reversal of status between the teachers and the students in the teaching and learning process, and calls for reasoned judgements on the part of the students based upon their direct experiences. To some students, this is not an easy task, especially when there are important consequences attached to their evaluations. It is therefore not surprising to find that some students had anxieties about evaluating their teachers.

One of the anxieties arises as a result of the students' desire to have more influences on the teaching of their lecturers on the one hand, and their worries about the fairness and objectivity of their own ratings on the other. Students believed that their feedback should be seriously considered and listened to by the lecturers to ensure that they could receive high quality teaching:

Only with the feedback from students can teaching be improved. If students do not have the chance to evaluate their teachers, they will be at the mercy of their lecturers. (K26 ENGL)

Most students hoped to see that their feedback has some influence on their lecturers, especially in personnel decisions concerning contract renewals or promotions. However, they also felt uneasy that their feedback might affect the careers of the lecturers, as shown in the quotations below:

I... don't want to jeopardise [the lecturers'] career. (L25 ENGL)

I don't know whether our ratings will have any effect on the teacher's promotion. I've read the news that some university teachers' contracts were terminated because of the poor student evaluations they got. I am worried about this. Everyone has some weaknesses... (L27 RS)

This led to conflicting emotions among students, particularly those who felt that some of their classmates were too biased and subjective in their evaluations. They felt that it might be unfair to the lecturers:

I have conflicting views myself. I feel that most of the students at university do not study for learning. They may have a biased view of what a good teacher should be like. Thus, I sometimes fear that this system is not very fair to the lecturers. (K34 ENGL)

The concern for fairness to the lecturer also led some students to question their own ability to evaluate the teaching of their lecturers in an accurate and objective manner. Their worries are exemplified in the following two quotations:

Sometimes I feel that I myself am not that objective in my ratings, because my evaluation of the tutor may depend a lot on whether or not he accepts my project ideas. (K10 SD)

I understand what the item is trying to address, but there is no indicator to help me make my decision. I don't know on what I should base my judgement. (L02 SD)

Students sometimes found it difficult to decide to whom or what they should attribute their difficulties in learning. They recognised that sometimes it was the students rather than the lecturers that caused problems and thus, it would not be fair to give a low rating even if they were dissatisfied with a particular learning experience. The arbitrariness of the rating scale was another source of anxiety for some students:

I am quite confused sometimes because I am not sure whether it is the lecturer or me who has caused the problem. For example, when I cannot understand the subject matter, I may wonder if it is due to my own problems such as language deficiency, or the poor teaching of the lecturer. That is why in many cases, I hesitate to give very negative ratings... in the fear that I am unfair to the lecturers. (K39 BUSS)

I think that attitude towards a teacher is a very personal thing. I am not sure about the differences between 'disagree' and 'strongly disagree'. I find it difficult to make a decision in terms of five levels. (L02 SD)

Students also expressed anxieties over how their teachers might react to their ratings or comments, even though their response was anonymous and they were assured that their feedback would not influence their results. They were afraid that if they gave negative feedback on their teachers, it might ruin their relationship or even lead to retaliations from their teachers. Their fears are reflected in the comments below:

I am worried that the lecturer may know that it is our class that gives her a low rating. (L25 ENGL)

Since teachers have special status, students tend to be more reserved in giving their opinions... Teachers need to 'save faces' and so do the students. We have to be cautious... Moreover, we are worried that our handwriting can be recognised [by the teachers]. (L01 SD)

Although students were temporarily put in the position of evaluators when the teacher evaluations were conducted, their new role was short-lived and contradictory to the normal role they take in the teaching and learning process. As the teachers had some real 'power' over the students in the daily teaching and learning process, the worries of the students were not totally uncalled-for.

CHAPTER 6 STUDENTS' STRATEGIES FOR MAKING RATINGS

This chapter reports the findings concerning the general strategies that students used to derive the ratings for their teachers. The first section describes the pattern of the actual ratings that students gave in the ratings simulation. The second section attempts to characterise the various strategies or approaches that students employed in deriving their ratings for their teachers. The third section examines the underlying reasons for some students to avoid making extreme scores. The last section explores the extent to which students are influenced by the views of their peers in the rating process.

6.1 Student ratings for 'good' and 'poor' teachers

In the interview, students were asked to think of a 'good' and a 'poor' teacher they had encountered and rate them separately on six items drawn from the University's evaluation form before they were asked to report verbally how they made the ratings⁷. This section examined the actual ratings they made in the simulation. The purpose was to identify any patterns or themes that could be used to check against the results of the subsequent analysis of the interview data. One student refused to give ratings for a 'poor' teacher on the ground that he could not think of any. As a result, only 70 pairs of ratings were obtained and subsequently analysed.

The descriptive statistics of the ratings for the 'good' and the 'poor' teachers for the six items are summarised in Table 6, and the distributions of the various ratings are shown in Figure 2. Results showed that the mean ratings for the 'good' teachers were consistently higher than those for the 'poor' teachers on all of the six items. The mean ratings for 'good' teachers ranged from 4.4 for Item 4 to 3.6 for Item 6. The mean

7. See pp. 38-9 for a description of the method, and Appendix D for a sample of the rating form.

ratings for the 'poor' teachers, on the other hand, ranged from 2.6 for Item 2 to 2.0 for Items 4 and 5.

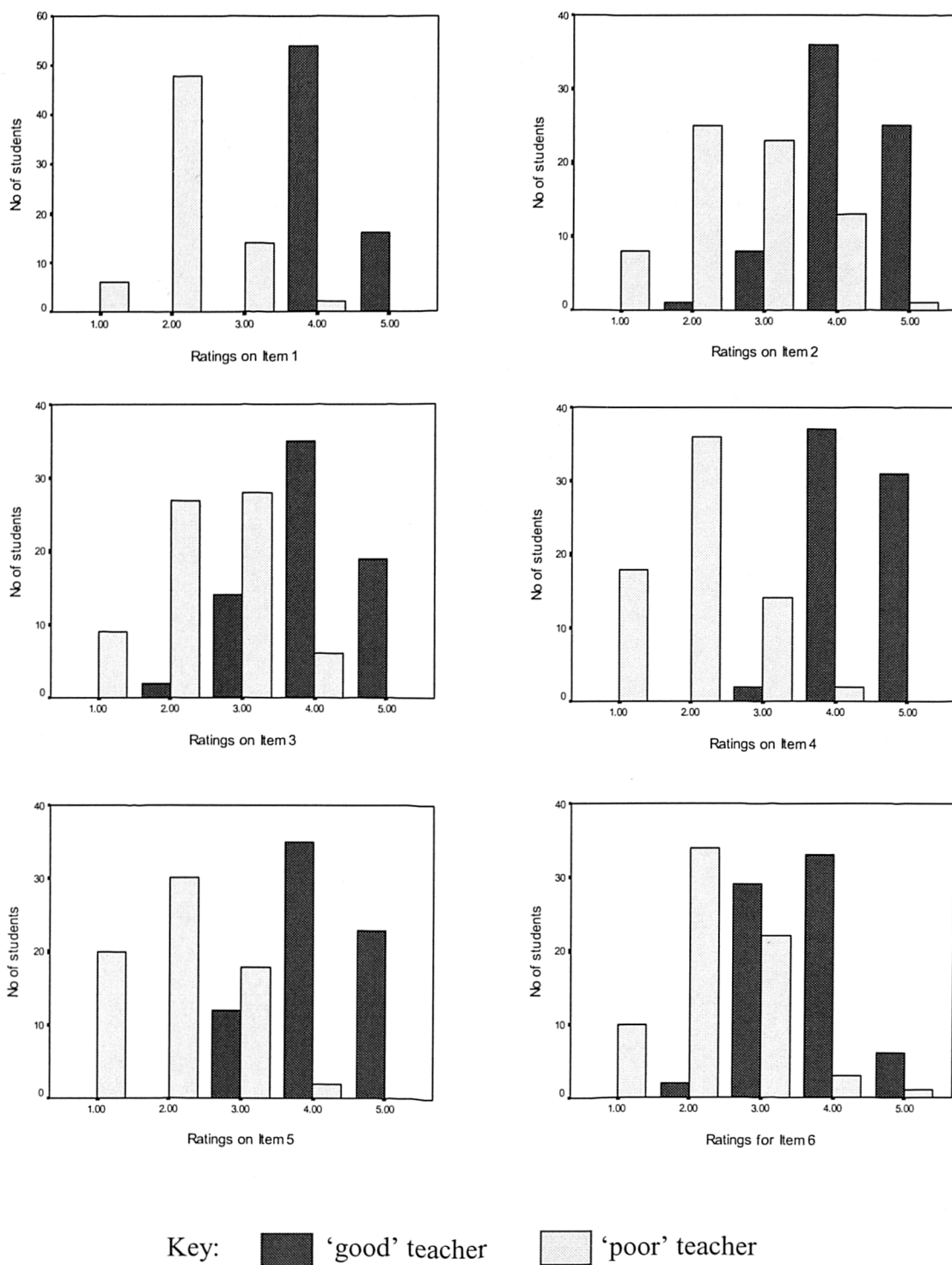
Table 6: Descriptives of the ratings for the 'good' and the 'poor' teachers

Item (paraphrased)	Teacher	Mean	SD	Max	Min
1. Method of teaching helped understanding	'Good'	4.2	.42	5	4
	'Poor'	2.2	.61	4	1
2. Encouraged active participation in class	'Good'	4.2	.70	5	2
	'Poor'	2.6	.97	5	1
3. Provided appropriate help for students	'Good'	4.0	.77	5	2
	'Poor'	2.4	.83	4	1
4. Presented subject materials clearly	'Good'	4.4	.55	5	3
	'Poor'	2.0	.76	4	1
5. Teaching stimulated interest in subject	'Good'	4.2	.69	5	3
	'Poor'	2.0	.82	4	1
6. Gave regular feedback on progress	'Good'	3.6	.68	5	2
	'Poor'	2.3	.82	5	1

N=70

That the 'good' teachers in the eyes of students as a whole tended to receive higher ratings than the 'poor' teachers did was also revealed in the distribution of the student ratings for the six items. The majority of the rating values for the 'good' teachers were in the range of '4' or '5', but those for the 'poor' teachers were normally at or below '3'. There were, however, noticeable differences in the distribution of the ratings across items. A significantly higher proportion of students chose a score of '3' for the 'good' teachers on Item 6 than on the other items. When the distributions of the ratings for the 'poor' teachers on the six items were compared, a relatively higher proportion of scores of '3' or above was found for Items 2, 3, and 6 than for the other items.

Figure 2: Distributions of the ratings for the 'good' and the 'poor' teachers



The distributions of the paired differences in ratings that students gave for the 'good' and the 'poor' teachers on the six items are summarised in Table 7.

Table 7: Distribution of the paired differences in ratings

Item (paraphrased)	Magnitude of paired difference [rating for 'good' teacher – rating for 'poor' teachers]						
	4.0	3.0	2.0	1.0	0	-1.0	-2.0
1. Method of teaching helped understanding	1 (1.4%)	16 (22.9%)	41 (58.6%)	10 (14.3%)	2 (2.9%)		
2. Encouraged active participation in class	5 (7.1%)	13 (18.6%)	18 (25.7%)	19 (27.1%)	13 (18.6%)	1 (1.4%)	1 (1.4%)
3. Provided appropriate help for students	3 (4.3%)	9 (12.9%)	27 (38.6%)	19 (27.1)	10 (14.3%)	2 (2.9%)	
4. Presented subject materials clearly	11 (15.7%)	18 (25.7%)	31 (44.3%)	9 (12.9%)	1 (1.4%)		
5. Teaching stimulated interest in subject	10 (14.3%)	13 (18.6%)	29 (41.4%)	12 (17.1%)	6 (8.6%)		
6. Gave regular feedback on progress	1 (1.4%)	8 (11.4%)	24 (34.3%)	20 (28.6%)	14 (20.0%)	2 (2.9%)	1 (1.4%)
All items	31 (7.4%)	77 (18.3%)	170 (40.4%)	89 (21.2%)	46 (11.0%)	5 (1.2%)	2 (0.5%)

The results again confirmed that students generally gave higher ratings for the 'good' teachers than for the 'poor' teachers. In 25.7% of the cases, the paired differences were larger than +2 points on a 5-point scale whereas in 40.4% of the cases, there was a paired difference of +1 point. The findings were hardly surprising because 'good' teachers were normally expected to attract higher ratings from students. However, in 11.0% of cases there was no difference in the pairs of ratings, and in 1.7% of the cases the 'poor' teachers received a higher rating than the 'good' teachers. Again, the distribution of the paired differences in ratings varied among items. For Items 1, 4 and 5, all the paired differences were larger than or equal to 0, and the proportion of paired ratings with a difference of 0 points was quite small, ranging from

1.4% (Item 4) to 8.6% (Item 5). On the other hand, a relatively larger proportion of students gave the same or lower ratings for the ‘good’ teachers compared to what they did for the ‘poor’ teachers for Items 2, 3, and 6.

To examine further if there were significant differences in the pairs of ratings for the ‘good’ and the ‘poor’ teachers on the six items, a series of paired *t*-tests were performed on each of the items. The results are summarised in Table 8. The mean paired differences in ratings ranged from +2.41 for Item 4 to +1.31 for Item 6, and all of the differences were statistically significant at the .01 level. The effect sizes ranged from 2.8 for Item 1 to 1.1 for item 6.

Table 8: Results of the paired *t*-tests of differences in ratings

Item (paraphrased)	Mean diff	SD	<i>t</i>	df	p	ES
1. Method of teaching helped understanding	2.06	.74	23.26	69	<.001	2.8
2. Encouraged active participation in class	1.58	1.30	10.19	69	<.001	1.2
3. Provided appropriate help for students	1.57	1.11	11.84	69	<.001	1.4
4. Presented subject materials clearly	2.41	.96	21.14	69	<.001	2.5
5. Teaching stimulated interest in subject	2.13	1.13	15.78	69	<.001	1.9
6. Gave regular feedback on progress	1.31	1.14	1.59	69	<.001	1.1

An examination of the pattern of the student ratings further revealed that students differed in the range of scores they actually used in rating their teachers. A significant proportion of the students avoided giving extreme scores, particularly at the lower end of the scale. Of the 70 students, only thirty-five (50%) gave ratings that

ranged from the highest score '5' to the lowest score '1' on at least one of the six items. Twelve (about 17%) of them did not include any extreme scores in their ratings on any of the items, while twenty of them (about 28.6%) gave ratings that excluded the lowest score. The remaining four (about 6%) gave the lowest scores to their teachers but not the highest score.

In summary, the pattern of ratings given by the students in the rating simulation showed that there were significant differences in the ratings students gave for the 'good' and the 'poor' teachers. The results implied that it was possible to broadly differentiate between 'good' and 'poor' lecturers in the eyes of the students by the ratings they gave. However, differences were found in the distribution of the ratings across items. Students seemed to be more able to differentiate between the 'good' and the 'poor' teachers on some items (e.g. Items 1, 4 and 5) than on the others (e.g. Items 2, 3 and 6) in terms of the ratings they gave. There was also evidence to suggest that students, at least some of them, would not routinely give high ratings for a teacher whom they thought was 'good'. In some cases, they might give the same or even lower ratings for the 'good' teachers they had in mind than the 'poor' ones. Finally, there were differences among students in the range of scores they used in rating their teachers. While some students used the full range of scores in their ratings of their teachers, a significant proportion tended to avoid giving extreme scores, particularly those at the lower end of the scale.

The analysis above examined what rating values students gave for different teachers but not how or why. In the following section, the general strategies that students used to derive the ratings are examined. The specific criteria and evidences that students used in making ratings for the individual items, and the meanings underlying their ratings will be investigated in Chapters 7 and 8.

6.2 General strategies for deriving ratings

Most students maintained that their ratings were not spurious but rather, based upon their conscious decisions. They claimed that they had tried to be fair and accurate in rating their teachers, and reported to have used a number of strategies to arrive at the specific rating values for the items. It is possible to characterise students' strategies in rating their teachers by four major dimensions:

- General criteria: relying on global impressions versus considering item-specific performance of the teacher
- Reference point: comparing the performance of the teacher with that of other lecturers versus comparing it against students' own interpretation of the items
- Focus: focusing on own experience versus considering other students' reactions
- Disposition to giving extreme ratings: using the full range of the scores versus avoiding extreme scores

It must be pointed out that for each of the dimensions described above, the strategies should be conceived as the opposite poles of a continuum rather than two discrete categories. In the interviews, students did show varying degrees of adherence to the strategies. In some cases, the same student might change the way they gave their ratings, depending upon the conditions under which the ratings were made.

Global impression versus item-specific considerations

For this dimension, students at the global impression pole tended to rate their teachers according to their overall impression of their teaching, as illustrated in the quotation below:

When giving ratings, I rely mostly on my overall impression of the teacher. I seldom examine the individual items in detail. (K17 BUSS)

Students adopting such an approach to making ratings would give a high score to lecturers whom they felt good about; they might even disregard what the item purported to ask. The following quotation clearly illustrates this point.

I would base it on my impression of the lecturer. If I thought that the lecturer deserved a '5', I might give her '5' on every item, irrespective of what the item was specifically asking about. (K17 BUSS)

Students at the item-specific consideration pole, on the other hand, focused on the teaching behaviours of the teacher and made judgements according to what they thought the item was about:

[My ratings] depend on what the items on the evaluation form asked about. The different items clearly asked about different aspects of the teaching of the lecturer, for example, whether the staff member presented the subject materials clearly, etc. Then, I would consider how the lecturer had communicated the materials to me in the lessons, and gave an appropriate rating. (K07 SD)

Students who adopted this approach tended to rate the items independently, and tried not to let their rating on a particular item be influenced by their overall impression of the teacher, or the performance of the teacher on other aspects of the teaching, as explained by one student:

There are many items on the evaluation form. I will rate each item independently. I won't give low ratings to teachers on the ground that I don't understand their teaching or I have got low marks in the subject. I'll consider the items separately. Even for a teacher who does not teach very well, if he has encouraged us to participate actively in class, I will rate him high on this item accordingly. (L19 BUSS)

While most students tended to resort to either one of the two approaches described above, some students reported a combination of the two approaches in making their ratings. One student explained how she used a combination of the two strategies in forming her ratings:

I have already formed in my mind impressions of the different teachers, and assigned a grade mentally to each of them according to my impression. My

ratings are probably based first on my global impression of the lecturer, and then with reference to the specific items on the feedback form. I'll generally give higher ratings for the 'Grade A' lecturers and lower ratings for the 'Grade C' lecturers. (K13 SD)

Furthermore, there was evidence that the more ambiguous the item to the students was, the more likely the students would rate the item according to their global impressions of the teacher. We shall discuss this further in the following chapter when we examine students' interpretations of the items.

Other lecturers' performance versus students' own standards

Students need a reference point for determining the specific rating value to give for a particular teacher on an item. Analysis of students' responses suggested that students turned to either one of the following two reference points in the process of rating their teachers: (a) the performance of other lecturers, or (b) the students' own standards resulting from their interpretations of the items.

Some students rated their teachers by comparing their performances with the other lecturers they had encountered, either in the past or in the semester they did the evaluation. Their ratings, thus, reflected their perceptions of the *relative* merits of the teaching of the lecturer concerned as compared to other lecturers:

I would definitely compare the lecturer with other lecturers when I made the ratings. It would be difficult to rate the lecturers separately. When you compared the performance across lecturers, you could easily tell who was a good one and who was really poor in teaching. (K24 ENGL)

In this case, the actual ratings they gave would be determined not only by what the teacher did or did not do, but also by the common practice adopted by other teachers at the university.

Other students tended to make ratings according to their perceptions of the extent to which the performance of the lecturer matched the requirement of the item, according to their own *personal* interpretations of what a 'good' teacher should do in that aspect of teaching. One student explained:

I would not compare a tutor against others but rather, make judgements on them independently of one another... I tended to compare the performance of the lecturer with my own ideas of what an 'ideal' lecturer should do... and examine the extent to which he met my expectations. If he didn't, I would give him a lower rating. (K06 SD)

A high rating that students gave, using this strategy, reflected a match between their expectations and the actual practice of the teachers, while a low rating signified a mismatch between the two. In other words, these students' implicit theories of teaching played a significant role in determining the actual ratings they gave for their teachers.

Cross-tabulation of students' strategies according to the two dimensions described above suggested that students who focused mainly on their global impressions of teachers were more likely to compare the performances of the teachers with that of the other teachers to arrive at their rating. On the other hand, students who focused mainly on the item-specific behaviours of the teachers tended to compare the behaviours of the teachers against their own expectation of how the teacher should perform with respect to the item. However, given the small sample size and qualitative nature of the data, the relationships observed are merely indicative and need support from further research.

Self experiences versus other students' reactions

There were differences across students in the extent to which they took into consideration the reactions of other students in class when making ratings for their teachers. Some students focused primarily on their own feelings and experiences – for

example, whether they *personally* liked the teaching or found it helpful to their understanding of the subject matter:

When determining the actual ratings for a lecturer, I would consider if the lecturer is good or not... I would base my ratings on my own experience – for example, whether I think he is a good lecturer, or whether I could understand his teaching or not. (K39 BUSS)

Other students believed that their own experiences alone might not be comprehensive or typical enough, and said they would consider the experiences or reactions of other students when deciding on their ratings. They argued that it might not be wholly the responsibility of the teacher if students had problems in learning and so, they should take the stance of the whole group when giving ratings:

I based my ratings not only on how the teacher interacted with me, but also on his general attitude and ways of interacting with other students... I think that in teaching surveys of this sort, we should evaluate the teacher from the point of view of the students as a group, not as individuals... Even if I personally found his teaching not good, I would rate him according to how he taught the whole group rather than how he taught me. (K01 SD)

Using versus avoiding extreme scores

In Section 6.1, we have shown that there were distinctive differences among students in the range of rating values they gave for their teachers. The finding was supported by the interview data. Students reported different dispositions to using the extreme scores in their ratings. Some students said that they would use the full range of scores in rating the ‘good’ and the ‘poor’ teachers:

I believe that I have given both ‘5’s and ‘1’s for some of my lecturers... I am not worried about giving extreme scores if I think that they are very good or extremely poor in teaching. (K29 RS)

I don’t hesitate to give either the highest or the lowest scores for any teacher, as long as he was not present when I was completing the form... (K37 ENGL)

Other students were more hesitant about giving extreme scores. They tended to avoid giving the highest or the lowest scores for their teachers:

My responses were not extreme. I seldom gave a '5' or a '1'. I cannot make myself do it because... they are after all my teachers, I should not be too harsh on them. I usually give ratings from '4' to '2'. If the lecturer is good, I will give him a '4'. (K15 BUSS)

The differences among students in their dispositions to giving extreme scores will have no effect on differentiating between 'good' and 'poor' teachers in the eyes of students when student ratings were examined *individually*. However, to the extent that students using extreme ratings will have more influence than those who did not on the class-average ratings, the composition of the students with different dispositions to give extreme ratings will have a significant effect on the actual class-average ratings received by lecturers.

6.3 Why students avoided giving the minimum rating value

That student ratings of university teaching were positively biased is a well-established finding in the literature. Yet, few studies have examined why this occurs. A number of reasons for not giving the lowest ratings for their teachers have been suggested by the students in the study. One possible reason, of course, was that the teaching was generally satisfactory and thus, did not warrant the minimum score:

I have never given a rating of '1' or '2' for any of my lecturers before. Maybe, none of the lecturers that I have encountered so far were that poor in teaching. (K33 ENGL)

Yet, the results also revealed that even when students were dissatisfied with their experiences of learning, they might still refrain from giving the minimum score for their teachers, for the following reasons:

- Students didn't want to be too mean or harsh with their ratings. They were concerned about 'saving the face' of their lecturer. For example:

I hesitate to give '1' or '2' because in the Chinese culture, we feel that we need to 'save the face' of our teachers. If I give a score of '1', it seems to be too mean. (L25 ENGL)

- Students were not sure about the meaning of the item and thus, chose the middle-range scores to 'play safe':

If the item is so clear that I can determine if the performance matches the description of the items clearly, I will choose a score that truly represents my view. But if the item is not clear, I will choose 'neutral'. (L05 BRE)

- Students thought that it might be (at least partly) their own fault that had caused the problem. Thus, they rated the teachers more leniently so as not to be unfair to them, as shown in the following comments:

Even if I feel that the lecturer is not teaching well, I tend to choose '3' (neutral) because I am not sure if it is my own problem that I fail to adapt to the teaching of the lecturer. It may not be the lecturer's fault. (K39 BUSS)

I believe that most lecturers have made some effort in their teaching. The main issue is whether students find the method acceptable to them. The fact that I do not accept the method does not mean that the lecturer has done something wrong in her teaching. Thus, I always hesitate to give a rating of '1'. (K30 ENGL)

- Students recognised that some of the problems were a result of the contextual factors which were outside the control of the lecturer. To be fair, they avoided giving a low rating even if the performance fell short of what the item required:

One of the items asks if the lecturer is able to provide individual help. For a large class, it is impossible for the lecturer to achieve this but we are still required to answer it. In this case, I would give a '3'. (K25 ENGL)

- Students found it difficult to interpret the values on the rating scales, particularly the meaning of the highest and lowest scores. As a result, they tended to choose the scores in the middle and avoided both the highest and the lowest scores:

I feel that the rating values are rather abstract. I don't really know what they really stand for... It is hard to tell exactly what '1' means and what '5' means.

That is why I usually give '3' or '4' unless there is something extraordinary about the teacher. (K06 SD)

The findings clearly showed that some of the students had a tendency to avoid giving the highest and in particular the lowest ratings, even if they were very satisfied or dissatisfied with the teaching of their lecturers. However, it was not possible to tell from the data which of the reasons above contributed most to the observed positive biases in ratings. It is logical to speculate that some or all of them have some influences on the actual ratings.

6.4 Peer influence on student ratings

The study also examined whether and how the students were influenced by their peers in their ratings. Students were asked whether they discussed with their classmates the teaching of their teachers in their daily interactions as well as during the rating process. They were also asked whether, or to what extent, they thought their ratings were influenced by the views of their classmates.

Students stated that it was highly uncommon for students to consult each other in the process of making ratings. The majority (over 60 percent) of students reported that they did not discuss their ratings with their classmates when completing the feedback forms. Some (about 10 percent) students reported that they might look at each other's ratings, but only after filling out the form. Some students (about 8 percent) said that they would discuss with their classmates when they had doubts about the meaning or interpretation of the items, but not the ratings themselves. A few students reported that on some occasions, they heard classmates shouting to their classmates that they should give a low rating for a certain teacher. But they perceived those suggestions as expressions of frustration and dissatisfaction rather than an intention to actively

influence other classmates' ratings. Only a minority of students (about 15%) admitted that they discussed their ratings with their classmates.

On the other hand, almost all students reported that they did talk about the teaching of their teachers in their daily interactions, e.g., during informal chats with their classmates between lectures or over lunch, normally with their close friends or classmates in the same work-groups for tutorials or projects. The interesting thing they observed was that most of the time, their conversation centred on the poor teaching of some of the teachers rather than the good practices that some other teachers adopted. Most students said that they seldom talked about the good teachers or good teaching in their conversations:

We seldom talked about the good performance... We won't say things like "this teacher is very good, she makes good use of transparencies." We don't have this kind of conversation. We only talk about the teachers when we want to air our grievances. (L30 ENGL)

We definitely talk about our teachers – especially when the tutor is irresponsible, talks nonsense, or wastes our time in the tutorials. We will criticise the tutor in our conversation. (K06 SD)

Students felt that it was perfectly all right to criticise their teachers in front of their classmates, but it would be grossly improper if they praised the teaching of their lecturers in such conversations, even if they personally found the teaching really good. They were worried that other students would be suspicious of their intention and so, they refrained from talking about good teaching/teachers openly with their classmates, as explained by one student:

We usually talked more openly about poor teaching than about the good ones. It is because different students may have different ideas of a good teacher. If you say in public who you think is a very good teacher, other students may think that... you want to become the teacher's pet. But if you criticise a teacher, other students may agree and join in the discussion. (L05 BRE)

Most students believed that their ratings were not influenced at all by the views of their classmates. They argued that since they had the direct experiences of the teaching, they would not be easily affected by other people's views, particularly if the views did not match with their own experiences. Furthermore, they believed that as university students, they were mature enough to make independent judgements. Even students who felt that their ratings might be influenced by the views of their classmates believed that such effects, if they existed at all, would be quite small:

As university students, we are mature enough to have our own opinions and viewpoints... I believe that every student has his/her own views, and nobody is able to influence his/her evaluation of the teacher. (K01 SD)

I think that other students may influence my ratings, but only to a small extent... Besides, we usually base our views on our own experience of interacting with the lecturer. If we do not know the teacher, we may be influenced by others. But if we know the teacher, we shall base our ratings on our own experience and impression. (K39 BUSS)

The results certainly do not imply that students are completely immune from the influence of their peers in their rating processes. Instead, they suggest that the influence might not be explicitly recognised by the students themselves, and it was not normally effected through peer discussions in the process of rating. Given that students often discussed the teaching they received with their close friends and classmates in their daily conversations, it is highly probable that students may develop a shared implicit theory of teaching, which indirectly influences the ratings they give when they are asked to evaluate their teachers. Students admitted that their views concerning their teachers had been generally consistent with those of their classmates. Besides, there was pressure for them to conform, as seen from the following comments:

We sometimes differ in our views, but this seldom happens. Most of the time, we have the same feeling towards the teachers. We've shared the same experience. It is easy to differentiate between good and poor teachers. For example, there are lazy teachers who give us poorly organised worksheets, or leave half an hour earlier than the scheduled time. There are also good teachers

who are knowledgeable about the subject matter and able to present it clearly to us. (L23 ENGL)

Our opinions were quite similar. Even if we have different views, we seldom voice them out. We'll listen to the majority view. We feel embarrassed to express a different opinion. It is not worthwhile to argue over these things. (L11 BUSS)

The discussions so far focused primarily on the general strategies that students adopted for rating their teachers. The actual criteria, evidence, and standards that students used for making ratings for the 'good' and the 'bad' teachers on each of the evaluation items will be described and discussed in Chapters 7 and 8 below.

CHAPTER 7 CRITERIA AND EVIDENCE THAT STUDENTS USED FOR MAKING RATINGS FOR INDIVIDUAL ITEMS

How students actually arrive at their ratings is an important but unresolved question in the literature on student evaluations of university teaching. This will be examined in Chapters 7 and 8. This chapter focuses on the criteria and evidence that students used to derive the specific ratings for the six items on the evaluation form. It reports the findings for each of the items, one at a time. For each item, the criteria or reasons that students gave for determining their ratings are examined. Key words or phrases that characterised their positive and negative evaluations are summarised and reported. Major themes or participant constructs that emerged from the interviews are also identified and discussed. Relevant quotes from the students' interview transcripts are used for illustration, wherever appropriate.

In Chapter 8, we shall examine students' interpretations of the rating scales, and discuss some of the issues in relation to how students arrive at their ratings.

Item 1: The staff member's method of teaching has helped my understanding.

Table 9 summarises the reasons that students gave for making high or low ratings for their teachers on Item 1. Students broadly used two different approaches to making their ratings for this item. Some relied mainly on their assessment of the extent to which they could understand the subject matter taught by the teacher in class, or whether they had learned something useful to them. They focused more on the degree of understanding they had attained of the subject materials than on what the teachers did in their teaching. Very often, this involved subjective judgements on the part of the students concerned:

Table 9: Criteria/evidence that students used in making ratings for Item 1

Item 1: The staff member's method of teaching has helped my understanding	
<i>Reasons for high ratings</i>	<i>Reasons for low ratings</i>
<ul style="list-style-type: none"> • Students' subjective feeling that they can understand the subject matter/ follow the teaching (24) • Teacher able to explain/express ideas clearly (19) • Teacher giving useful, real-life examples in teaching (16) • Teacher implementing student-centred activities that stimulate students to participate and think (13) • Teacher well-prepared, providing useful notes / learning materials (11) • Teacher conscientious and enthusiastic towards teaching, showing care and concern for student learning (10) • Teacher being systematic and organised in teaching (8) • Students' subjective feeling that they can gain something useful from the teaching/ able to apply or master the subject knowledge (6) • Teacher using a variety of methods audio-visual aids in teaching (6) • Teacher able to make students see the reasoning behind theories (3) • Teacher able to develop a class atmosphere conducive to learning (2) 	<ul style="list-style-type: none"> • Students' subjective feeling that they cannot understand the subject matter/ follow the teaching (17) • Teacher just talking/ reading out notes in class (13) • Teacher giving disorganised, confusing presentations (10) • Teacher not serious nor enthusiastic towards teaching / not caring if students understand or not (10) • Teacher not receptive to students' views or difficulties (7) • Teacher having poor presentation skills, unable to explain / express ideas clearly (6) • Poor preparation and content, poor notes and / or overhead slides (5) • Teacher unable to relate to real-life situations (2) • Students' subjective feeling that they can gain nothing / only superficial learning (2)

Notes: 1. The number in brackets represents the number of times the reason appeared in students' responses

2. Only items with number of mentions of two or above are included

I would consider if I could master the knowledge taught by the lecturer – how much I have learned, and the extent to which I can apply the knowledge. If, after the lecture, I still didn't know what I was supposed to learn, or did not understand the materials taught in class, I would think that the method of teaching has not helped my understanding. (K24 ENGL)

Other students tended to base their ratings on their observations of what the teacher did in their teaching to facilitate their understanding. They often quoted evidences that related to the teacher's behaviours inside or outside the classroom to account for the ratings they gave:

The 'good' teacher used many different methods, e.g., newspaper cuttings, group discussions, videos, etc. in her teaching... She collected news about the current affairs and economy, and led us to investigate those issues. The 'poor' lecturer used only one method – following strictly what was in the textbook. (L23 ENGL)

I would think about whether the lecturer could explain the subject matter clearly, whether he emphasised the reasoning behind the theories, whether he would spend a lot of time preparing diagrams and notes to aid his explanation, and ask students to raise questions and discuss ideas in class. I would give a teacher a high rating if he did the things above. However, if the lecturer simply talked in class and did nothing else to help us learn, I would give him a lower rating. (K15 BUSS)

Although the item asked specifically about the teacher's methods of teaching, a lot of students focused on the skilfulness of the teacher in implementing the methods rather than the methods *per se*. In fact, most of the comments made by the students were related to the conventional teacher-directed lecture and discussion methods of teaching; only a few students mentioned the use of the more innovative student-centred methods as the reason for their giving a high rating. The differences in ratings were primarily attributed to what the teacher did to make the lectures or discussions more effective in terms of helping them understand, rather than the particular teaching method that the teachers adopted. For example:

I would consider the presentation method of the lecturer – whether he just talked without any aids, whether he used rough hand-written transparencies or good

quality Power Point presentation, whether he made use of models, or gave demonstrations, etc. (K28 RS)

Students also made reference to the attitudes of teachers towards teaching and the students when determining the ratings for this item. They perceived a definite positive relationship between teachers' attitudes and their own learning:

I would consider the attitude of the lecturer towards teaching – what he wants us to learn and achieve in class, whether he uses daily and practical examples in his explanations, whether he was enthusiastic and interested in teaching, and whether he has prepared some notes for us... but I would focus... mainly on his attitudes. (K20 ITC)

The 'good' lecturer uses different methods and materials to help us understand. She really cares and keeps asking us if we understand. The 'poor' lecturer doesn't care if we can understand or not. She just goes through the materials and that's it. Even if we tell her that we don't understand, she will either ignore us or just explain it again in a very impatient manner. (K30 ENGL)

While teachers' enthusiasm and seriousness towards teaching and their empathy for students are likely to affect how they teach and consequently how well students learn, their inclusion as a criterion for judging teachers on this item appears to be not directly relevant. At least, it can be argued that teachers' attitudes are not part of the teaching methods of the staff member, which is supposedly the focus of this item.

A minority of students found it difficult to make their ratings on this item because they felt that their inability to understand the subject material might be partly attributable to factors unrelated to the teacher concerned. For example, the problem of the syllabus and textbook, the difficulty of the subject matter, or the lack of background knowledge or ability of the students, etc., may contribute to their lack of understanding:

The problem may not be with the lecturer. The textbook is poorly written ... the curriculum is problematic and the teacher is obliged to follow it. (L30 ENGL)

As a result, they tended to give a 'middle' rating even though they felt that they could not understand the teaching very well.

Item 2: The staff member encouraged active participation in class.

The major reasons that students gave for making high and low ratings for this item are summarised in Table 10. A number of themes emerged when students' responses to this item were examined. Unlike what they did for Item 1, most students focused on the presence or otherwise of certain teacher behaviours that encouraged students to participate in class when making ratings for this item. They tended to give higher ratings for teachers who made efforts to motivate students to participate in class activities and discussions. On the other hand, they gave low ratings to teachers who just talked and did not make any attempt to ask students questions or to encourage students to ask them questions. For example:

I would consider the teaching in class – whether the lecturer would ask students questions, encourage students to express their views on certain issues, or give real-life examples to stimulate students to think, rather than simply talk about the abstract theories from the textbooks... The 'poor' teacher is totally different. He never asks us any questions or encourages us to discuss in class. It seems to me that he is there just to go through the syllabus, and both the lecturer and the students are waiting impatiently for the class to end. (K20 ITC)

Some other students focused on the quality of the interactions rather than the chance for participation when making their ratings. Their ratings were based not on whether the teacher initiated any discussions or asked any questions in class but rather, the extent to which the discussions and questioning were useful to their learning or stimulated them to think. Some students also stressed the importance for the teachers to get the relatively inactive students to participate, and to treat students equally in the process.

Table 10: Criteria/evidence that students used in making ratings for Item 2

Item 2: The staff member encouraged active participation in class.	
<i>Reasons for high ratings</i>	<i>Reasons for low ratings</i>
<ul style="list-style-type: none"> • Teacher giving in-class tasks that encourage students to discuss and present ideas (25) • Teacher asking students questions in class (22) • Teacher taking active steps to find out students' difficulties in understanding, and providing chances for them to ask questions (19) • Teacher able to get students to participate, e.g., by inviting them by name to answer questions (12) • Teacher encouraging students to express their views in class (9) • Teacher able to stimulate students to think (9) • Teacher using active learning methods that motivate students to participate (6) • Other students appear to be attentive and actively involved in class (6) • Teacher having an open and friendly manner, and receptive to students' questions /answers (5) • Teacher encouraging students to attend the lecture tutorial sessions (5) • Teacher giving 'breaks' for students to think and ask questions (3) • Useful discussion and feedback in class, learn something useful (3) • Teacher encouraging students to get help from him/her after class (2) • Students' feelings of being motivated to learn more after class (2) 	<ul style="list-style-type: none"> • Teacher just talking, seldom asking students to discuss or express views (23) • Teacher asking questions but students were not interested / motivated to participate (15) • Discussions or tasks not useful / with no teacher feedback (9) • Difficult to achieve in mass lecture situations (7) • Teacher showing negative reactions to students' questions, e.g. showing impatience, mock at students' mistakes, etc. (6) • Teacher just asking students casually to ask questions if they have any problems (3) • Poor social relationship and atmosphere in class (3) • Teacher not caring if students are attentive / involved in class (3) • Teacher seeming not enthusiastic or able to help when students have questions (2) • Teacher 'forcing' students to participate by giving marks for participation or answering in-class questions (2)

Notes: 1. The number in brackets represents the number of times the reason appeared in students' responses

2. Only items with number of mentions of two or above are included

The following two quotations clearly illustrate such views:

The ‘good’ lecturer stimulates us to think, encourages us and treats every student equally – unlike the other lecturers who are nice only to the ‘good’ students and seem reluctant to answer questions from other students. He praises us for our good work and encourages us to think harder if our answer is not correct. The ‘poor’ lecturer also asks us questions, but the questions do not stimulate us to think. (K37 ENGL)

I would think about the tutorials in which some students stayed quiet and did not speak their mind. Some tutors did nothing about that, but some other tutors would do something to encourage them to take part, e.g., by calling them by name to answer questions. (K07 SD)

A significant proportion of students commented that teachers were quite alike in that most of them seldom encouraged students to participate in class. It is not clear, however, whether the perception is due to their inability to differentiate between the variations in practice among teachers, or that the teachers actually behaved as perceived. One student observed:

All the teachers are the same – they talk most of the time and there are few opportunities for participation. (K18 ITC)

On the other hand, some students pointed out that although teachers used essentially the same method, the reactions and levels of participation of the students were vastly different in different classes. In some classes, even though the teachers asked questions or provided opportunities for students to discuss or express views, students remained non-participative or just participated in a superficial way without rigour. Students claimed that their ratings were therefore based on the actual participation and reactions of the students, rather than what the teacher did in class. They reasoned that the differences in their reactions were attributable to their relationship with the teacher and the classmates, the usefulness of the input and feedback from the teacher, as well as their own problems such as the lack of interest, diligence, understanding, or initiative in learning the subject:

The 'good' lecturer always asks us to discuss during tutorials. Sometimes, she will ask us questions, e.g., what we think about what she has said. In this way, everyone has a chance to speak. The 'poor' lecturer also gives us the chance to speak, mainly through group discussions. However, our participation level is different. Our participation is more serious in the class of the former. In the case of the latter, we are not serious, and will discuss only in a superficial way. We know that the teacher is not enthusiastic. Even if we are serious in discussing the questions, it won't be helpful to our learning. (L23 ENGL)

Some students questioned the relevance of this item for evaluating the performance of teachers in mass lectures. They felt that as it is much more difficult to encourage student participation when lecturing to a large class and so, ratings for this item may not be fair or meaningful to those lecturers. As a result, they tended to give 'middle' ratings, or based their ratings on their overall impression of the teaching performance of the teachers:

It is difficult to decide. I don't know what active participation means. It is difficult to achieve this in mass lectures, and I can hardly recall any experience of teachers trying to encourage me to participate actively in class. To make ratings for this item, I try to base it on my experience. If there is none, I will rely on my impression. If the lecturer seems receptive to our questions, I'll tend to choose 'agree'. Otherwise, I'll choose 'disagree'. But I won't give any extreme scores. (L26 RS)

Some of the interpretations by the students were quite unexpected. For example, some students evaluated the teachers on this item on the basis of whether they encouraged students to attend the class. They gave higher ratings to teachers who made efforts to ensure high attendance. Others focused on whether the teacher was able to make students attentive in class. Still others based their ratings on whether the teacher motivated them to read and learn more after class, or if the teacher encouraged them to approach him/her for help outside the classroom. It may be argued that such personal interpretations are not wholly consistent with the focus of the item and thus, the validity of the ratings based on such interpretations are questionable.

Item 3: The staff member provided appropriate help for students with learning difficulties.

Table 11 summarises the criteria and evidence that students used in making ratings for Item 3. Students were quite consistent in their interpretation of this item. Most students focused on the interaction between the students and the teachers outside the classroom. The majority of students tended to give higher ratings for teachers who were available, willing, and able to help them when they had difficulties. They gave lower ratings for those who they perceived were reluctant or unable to help:

I would consider my interaction with the lecturer outside the classroom – whether she is willing and happy to help me solve my problems when I approach her outside class. Some lecturers may turn us away and tell us to first ask other classmates or think about the problem ourselves. Some may answer our questions in an off-handed or uninterested manner. The ‘good’ teachers will sit down with us and explain patiently until we understand. (K15 BUSS)

The ‘poor’ teacher seemed reluctant to meet us and answer our questions. He would use a lot of excuses for not meeting us, or for keeping the meeting short. In the discussion, he could not provide us with any useful feedback or guidance. (K08 SD)

However, students had different views of what constitutes useful or appropriate help. Some students focused on the ability of their teachers to give them useful hints and suggestions, or to guide them to find the answer and solve their own problems. Others expected their teachers to give them clear and direct answers immediately, and would give the teachers a lower rating if their expectations were not met. One student commented:

I would consider if the lecturer was able to answer our questions clearly... For the good lecturer, his advice is very helpful – he gives us guides and hints on *how to find the answer to our problems*. (K14 BUSS)

Table 11: Criteria/evidence that students used in making ratings for Item 3

Item 3: The staff member provided appropriate help for students with learning difficulties	
<i>Reasons for high ratings</i>	<i>Reasons for low ratings</i>
<ul style="list-style-type: none"> • Teacher willing and enthusiastic, making positive efforts to help students even when they are busy (29) • Teacher able to explain clearly / provide useful hints or suggestions for students to deal with their problems (19) • Teacher available outside class, easy to contact staff for help (16) • Teacher answering students' questions outside class when asked (13) • Teacher actively identifying students with difficulties and providing them with extra help (7) • Teacher responding to students' questions promptly and seriously (3) • Teaching being patient and helpful when answering students' questions (3) • Teacher providing opportunities for students to ask questions in class (3) • Teacher encouraging students to approach them for help when needed (2) • Teacher showing care and empathy for students' with problems (2) 	<ul style="list-style-type: none"> • Teacher not able to answer students' questions clearly, unable to give useful guides or suggestions to students to solve their problems (18) • Teacher not willing or interested to help students, giving half-hearted help (14) • Students seldom / never approached the teacher for help (13) • Teacher turning students away / not willing to answer students' questions even when asked (7) • Student feeling that the teacher might not be willing or able to help (5) • Teacher leaving it entirely to the students to take the initiative to ask (5) • Teacher scolding students for being inattentive, lazy, or stupid when they have difficulties (4) • Teacher not caring / knowing if students have difficulties (4) • Teacher being less helpful (2)

Notes: 1. The number in brackets represents the number of times the reason appeared in students' responses

2. Only items with number of mentions of two or above are included

Another student had a completely opposite view of what appropriate help meant:

The lecturer didn't answer my questions directly. I had to search for the answer on my own. This may be good to me, but personally I hope that he can answer my questions directly and save my time in finding the answers myself. (L18 BUSS)

Students also differed in their views concerning the role of the teachers vis-à-vis the students in providing help. Some opined that at the university level, students should have the responsibility for seeking help when they need it. The duties of the teachers were simply to make themselves available, and to try their best to answer students' questions when approached. Others felt that only teachers who made the extra effort to identify students with difficulties and provide extra help to them individually deserved a high rating:

Most teachers will certainly help when asked. It depends on whether the students will take the initiative to ask or not... I think that it is the students' responsibility to take the initiative to approach the teacher for assistance. If you ask more, the teacher will tell you more. As long as they are free, they'll help you. (L11 BUSS)

I don't feel that any of our teachers will actively try to identify which students have learning problems. (L07 BRE)

While most students agreed that all teachers would answer their questions when asked, they also confessed that they would not approach certain teachers for help even when they had difficulties, either because they had found their help not useful in the past, or because they felt that the teachers might not be willing or able to help them. Some students said that they hesitated to ask for help because some of the teachers would scold them for being lazy or inattentive when they asked them questions:

The lecturer just treats teaching as a task to be completed. I didn't feel that he would be able to help me... I will not approach him for assistance. (K20 ITC)

When we approached the teacher when we had questions, he would often scold us for not paying attention in class. I felt miserable about the experience. I jotted down my questions when revising the notes, and I was not asking for tips for the exam. I felt bad for being criticised for not listening in class. I believe that even if a student is attentive, there must be things that she may not fully understand. And when a student asks for help and is treated that way, the lecturer is definitely not providing appropriate help. (K22 ITC)

Some students commented that they found it difficult to decide on the ratings for this item because they had never approached the teacher for help before, and they did not know the experiences of other students who had tried to seek help from the teacher. Because of the lack of relevant experiences or knowledge, some students maintained that they would normally give a neutral rating for the teachers. Others admitted that as they were uncertain, they tended to make their ratings on the basis of their overall impressions of the teacher:

I think that I shouldn't be asked this question. How can I know which students have learning problems? If I've consulted the teacher before, I'll decide on the rating by how he responded to my questions. If I haven't had that experience before, I'll just give a neutral rating so as not to be too harsh on the teacher. (L02 SD)

I find it quite difficult to answer this questions because it depends on the students' own initiative to approach the lecturer. To be frank, my response to this question is sometimes quite arbitrary. I give a rating of '3' to the second lecturer because I know that I will not approach him for help. I give a '4' to the first one because I feel that he may be able to help me if I approach him with a problem. It is just a feeling. (K36 RS)

A few students interpreted the item in a way similar to Item 2 above. They concentrated on the help provided by the teacher within rather than outside the classroom. They based their ratings on whether the teacher provided opportunities for them to ask questions, and whether the teacher was sensitive to their problems and gave further explanations and extra help when they had difficulties in understanding.

Item 4: The staff member presented the subject materials clearly.

The criteria and evidence that students used for making ratings for this item are summarised in Table 12 below.

Table 12: Criteria/evidence that students used in making ratings for Item 4

Item 4: The staff member presented the subject materials clearly	
<i>Reasons for high ratings</i>	<i>Reasons for low ratings</i>
<ul style="list-style-type: none"> • Students' subjective feeling of being able to understand the subject matter / follow the teaching (24) • Same criteria as Item 1 (19) • Teacher able to explain / express ideas clearly (17) • Systematic and organised presentation (14) • Teacher giving useful, real-life examples (11) • Teacher <i>highlighting and focusing on</i> main points (10) • Teacher providing useful / well-organised handouts (7) • Teacher well-prepared (7) • Teacher using a variety of aids and resources to aid understanding (6) • Students' feeling of having learned something new / useful (4) • Teacher using Chinese to aid explanation when needed (4) • Teacher linking knowledge (2) • Teacher presenting materials at a level appropriate for students (2) • Lively presentation (2) • Useful content (2) 	<ul style="list-style-type: none"> • Students' subjective feeling of not being able to understand / follow the teaching (13) • Teacher just reading out from textbooks / notes (11) • Poorly-organised / confusing presentation (6) • Students having difficulties in understanding because of the strong accent of the expatriate staff (6) • Content too superficial / not useful (5) • Poor, disorganised handouts (4) • Teacher not able to express ideas clearly (3) • Boring presentation (3) • Materials not relating to real-life situations (3) • Teacher insensitive to students' problems or difficulties (3) • Poor irrelevant examples (2) • Teacher with poor verbal skills (2) • Unclear expectations and/or assessment criteria (2)

Notes: 1. The number in brackets represents the number of times the reason appeared in students' responses

2. Only items with number of mentions of two or above are included

Items 1 and 4 seem to be providing overlapping if not redundant information on teaching, at least from the students' perspectives. Many students admitted that they used criteria that were similar to those for Item 1. A comparison of the students' responses to the two items confirms that the criteria and evidence that student used in deciding on the ratings for their teachers for both items are highly similar:

I think that this item is very similar to Item 1. I would consider if the lecturer is able to explain the main points clearly enough for me to understand and learn. Some lecturers just read from the textbooks, but reading is not the same as explaining to me what I need to understand. Good lecturers usually give a lot of useful examples, and give us tasks to do to check our understanding... (K24 ENGL)

The major identifiable difference between the two is that a lot of students included in Item 1 the attitudes of the teacher towards teaching and the students, but very few students included this in Item 4. The result suggests that students thought that while teachers' attitudes made a significant contribution to their learning the subject, this were conceptually unrelated to the clarity of their explanations.

Students also mentioned language as a factor affecting the clarity of explanation of teachers. Some students found that if a teacher used Chinese⁸ (or more accurately, Cantonese) to supplement the explanation in English for the more difficult subject matter or concepts, they could understand much better. At the same time, some students reported difficulties in comprehending the accented English spoken by some expatriate staff whose native language is not English:

I find that using a mixed medium of instruction (i.e., using both English and Cantonese) is much more helpful to our understanding... When the teacher is a foreigner, we tend not to ask questions. Moreover, some of the teachers speak accented English, which is very difficult for us to comprehend. We can understand better when the teacher used English in his teaching but

8. As mentioned in Section 3.1, the majority of the population in Hong Kong are Chinese, speaking the Cantonese dialect. However, English is still the official language of instruction in most Hong Kong universities, even after the handover in 1997.

supplemented it by further explanation in Cantonese, or use Cantonese alone. (L29 RS)

Other students insisted, however, that their problem is not about learning in English. They argued that they could understand well as long as the teacher speaks good English, and if their teaching was systematic and well-organised. The main issue, to them, is whether the teaching was systematic and clear:

We have no problem listening to the expatriate teachers because their spoken English is better. We have no problem in coping with a class conducted in English as long as the teacher speaks English well. It depends on whether the teacher has the ability to express his ideas clearly. (K03 BRE)

Item 5: The staff member's teaching stimulated my interest in the subject.

Table 13 summarises the criteria that students used for arriving at their ratings for Item 5. It is clear from the results that although the item focuses on whether the teaching stimulated students' interest, students' ratings on the item were heavily influenced by the understandableness of the teaching to the students. The majority of the students maintained that if the teacher was able to help them understand the subject matter clearly, they would be more interested in the subject because they felt that they could learn the subject well and develop a sense of achievement.

On the other hand, they claimed that if they could not understand what was being taught in class, they would feel frustrated about their learning and eventually lose interest in the subject. In other words, understanding is a pre-requisite for their developing an interest in the subject matter.

Table 13: Criteria/evidence that students used in making ratings for Item 5

Item 5: The staff member's teaching stimulated my interest in the subject	
<i>Reasons for high ratings</i>	<i>Reasons for low ratings</i>
<ul style="list-style-type: none"> • Good teaching that enables clear understanding of the subject matter (29) • Students' own prior interest in the subject (21) • Students' feeling that they had learned something 'new' or useful (15) • Teaching that relates to real-life situations or practice in real work settings (14) • Chance for active involvement or participation in the learning process (10) • Students' subjective feeling towards the teacher / teaching (9) • Teachers' use of humour / jokes (8) • Lively, interesting presentation (5) • Giving examples (3) • Friendly and good relationship (3) • Students' feeling of being able to learn the subject well (3) • Using audio-visual materials (2) • Materials that relate to students' interest (2) • Teacher sharing his her personal experience (2) • Teachers' enthusiasm and positive attitude towards teaching (2) • Teacher helpful to learning (2) 	<ul style="list-style-type: none"> • Poor teaching / failing to help students understand the subject matter (34) • Teacher just reading from notes or books / focusing on abstract theories / unable to relate to real-life practice or application (14) • Learning not useful / too superficial (6) • Students' lack of prior interest in the subject (6) • Boring, monotonous presentation (5) • Students' subjective feeling of not interested in attending the class (3) • Teacher not caring for students' learning (2) • Teacher not serious / enthusiastic towards teaching (2)

Notes: 1. The number in brackets represents the number of times the reason appeared in students' responses

2. Only items with number of mentions of two or above are included

One student commented on how the understandableness of the teaching affected his interest in the subject:

The 'poor' teacher could not help me understand the subject materials, and therefore could not stimulate my interest in the subject. I was quite interested in the subject before studying the course, but he simply could not make me

understand. When I asked him questions, he always replied that they could not be explained. That is what I don't like about his method of teaching. As there were more and more things that I didn't understand, I lost interest. Even when I had time, I would rather spend it on studying other subjects that I could make sense of. (K19 ITC)

Another important factor that affects students' ratings on the item is the extent to which the teacher is able to relate the concepts and theories to real-life situations, actual practice in work settings, or their applications in solving problems.⁹ Students felt empowered with the knowledge gained when they could see its relevance, and would be motivated to learn more about the subject. On the other hand, students gave low ratings to teachers who just talked in the abstract without linking the materials to real-life examples or applications. They would not be interested in the subject because they viewed the knowledge as superficial and not useful to them:

If I can see how the subject matter can be applied to real-life situations, I may find it more interesting. Thus, if the lecturer can relate the subject matter to real-life examples and practice in the work setting, I will become more interested in it. (K20 ITC)

However, as some students have pointed out, some subjects are by nature more 'applied' and related to their careers than others. Thus, the ability of the teachers to motivate them is limited to some extent by the nature of the subject matter. In those cases, the teacher should not be held accountable:

I would give a higher rating if the lecturer presents something that is new or up-to-date, or something that relates to actual practice. Again this is limited by the nature of the subject. Some subjects are more theoretical in nature and thus, more difficult to relate to practice. (K17 BUSS)

Students' ratings on the item were also moderated by their initial interest in the subject prior to attending the course. Some students maintained that they had

9. The emphasis of the students on applicability and relatedness of the teaching to real-life situations may, to some extent, be a result of the nature of the programmes offered in the University. The University, being formerly a polytechnic, has the tradition of offering mostly programmes leading to professional qualifications in various applied disciplines.

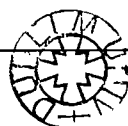
differential interests in the subjects even before starting the course. They argued that the teachers had very little or no influence on their interests. Some students would therefore give middle ratings to most subjects. Some others made ratings according to their own level of interest in the subject rather than the ability of the teacher to stimulate their interest:

I think that the lecturer has little influence over my interest in the subject. I would consider whether the subject is of any interest to me rather than whether the teaching of the lecturer has stimulated my interests in the subject. Thus, I tend to give a 'neutral' rating for this item. (K39 BUSS)

I think that interest is personal. No matter how well a teacher teaches, I still won't develop interests in a certain subject if I am not interested in it. Good teaching may help me get a better grade, but it has no effect on my interest in the subject... I give a higher rating for subjects that I personally like. I am very biased in this respect. That is my perception. I am sympathetic with teachers who have to teach the boring subjects. They need to stimulate the interests of an unmotivated class of students. It is very difficult. (L11 BUSS)

Other students, while admitting their own interest was important, argued that the teachers and their teaching could, and did, make a difference. They claimed that good teaching helped to stimulate their interests in subjects that they did not have any prior interest in, and poor teaching definitely killed their interests. Some even asserted that they would feel particularly bitter if a subject that they were originally interested in was poorly taught. They might give the teacher an even lower rating on the item for this reason alone. For example:

Our interest is not only affected by our own interest in the subject matter, but also influenced by whether the lecturer is able to demonstrate to us that the subject is useful and not boring. Sometimes, even for a boring subject, a good lecturer can make it less boring for us. This requires some skills on the part of the lecturer... They may think of examples to show us that the material is useful, organise some talks, or give us some reference to read... I think that if the lecturer is well-prepared and has a good attitude, he will be able to make us interested. On the other hand, even if I am interested in the subject, if the lecturer simply reads from the textbook, and cannot provide me with what I want to learn, I may lose interest very quickly. (K24 ENGL)



If I am interested in the subject initially but the teaching is very poor, I will feel extra bitter towards the teacher. My interest will decrease. (K15 BUSS)

Students had divided views on the role of humour, audio-visual aids, or multi-media presentations in stimulating their interest in subjects. While some students found these stimulating and useful in arousing their interest and gave high ratings for teachers who made use of the techniques, others argued that they were not essential and the effects were often short-lived and slight. They contended that such practices had little influence on how they rated their teachers on the item:

The use of humour or audio-visual materials did make some differences in students' interest. But it is highly individual... If the lecturer showed us some videos or talked about some interesting things in class, I might feel interested at that moment. But if I could not understand the materials, I might wonder what the purposes of the videos were, and feel that it was a waste of time... (K22 ITC)

It is not important if the lecturer tells any jokes or uses any computerised multimedia presentation or not. The most important thing is the clarity of the presentation... (K36 RS)

Some students resorted to their subjective feeling of whether or not they were motivated to attend the class when making ratings for this item. They gave high ratings when they felt that they were motivated to attend the session, and gave low ratings when they felt reluctant and unmotivated to attend.

The importance of interest and motivation to student learning is also evident. Students alleged that if the teaching was not understandable to them, they would not be interested in learning more about the subject. They would, instead, try to memorise the content and study merely for the sake of passing the examination:

If the teaching is poor and we do not understand the subject materials, we may just want to pass the exam. We are not interested in the subject itself; we study for passing the exam, not for learning. (L25 ENGL)

Item 6: The staff member gave me regular feedback on my progress.

This is by far the most problematic item from the students' perspective. A large proportion of students admitted that they found it difficult to rate and differentiate the performance of their teachers on this item. This is also reflected in the relatively large proportion of students who gave a rating value of '3' for their teachers in the simulated rating exercise described in Section 6.1. The criteria and evidence that students used for deciding their ratings for this item are summarised in Table 14.

Table 14: Criteria/evidence that students used in making ratings for Item 6

Item 6: The staff member gave me regular feedback on my progress	
<i>Reasons for high ratings</i>	<i>Reasons for low ratings</i>
<ul style="list-style-type: none"> • Teacher giving individual comments on students' performance in tests / submitted work (22) • Teacher giving students regular tests or assignments to test understanding (20) • Clear teaching schedule and objectives for students (7) • Teacher asking students questions in class on what has been learned before to check understanding (6) • Teacher encouraging students to work according to schedule (4) • Teacher showing interest in students' progress (3) 	<ul style="list-style-type: none"> • Teacher giving just a mark or grade on assignments or tests without any comments or feedback (9) • Few assignments or tests given (7) • Students' feeling of making no progress in learning because of not understanding the subject matter (5) • Students not knowing own progress (4) • Unclear about the teaching schedule (4) • Feedback too vague / not useful (3) • Long time lag in returning the marked assignments (2)

Notes: 1. The number in brackets represents the number of times the reason appeared in students' responses

2. Only items with number of mentions of two or above are included

A lot of students expressed that they had difficulty in deciding on the ratings for this item. Some students commented that the item is ambiguous, and they did not know what it actually asked about. They therefore tended to give a middle rating on this item:

I don't know how to rate. It is very difficult to tell if the lecturer gives me feedback on my progress or not. This is the most difficult one, so I choose 'neutral'. (L26 RS)

Others claimed that the only way for them to obtain regular feedback on their progress was through regular tests, assignments and in-class questions, and the comments made by their teachers on their individual performances in those tasks. However, many of them admitted at the same time that few of their teachers could afford to do that because of the various constraints such as class size. They found that most teachers were quite alike in this aspect of their teaching and so, it was not easy for them to differentiate among their teachers and make ratings accordingly. In other words, the item is not very useful in discriminating between the performance of the 'good' and the 'poor' lecturers, as explained succinctly by one student:

I think that this question is difficult, because under the existing arrangement there are so many students in a class that it is virtually impossible for the lecturer to pay attention to the progress of the individual students. The only way that we can know our own progress is through tests, presentations, and in-class questions, but I find that there aren't any big differences between a good lecturer and a not-so-good one in this respect. Even a good lecturer cannot do much in this aspect. I think that there are a lot of constraints here. (K24 ENGL)

Yet, among this group of students, some still gave a consistently lower rating for the 'poorer' teacher, which suggests that their ratings might be influenced by their overall impression of the teacher. Thus, the ratings they give may not be true representations of the performance of the teacher with regard to the aspect being evaluated:

I feel that this item is very ambiguous. I think that it is difficult for the lecturers to make me understand my progress. But very often, when I looked back at the

ratings I gave for the lecturer on the other items and realised that I had rated the lecturer as a good one, I would tend to give a higher rating for this item as well. (K17 BUSS)

Given the ambiguity of the item, some students focused more on their own progress in learning the subject rather than the feedback given by the teacher in their interpretation of this item. They gave a low rating when they felt that they were not making any progress in their learning, often as a result of their inability to understand the subject matter:

As for the subject taught by the ‘poor’ lecturer, I don’t know what I’ve learned from the lessons. I find myself making no progress at all in learning, so I choose ‘disagree’... (L15 BUSS)

A few students put forward criteria or evidence not directly relevant to the item. The wide variety of personal interpretations ranged from whether they felt motivated to revise the subject materials at home, whether the teacher gave them the ‘model answers’ to the tests, to whether the teacher reminded them of the coming of the examination and urged them to prepare well for it. The following comments clearly show this:

The other thing that I consider is my own interest in the subject. If I am interested in it, I will revise the materials regularly and know my own progress... The ‘good’ lecturer was able to stimulate me to revise his notes after each class. (K14 BUSS)

I give the lecturer a low rating because he refused to tell us the answers after the test. He said that those were open-ended questions, and we might answer in different ways. But I am not happy with that because without the answers, I don’t know what approach I should take when tackling the questions. (L21 ITC)

The ‘good’ teacher always reminds us that the exam is coming and urges us to study harder. He may also tell us that the progress of our class is behind that of the other classes, and will speed up a bit in his teaching. (K11 BRE)

It should be noted, however, that the students were asked to explain the criteria and evidences they used *after* they had made the ratings for the individual items. It is

possible that their ratings were influenced by their overall impressions of the teachers, and what they suggested above were merely anecdotal justifications for the ratings they made, rather than the actual criteria they used to derive the ratings. More studies are needed before we can fully understand how students make their ratings for their teachers.

CHAPTER 8 MEANINGS UNDERLYING THE NUMERICAL VALUES OF STUDENT RATINGS

This chapter explores further how students make their ratings on the various items on the student feedback form. The first section examines the meanings that students attached to the different rating values. The second section investigates the standards that students applied in making their ratings. The third section looks into the relationship between students' conceptions of 'good teaching' and their ratings of university teachers. The fourth and fifth sections examine the inter-relationships between students' ratings on different items, and the issue of subjectivity of student ratings.

8.1 Students' interpretations of the rating values

One important question in understanding student ratings of university teachers is: What are the underlying meanings of the rating values that the students give? To be able to properly interpret the ratings, we need to know what the students really mean by giving a value of '5', '4', '3', '2' or '1' for a specific item on the rating form.

The original labels of the rating scale that the students used in the rating simulation are identical to those of the University's student feedback exercise. Students were asked to indicate their views on each item on a five-point Likert scale, ranging from '5' (strongly agree) to '1' (strongly disagree). However, analysis of students' responses showed that most students interpreted the rating scale in an *evaluative* rather than a descriptive sense, and used the various points on the scale to categorise the merits of their teachers. They attached different qualitative judgements about the teacher such as 'exceptionally good', 'good' or 'bad' to the different ratings given, as illustrated by the following quotation:

If the lecturer is on the whole okay, I would give a '4'. If she is really good, I would give a '5'. If I have some reservations about the performance of the lecturer and feel that it is about average, I would give a '3'. For those whose teaching is really bad, I would give a '2' or '1'. (K36 RS2)

Very often, the ratings that students gave for a particular lecturer on a specific item are not based solely on their observation of the presence or otherwise of certain teaching behaviours demonstrated by the teacher. Instead, they tend to take also into account their subjective evaluation of the effectiveness of the teacher in achieving what they thought is desirable regarding the particular aspect of teaching. In many cases, students gave different ratings for teachers, not because of the differences in what the teachers did, but because of the differences in the perceived quality of the effort and the resulting effects on their learning. A student explained why she gave different ratings ('5' versus '2') to two different teachers on encouraging active participation in class:

I felt that most lecturers now have the habit of getting students involved by asking questions and giving chances for students to express their ideas in class... The first lecturer stimulated us to think, encouraged us, and treated us equally... The second lecturer also asked us questions, but the questions could not stimulate us to think. (K37 ENG)

Another student quoted an example in which the teacher tried to encourage active participation in class by giving marks to students who responded to her questions. Yet, the student gave a low rating on the item because

... she was treating us like small kids. Those students who were eager for getting more marks would rush to answer his questions. This was not good. Students were doing it for marks, not for learning. If no marks were given, there would not be any responses. (L16 ITC)

It is also clear that different students may have different interpretations of the rating values. As explained in Chapter 6, some students used the rating values to represent how well the teacher's performances have met their own expectations and

standards. Others made use of the ratings to indicate the relative merits of the teacher as compared to other teachers. The following three comments help to illustrate this:

I would not give my teacher a '5' even if I thought that his teaching was good. I would only give a '4'. I would give a '5' only when the teacher was exceptionally good. And if I found his teaching not good, I would give a '3'.... I seldom gave '1' or '2'. To me, '3' is already indicating that the teaching is not very good. (K21 ITC)

To me, '5' means 'perfect or extremely good'. '4' means 'good but there is still room for improvement'. '3' means 'nothing special, just average'. '2' means 'not good', and '1' means 'extremely poor'.

If the teacher is better than the others, I would give him a '4' or a '5', and if he is not as good, I would give a rating of '1' or '2'. (K14 BUSS)

The quotations above show that the same rating value may mean different things to different students. For some students, '4' and '5' represent 'better than average' and '1' and '2' represent 'not as good as the other teachers'. For others, '5' means 'exceptionally good', '4' means 'good' while '3' means 'not good'. Still others equated '5' to 'perfectly good', '3' to 'average', and '1' to 'extremely poor'.

Some students admitted that while they could broadly differentiate between 'good', 'average' and 'poor' teaching, they had great difficulties in making the finer distinctions between scores of '4' and '5', and between '2' and '1':

I find it quite difficult to differentiate between '1' and '2', and between '4' and '5'. I can only think in terms of broad categories of 'good', 'average' and 'not good'. (K26 ENGL)

The meanings that students attached to the mid-point score '3' are even more varied. In the interviews, students revealed widely different interpretations of the middle score in their actual rating practice, ranging from 'average' to 'quite poor'. Students offered that they gave a rating of '3' when:

- they thought that the teaching was just 'average',

- they had ‘no strong feeling about the teaching one way or the other’,
- they could not decide what rating to give either because of the ambiguity of the item or their lack of relevant experience to judge,
- they felt that it was difficult for the lecturer to achieve what was expected of him/her under the contextual constraints,
- they found the teaching ‘poor but tolerable’, or
- they found the teaching really poor but it is the lowest score they are willing to give.

Students commented that they could not fully express what they felt about the teaching by the numbers. Many preferred a more direct dialogue, and suggested alternative ways for collecting feedback such as interviewing or giving students space below each question for them to express more fully what they think.

From the discussion above, it seems reasonable to conclude that while a high rating broadly indicates positive reactions of the students to the teaching and a low rating broadly indicates negative ones, the exact meanings of the rating values are *personal* to the students who give them. Given that different students may have different interpretations of the rating values, averaging the ratings across students within a class may further obscure the true meanings underlying the numerical ratings, making the interpretations of the ratings even more difficult.

8.2 Standards for making ratings

As shown in Chapter 7, students are generally able to identify relevant criteria and evidences for making their ratings. Yet, some students found difficulties in operationalising them to derive a clear standard for giving their ratings on the items. Given the diversity of evidences relevant to a single item, students faced the problem of

how to combine the different evidences or observations to come up with a single rating that best represents their views. One student commented:

The main problem is that for each item, we can think of many relevant points and interpretations, and it is really difficult to arrive at an 'average' rating for all the points. So, we may focus only on one or two things that we feel are the most important, and neglect the others... (K38 BUSS)

As students were usually given very little time to fill out the rating forms, they needed to find short-cuts to deal with the task at hand. Some tended to rely on their global impressions of the teacher, and made slight adjustments to their ratings according to the item. Others simply resorted to the few things that first came to their minds, and made their ratings accordingly. In both cases, students' spontaneous 'gut reactions' may have influenced their ratings.

Students also revealed that they applied different standards for making ratings, some of which are rather naïve or unreasonable from the standpoint of the lecturers being evaluated. One student commented that although the teaching of a lecturer was good, she did not give a '5' for the item on understanding of the subject matter because '5' means "really, really good and I can understand *every* detail". Another student said that for the item on individual help, none of the teachers deserved a '5' because they might not be *available at the time* when he approached them for assistance. Yet another student justified her not giving a '5' for a teacher on encouraging active participation by the observation that "some of the students did not turn up for the tutorials because they hadn't finished their work."

The lack of a commonly agreed standard is a problem not only for the students, but also for the teachers and administrators who need to interpret and use the information for making curricular or personnel decisions. Even if students share common criteria in evaluating teaching, that they may apply different standards in

making ratings implies that the numerical values are far from being a precise measure. Class-average ratings, thus, must be interpreted with extreme sensitivity and care.

8.3 Student ratings and conceptions of 'good teaching'

Essentially, student ratings reflect the extent to which the teacher's behaviours have matched the standards as perceived by the students. Thus, students' conceptions of what constitutes good teaching will inevitably influence their perception of what a teacher should do in their teaching and hence, the ratings they give.

Students' view on the importance of providing notes in teaching is a case in point. For example, a student who thought that good teaching meant "giving students tidy and systematic notes and not wasting time to talk about things that are unrelated to the examination" remarked that "if a lecturer does not give us good notes while other lecturers do, then in comparison we would think that he is really 'bad' in teaching... we may think that the one who gives good notes is more serious in his teaching." When asked about why he gave a higher rating for a certain lecturer on the clarity of presentation, he justified it by the following reason:

It has something to do with the notes. If the lecturer gives good notes, we might find it easier to understand and revise. But if we were asked to take our own notes, we might not want to read them later. (K21 ITC)

On the other hand, students with a different conception of good teaching may have a different view on the importance of notes. A student from the same department who thought that an important part of teaching is to "stimulate students to think and apply the knowledge in practice" opined that the best way to teach is to "guide us to think through the subject materials, and to develop our thinking skills." When asked to rate two teachers on their clarity of presentation, he gave a rating of '2' to one who

gave them detailed notes and a rating of '5' to the other who provided them little notes, and made the following comment:

Giving notes did not guarantee that we could understand and not giving notes did not mean that we couldn't... It depends on whether the examples are good.
(K23 ITC)

Students' expectation for the kind of help they can receive from their teachers is another example. Students who believe that good teaching is about efficient transmission of knowledge tend to expect their teachers to provide clear and direct answers to their questions. For example, a student who described a good teacher as one who can present clearly and explain the subject materials to students in a step-by-step manner stated that she gave a lower rating for a lecturer "who teaches very well" on the item about individual help because

... the lecturer didn't answer my questions directly. I had to search for the answer myself. This may be good for me, but personally I hope that he can answer my questions directly and save my time in finding the answers myself.
(L18 BUSS)

On the other hand, students who believe that teaching is about helping students develop the skills to apply knowledge from different topics to analyse new situations stress the importance of guiding students to discover the answers themselves. A student holding this view gave a rating of '5' for one of his teachers not because the teacher gave him direct answers, but because the teacher was able to give him "useful guides and hints on how to find the answer to the problem."

The findings suggest that students' ratings are inseparable from their conceptions of good teaching. The same teaching behaviour may elicit different responses from students which in turn, lead to different ratings, dependent upon what they think are the appropriate things that the teacher should do to help them learn. It

follows that to interpret what the ratings really mean, it is necessary to understand the implicit theories of teaching held by the students giving the ratings.

8.4 Halo effects on student ratings

Another important question about student ratings is: To what extent is a student's rating on one aspect of teaching influenced by his/her evaluation of the teacher on another? In fact, one criticism of student evaluations is that student ratings are often biased in the sense that the warmer and more expressive teachers always get higher ratings, even for items unrelated to the observed teachers' characteristics.

There is evidence that students' ratings on the individual items are inter-related. This, however, does not necessarily mean that there has been a strong halo effect on student ratings. Instead, many students declared that although the items on the evaluation form purport to measure different aspects of a lecturer's teaching, the lecturer's performance in one aspect is inextricably linked to his/her performance in the others and thus, affects the respective ratings for those items. For example, the clarity of the lecturer's teaching affects not only students' understanding of the subject matter, but also their participation in class, their help-seeking behaviour, as well as their motivation for studying the subject:

If the method helped my understanding and made me interested in the subject, my participation in class would be automatically higher... [Item 2] ... If I am involved actively in learning a subject, I would have more questions to ask the teacher. Then I might feel that the teacher provided me with more help... [Item 3] ... I could understand better because the lecturer's presentation was clear... [Item 4] ... Even if you were interested in a subject initially, if the teaching could not make you understand, you would not develop interest in the subject... [Item 5] (K19 ITC)

In this case, students' ratings for the different items are inter-related not because of the halo effect but rather, as a result of the fact that the various aspects of their

learning experiences are necessarily inter-related. While it is conceptually possible to distinguish between the different dimensions of teaching, in actual practice, it is difficult to treat each aspect in isolation and ignore how they interact with each other to shape the *total* learning experiences of the students.

It is undeniable that students' overall impressions of teachers may influence, to some extent, their ratings for individual items. This is particularly true for items which are ambiguous to the students or which they may not have direct relevant experiences to judge. But even in those cases, students argued that their ratings were based on their holistic judgement about the entirety of the teacher's teaching rather than purely their personal liking for, or relationship with the teacher. Students were able to differentiate between a 'nice person' and a 'good teacher', as illustrated by the two comments below:

I give ratings according to the overall performance of the teacher – e.g., whether the content of his teaching is substantial, not abstract; whether he is able to give a lot of real-life examples to illustrate the application of the materials, etc. Briefly, whether he uses different methods to make us understand... I feel that every teacher is nice. But being nice does not necessarily mean that they are eager to teach and want to teach the students a lot of things... (L14 BUSS)

The teacher is a really nice person... but his teaching is quite confusing sometimes. (K03 BRE)

It is also interesting to observe that while none of the item asks about the teachers' attitudes such as their enthusiasm towards the subjects and teaching, or concerns for student learning, these attributes have come up in many of the students' responses on a number of items as one criterion they considered in making ratings. For example, for Item 4 on whether the teacher presented the subject materials clearly, a student gave a high rating for a teacher with the following justification:

In our discipline, the methods of teaching adopted by teachers are almost the same. There are not any big differences. The only difference is that the 'good'

lecturer is conscientious and enthusiastic in her teaching, while the other lecturer is not. (K04 BRE)

This observation suggests that teachers' affective characteristics have strong influences on students' perceptions of the effectiveness of the teaching which, in turn, affect student ratings of their teachers. It also implies that students do not necessarily interpret the individual items strictly in terms of the wordings of the items but rather, may make use of evidence that has *personal* meanings to them. Given that teaching is an inter-personal activity, it is highly likely that teachers' attitudes towards teaching and their empathy for students will influence how they teach as well as how their students learn. It would therefore be a mistake to label them as biasing factors and dismiss the possibility that teachers' affective characteristics may have real impacts on students' learning.

8.5 Subjectivity of student ratings

To the extent that different students may attach different meanings to the same evaluation item on the basis of their own perceptions of what the teacher should do in their teaching, student ratings must be subjective in nature. Many students recognised that their ratings might sometimes be affected by situational factors such as their moods or the kind of feedback they have obtained from the teacher on their work, but almost all claimed that they have tried to be as fair and objective as possible in giving their ratings.

It is also apparent that the degree of subjectivity of students' ratings depends on the focus and/or wordings of the specific items. If the meaning of the item was not clear to the students, or if it asked about an aspect of teaching that the students did not have any direct relevant experience of, students may have problems in determining the appropriate rating. Faced with this problem, students might adopt different strategies to

derive their ratings. Some students tended to choose the neutral rating '3' when they had doubts about the interpretation of the item, or when they did not have the experience to judge:

I don't know what is being asked here for this item. So, I give a neutral rating. (L02 SD)

It is difficult to say. I myself have never approached a lecturer for help before because... I tended to solve my problem myself. So, I usually give a neutral rating. (K17 BUSS)

Other students based their decisions on their overall subjective impressions of the teaching of the lecturer concerned. For example, to Item 3 on whether the teacher provided appropriate help to students with learning problems, a student replied:

It is difficult to decide. I won't ask the lecturer even if I have problems in learning a subject. So, I don't really know if the teacher will provide me with appropriate help or not if I approach him. I rely mainly on my overall impression. (L26 RS)

Another student explained how she made her rating on Item 1:

I think that it is quite difficult to decide on the ratings for this item because it is very vague. I don't really know what it asks. So I consider mainly the overall performance of the lecturer – whether she is serious and conscientious about teaching, and whether she really wants to help us learn the subject material well. (K25 ENGL)

In both cases, it is hard to dispute that their ratings were subjective in nature. However, this does not mean that their ratings are totally erratic, either. As some of the students have argued, their subjective feelings were often based on their direct experiences of the teaching within and outside the classroom over a substantive period of time. Furthermore, as most of them shared the same feeling about their teachers and gave similarly high or low ratings on them, there must be some degree of objectivity in their *collective* observations. At least, the rating cannot be simply attributable to the personal prejudices of the individual students.

CHAPTER 9 SUMMARY AND DISCUSSION

This study set out to examine how university students go about making ratings for their teachers. It explored students' perceptions of and attitudes about the rating process, their strategies for making ratings, and the criteria, evidence, and standards students used to arrive at specific ratings. This chapter summarises the major findings, and discusses their implications for theory and practice. The last part of the chapter concludes the study, and suggests areas for further research.

It should be pointed out at the outset that as the study involved only 71 students from one university in Hong Kong, the findings must be viewed as tentative rather than definitive. The applicability of the findings to other institutions, especially those with different institutional contexts or cultures, should be viewed with great caution. Yet, the study has contributed significantly to the huge body of literature on student evaluations of university teaching in that it filled an important gap in the existing knowledge base by inquiring into students' mental processes in giving ratings for their teachers.

9.1 Students' perceptions of teaching evaluation

The study found that most students demanded to have the opportunity to participate in teaching evaluations, but perceived that their feedback had little impact on teaching improvement. The findings were generally in line with prior research (Jacobs, 1987; Marlin, 1987; Taylor & Ricketts, 1982). Moreover, this study revealed that students had different perceptions of the purpose of teacher evaluation, some of which are associated with students' uninspired approaches to giving ratings. Arguably, if students perceive that the evaluation is a ritual to satisfy external demand for accountability or a strategy for the university to enhance its status and competitiveness rather than a genuine attempt on the part of the university to improve teaching and

hence, student learning, they will find the exercise of little personal relevance and thus be more likely to treat it casually — there is no personal motivation for them to do it otherwise.

Moreover, students reported different personal reasons for participating in the rating process. The finding that some students took the evaluation as a channel for making consumer complaints or venting their anger or dissatisfaction is a matter of great concern. For student feedback to have any beneficial effect on teaching, it must be seen by the teachers who received it to be credible and reasonable (Coe, 1998). Harsh and emotionally-charged comments from students will only reinforce the scepticism of the teachers, and fuel the distrust and antagonism that may exist between the two parties, leading to a culture of blame rather than a culture of problem-solving which is needed for improving teaching.

There is no dearth of suggestions for using student evaluations. Most of them focus on the procedure, instrumentation, and appropriate use of the data (Braskamp & Ory, 1994; Ramsden & Dodds, 1989). Few, however, have emphasised the important role played by the students in the process, or the need for getting them committed and involved in the system. Results of the study suggest that students' understanding of the purpose of evaluations and their perceptions of the role they play in the evaluation process may have a strong influence on how they react to the evaluations. Unless students have a clear understanding of the purpose of, and their role in, the teacher evaluation system, it is unlikely that they would be willing to give their feedback in a thoughtful manner that is so important for either formative or summative evaluation purposes.

9.2 Attitudes of the students about the rating process

Previous studies showed that most students reported to have taken the evaluation seriously and tried their best to be fair and accurate in their ratings (Jacobs, 1987; Marlin, 1987). The present study, however, revealed that students had different personal meanings of 'seriousness' and reported quite different rating behaviours. Furthermore, the inconsistency and discrepancy between their self reports and their depictions of their classmates' behaviours suggested that their claims could not be accepted at face value. Future research should focus more on the thinking processes and actual rating behaviours of the students rather than relying exclusively on their own claims of seriousness in rating.

The study also found that students' attitudes about the rating process were heavily influenced by the context of the evaluations. Students tended to become less serious in their ratings when they were asked to complete too many forms, to answer questions that appeared to be ambiguous or irrelevant to their contexts, or to rate teachers of subjects that they considered of low personal importance to them. Furthermore, students who were suspicious of the purpose of the exercise were less inclined to do the evaluations seriously. Most importantly, the study showed that students' attitudes in the rating process were not fixed but rather, changed according to their experiences and perceived impact of their ratings. If students perceived that their feedback would be disregarded by their teachers and had no effect whatsoever in their teaching practices, they would start to question why they should make the effort to do the evaluations. This is probably why many evaluation systems started with good faith have degenerated into pure bureaucratic exercises which both students and teachers dread and to which they give little thought (McKeachie & Kaplan, 1996).

Paradoxically, the believability of the students' ratings is influenced by the believability of the teachers and administrators on the part of the students (Arreola, 1987). If students do not have any faith that their feedback will be taken seriously by the teachers or the administrators, they will not be serious in giving their ratings. Thus, building up a culture of mutual trust between students and lecturers is an important component of evaluating and improving teaching.

9.3 Students' general strategies for making ratings

The study showed that students employed a number of strategies to arrive at the ratings for their teachers. Their ratings were largely the results of deliberate choices rather than spurious, erratic decisions. There were, however, variations in the strategies that students used in arriving at their ratings in terms of the criteria (i.e., global versus item-specific), reference point (i.e., other teachers' performances versus students' own standards), focus (i.e., self versus whole-class experiences), and disposition to giving extreme scores (i.e., using versus avoiding extreme scores).

While the study has not shown why students adopted different strategies or how the strategies influenced the specific ratings they gave, the findings suggest that student ratings, by virtue of how they are derived, may mean different things to different students. For example, they can be a criterion-referenced measure of the extent to which the teachers' performances match students' definitions of good teaching, or a norm-referenced measure of the relative merits of the teachers from students' perspectives. Also, the ratings may involve different degrees of subjectivity on the part of the students, dependent upon the stance that students take when determining the ratings.

The implication is that student ratings cannot be interpreted in a straightforward manner as if they were objective and precise measures of the effectiveness of university teachers. To interpret and use student ratings in a meaningful way, it is crucial to understand how they have been made, and what they actually represent. Users of student evaluation data should be made aware of the contextual complexity underlying the numerical ratings, and take them into consideration in their interpretation and use of the data. There is also a clear need to coach the students in the way they should approach the task of making ratings, to ensure that the data they generate are what the evaluators intend to collect.

9.4 The criteria and standards that students use for making ratings

A number of major findings have emerged in analysing students' reports of how they made their ratings for the individual items. First, students were generally able to differentiate between 'good' and 'poor' teachers from their perspectives, and their ratings were broadly indicative of their views. Moreover, their ratings were largely based upon their holistic evaluation of how well their teachers had been able to help them learn the subject matter rather than the personal attributes of the teachers such as 'niceness'.

Secondly, students used a variety of criteria and evidence, and applied different standards when deciding on the specific rating values for the items. Furthermore, they had widely different interpretations of the rating values, particularly that of the middle rating '3'. The result is broadly consistent with prior research (for example, Benz & Blatt, 1996; Talley & Timmer, 1992). On the other hand, it is also clear that common themes could be identified in the criteria and evidence that students used for evaluating the items, suggesting the existence of shared implicit theories of teaching among

students (Harrison, Ryan, & Moore, 1996). However, the fact that students make their ratings on the basis of identifiable common criteria does not imply that the criteria must be valid or important. Prior research has shown that student ratings correlate positively but only moderately with measures of student learning (d'Apollonia & Abrami, 1996; Cohen, 1987). While most of the identified criteria appear to be relevant and sensible, their validity as a measure of teaching effectiveness still needs to be demonstrated.¹⁰

Thirdly, students' ratings were influenced to some extent by their individual conceptions of 'good teaching'. Depending upon their personal conceptions of 'good teaching', different students were found to have different views of what the teacher should do, and gave ratings with regard to different criteria or standards. Thus, the same activity can be portrayed positively or negatively by students, resulting from the position they take. In other words, student ratings may be composed of endorsements of quite different behaviours, or be based on quite different criteria of 'good teaching' as Entwistle & Tait (1994) and Kember & Wong (2000) have argued.

Fourthly, there was high inter-correlation among student ratings on different items, but there are reasons to suspect that not all of it was attributable to the 'halo effect'. Students suggested that part of the inter-correlation could be a logical result of the complex interactions between different aspects of a lecturer's teaching in

9. While many of the criteria mentioned by the students seem to be good practices, they are not unambiguous evidence of effective teaching because of the extreme complexity of the teaching process. For example, while calling on students by name to answer questions is seen by many students in the study as a good practice to encourage participation in class and thus is rewarded with high ratings, it is viewed by some as threatening and a form of harassment. Most educationists agree that there is no one way of teaching that is effective for all situations (Elton, 1984). Thus, the definition of effective teaching must depend on the goals, the subject nature and characteristics, the learner characteristics, and the contexts. Hoyt & Cashin (1977), for example, revealed that teaching behaviours associated with learning factual information were quite different from those aimed at promoting students' critical thinking and problem-solving skills. Furthermore, it is commonly agreed that good teaching exists in many forms. That some teachers who are regarded by students as highly effective do not necessarily adopt all of the 'good' practices suggests that the criteria are neither necessary nor sufficient for defining and measuring effective teaching.

determining their mutual effectiveness. This offers a plausible alternative explanation for studies which showed that a drastic improvement in one aspect of teaching would lead to higher ratings for other items (for example, Williams & Ceci, 1997). It is logically probable that a more expressive or enthusiastic teaching style will lead to increased motivation and interest of the students, which in turn will lead to increased understanding, more active participation and help-seeking behaviour, and more feedback from the teacher.

Fifthly, student ratings were inevitably subjective in nature, particularly when the items were ambiguous or when students did not have the direct relevant experience to evaluate the item. They were not independent of the contexts in which the teaching was carried out, either. Class size, appropriateness of the textbook and syllabuses to students' needs, the nature and characteristics of the subject, students' own interest in the subject, etc. were mentioned by students as factors influencing the ratings they give for a particular teacher. The observation supports the quantitative findings from prior studies that student ratings are a function of the course characteristics such as electivity and level of course, class size, and discipline areas, etc., which are outside the control of the teacher (Feldman, 1978; Kwan, 1999; Marsh, 1981).

To conclude, student ratings are clearly not a very precise, objective measure of the performance of university teachers. Individual students' interpretations of the items and the rating values, standards, conceptions of 'good teaching', etc. will influence their perceptions of what the teachers do and should do, and consequently, their evaluations. To compare the ratings received by individual teachers within a university without due consideration of the differences in contexts and the cognitive and affective characteristics of the students across subjects will lead to grossly unfair and inappropriate judgments about the teaching of university lecturers.

However, it is also important to recognise that student ratings are not entirely situational or spurious. They are often based on students' first-hand experiences of the teaching and learning process over a substantive period of time, according to largely meaningful criteria which are shared by most students. Given the high level of consistency in students' views concerning the performances of individual teachers, their collective views must reveal something about the teaching and learning process beyond their individual personal prejudices. It is, therefore, equally wrong to hastily dismiss their credibility as a useful source of information for making informed instructional decisions and judgments on the grounds that they are merely subjective perceptions.

There is therefore a clear need for university lecturers and administrators to have a better understanding of the nature of student ratings – how they are made and what they really represent, so that they can interpret and make use of the information in a more appropriate manner. As Wragg (1993) points out: “systems are neither intelligent nor foolproof but rely on people to use and interpret the results.” Attempts to improve the usefulness of student evaluations should focus more on the proper understanding, interpretation, and use of the data, rather than on working towards developing a ‘perfect’ mechanism or instrument.

Talley & Timmer (1992) argue that the lack of shared meanings of instructional quality and evaluation between students, university administrators, and lecturers may have wide-ranging implications, given the importance of student evaluations in retention, promotion and tenure decisions. Perhaps, positive actions should be taken to engage students and teachers in a constructive dialogue on the nature and purpose of university teaching, and the effective ways to help and inspire students to learn. This will help to establish a common ground for students to make their ratings, and

contribute to the development of shared meanings of teaching quality between students and lecturers. It reduces the ambiguity felt by students about the meanings of the various evaluation items and their anxieties in the rating process, making the results more credible and acceptable to teachers as a source of data for improving and evaluating teaching.

9.5 Implications for developing student evaluation systems

The implications of the findings are clear. If we are serious about getting feedback from students, we need to convince them that the evaluation is important and worth doing. If we want students to be responsible and serious in giving their ratings, we need to treat their feedback in an equally responsible and serious manner.

Scriven (1981) points out that designing a student evaluation system involves much more than producing the piece of paper for students to fill out. Preparing students as evaluators is an important part. Likewise, Boyer (1990: 40) argues that it is a mistake “to ask students to fill out a form at the end of a class without a serious consideration of the process”.

Unfortunately, most designers/administrators of student evaluation systems have neglected the important role played by the students in the process, and few have considered how to convince students of the value of the exercise and make them a committed partner in the process. Even fewer have clearly thought of how the information collected will be eventually used.

Including a statement of purpose on the feedback form alone will not motivate students to treat the process responsibly, as students do not always read and understand the instructions on the form as expected (Tally & Timmer, 1992). Serious efforts must be made to gain students’ co-operation and support of the systems, prepare them for the

evaluation task, educate students regarding their role in the evaluation system (Marincovich, 1998), and establish and maintain the joint credibility of students with teachers, and teachers with students (Arreola, 1987). Ballantyne (1999) showed that students who received responses from their teachers on the feedback they gave tended to view the evaluation as more important, felt more empowered, and found their lecturers more approachable and concerned about their teaching. But to achieve this, the university must have a consistent and appropriate practices for valuing and rewarding good teaching, not just consistent and appropriate rhetoric (Scriven, 1981). It is also important to recognise that student evaluation systems can become banal over time. Universities must be aware of the need to periodically “re-examine the assumptions behind a practice gone stale” (Marchese, 1997: 4).

9.6 Implications for using student ratings for formative evaluation of teaching

To justify their use for formative purposes, it needs to be demonstrated that student evaluations can provide useful diagnostic feedback on teaching and learning that enables teachers to identify their strengths and weaknesses in teaching. Results of the study indicate that student ratings reflect students’ evaluation of the impact of the teaching on their learning rather than a mere measure of the likeableness of the teacher, easier courses, or leniency in grading. While they are not a direct, precise measure of the teaching effectiveness of a teacher, they can provide valuable information for understanding how well the teaching has been generally received by the students.

The finding that student ratings for different items were intercorrelated, however, may cause difficulties in interpretation when one attempts to identify the strengths and weaknesses of the teacher. There is evidence that students’ global impressions may influence their ratings for the individual items. It follows that any

item ratings that they make reflect not only the behaviour of the teacher but also the students' holistic impressions of the teaching, mediated by the subject nature, the context of the teaching, as well as their personal characteristics and value systems. Murray (1985) argues that low-inference items that focus on specific behaviours are less influenced by students' global impressions and thus, are more useful for teaching improvement purposes. Yet, findings of the present study show that even seemingly objective items are not totally immune from such effects. It implies that a meaningful and constructive interpretation of the feedback requires some understanding of the students, and the teaching and learning process. Given that most university teachers do not have any training in education, consultation is needed to make effective use of student feedback to improve teaching (Piccinin, Cristi & McCoy, 1999).

Students argued that the intercorrelations among item ratings are a logical consequence of the inexorable links between the different aspects of teaching in influencing their learning. For example, how well a teacher is able to encourage active participation and learning in class is determined by more than simply whether, or how often, the teacher asks questions or gives learning activities in class. The teacher's empathy for students and their learning, the ability to explain the subject matter clearly, success in stimulating interest and thinking, willingness and enthusiasm in helping students inside and outside the classroom, and the openness and supportiveness of the climate in class, etc., all contribute to the success or otherwise of the teaching. If we accept this argument, we must agree that teaching is not merely a series of disconnected activities, and that good teaching involves more than techniques (Palmer, 1999). The implication is that in any attempts to improve teaching, focusing too narrowly on the skills and behaviours related to specific aspects of teaching with low ratings without considering how they relate to the other aspects of teaching in effecting learning may

not be optimally effective. It is also important to recognise that teachers' personal qualities and characteristics, as much as their knowledge and techniques, are likely to determine their effectiveness in teaching (Banner, Jr. & Cannon, 1997), at least from students' perspectives.

Most training and professional development programmes for teachers use a model in which experts impart technical skills and knowledge to teachers in a context that is divorced from the actual classroom, usually not preparing them to engage in dialogue, reflection, and enquiry (Novice, 1996). To really improve teaching, teachers should be encouraged to re-examine the entirety of their teaching rather than to concentrate only on improving specific skills. The aim should be to enable teachers to become reflective practitioners (Schon, 1983) rather than to develop them into technicians of teaching. Ho (2000) demonstrates that a conceptual change approach to professional development is feasible and can be effective.

End-of-course numerical ratings such as those examined in this study have been criticised for coming too late and failing to reveal fully what students feel. Such sentiments are also evident in students' responses. Furthermore, while it is clear that students arrive at their ratings on the basis of personally meaningful criteria and standards, the rich meanings underlying the ratings are often lost when the ratings are collected, averaged and reported, unless special effort is made to solicit and document them. Very often, teachers know what ratings students give, but not why. They may be able to tell from the ratings if their teaching has reached the students educationally, but not what the problems are or how can they be alleviated if it has not.

The implication is that while the numerical ratings can be useful indicators for identifying what have or have not gone well in our teaching, they have obvious limitations in providing feedback for teaching improvements, and must be

complemented by evidence that gets beneath the numbers (Benz & Blatt, 1996) to enable us to understand why, and how the teaching can be improved. This does not mean that end-of-course ratings should be abandoned, but that we should not rely exclusively on this form of evaluation, allowing the simple numbers to drive out the richer forms of student feedback for the evaluation and improvement of teaching (Marchese, 1997). Angelo & Cross (1993) provide a wide range of classroom assessment techniques such as the “minute paper” which can be effectively used to obtain formative feedback for improving teaching.

9.7 Implications for using student ratings for summative evaluation of teaching

To be useful as an indicator of teaching quality for summative evaluation purposes, student ratings must be shown to be a fair, objective, valid, and unambiguous measure of the teaching effectiveness of the teachers being evaluated. To the extent that different students may approach the evaluations in different manners, employ different strategies to derive their ratings, have different interpretations of the items, and ascribe different meanings to the rating values, they are necessarily subjective and perceptual in nature. Moreover, they are invariably influenced by the cognitive, affective and motivational characteristics of the students, and the contexts of teaching. Thus, student ratings can never be precise, objective, and fair measures of the teaching performance of the teachers. While continuous efforts should be made to improve the evaluation instrumentation and procedures, it should be recognised that perceptual ratings by students can never reach the standard of measurement accuracy or validity of what McGaghie (1991) termed a “golden standard”. They are, at best, an imperfect and indirect measure of the performances of the teachers. To deny this and treat student ratings as if they were perfect measures of teaching effectiveness would only induce

hostility and resistance from teachers, and lead to “unfair and inconclusive distinctions among teachers” (Meech, 1976). This would defeat rather than support the ultimate purpose of collecting the feedback – to recognise, reward, and improve teaching.

The primary goal of teaching is to facilitate student learning, defined in the broadest sense. Ideally, the best measure of teaching effectiveness is the amount of additional learning that takes place as a result of the teaching. Unfortunately, valid measures of the value-added student learning are difficult to obtain because educational goals are multiple and contested (Cave, *et al.*, 1997), particularly across subjects and disciplines. Furthermore, a wide range of student and contextual factors apart from teaching affect learning, making it difficult if not impossible to separate unambiguously the unique contributions of teaching from other influences. The lack of explicit graduate standards and the absence of a common system for assessing student exit achievements make the task even more difficult, especially when university lecturers have the dual role of being both the coach and assessor of their students. As a result, most of the measures of student learning outcomes are too simplistic to be used as a direct measure of teaching quality (Johnes & Taylor, 1990). Much more research is needed before a direct measure of student achievement can be developed that is accepted by all as a valid indicator of teaching effectiveness.

In the absence of a perfect measure, decisions were, and still have to be made about the teaching quality of individual lecturers. The choice is: should the decisions be made on the basis of the personal opinions of the decision-makers without any support of data, or should it be based at least on some less-than perfect data? Fitz-Gibbon (1997: 314) argues that “the quality of data must be constantly challenged and we must seek for constant improvement in the quality of data, but to have no data at all is to leave people vulnerable...” Student ratings are definitely not perfect. They are

subjective and need to be interpreted properly and used cautiously. They are neither necessary nor sufficient for the evaluation of teaching merit, but they are one important source of data to enable informed judgements to be made (Scriven, 1994). The huge body of research on student evaluation of university teaching has shown that student ratings “tend to be statistically reliable, valid, and relatively free from bias or need for control; probably more so than any other data used for evaluation” (Cashin, 1995: 6). The findings of this study that students tend to evaluate teaching in terms of its impact on their gain in learning, and make ratings on the basis of largely sensible and meaningful criteria further attest to the credibility of the ratings as one source of useful (but imperfect) information for judging the quality of teaching.

What is important is for teachers and administrators to recognise both the value and limitations of student evaluation data, and interpret and use them appropriately. Student ratings of university teaching cannot be interpreted at face level; they must be understood in contexts, taking into account the course, the course objectives and the students’ characteristics. They must also be triangulated with other evidences of the quality of teaching of the lecturers. Moreover, the use of student evaluation data should not distract our attention from the urgent need for identifying and developing more direct measures of teaching effectiveness such as the learning gain of the students.

9.8 Appropriate interpretation and use of student ratings

Most researchers tend to agree that the fault of student ratings is not in the numbers, but in their interpretation and use (Kwan, 1999; McKeachie, 1997a, 1997b; Theall & Franklin, 1990). In many universities, student ratings are used as the only source of information for making summative judgements on teaching. There has also been concern over the abuse and indiscriminate use of the ratings for comparing

teachers across widely different contexts, and sometimes, solely for punitive purposes. These problems seriously undermine the usefulness of student ratings for evaluating teaching.

On the basis of the voluminous research, writers on teaching evaluations have provided lists of recommendations for the appropriate use of student ratings for evaluating teaching (for example, Braskamp & Ory, 1994; Cashin, 1990; Centra, 1993; Hoyt & Pallett, 1999). They suggest, among other things, the following important guidelines for the interpretation and use of the data:

- Multiple sources of information are needed for accurately evaluating and improving teaching.
- Student ratings data should be used as one, not the only, source of data about effective teaching, and they must be triangulated with other evidences of teaching quality.
- Student ratings should be viewed as information to be interpreted, rather than final verdicts. Students should be seen as providers of information, rather than judges of the teaching effectiveness of teachers.
- Student ratings must be interpreted in context, i.e., in relation to the nature of discipline, curriculum objectives, course features, and characteristics of the students in the instructional groups.
- Over-interpretation of student ratings should be avoided, and only crude judgements should be made on the basis of the ratings.
- To generalise from student ratings data to a lecturer's teaching performance, it is necessary to consider student ratings both across courses and across time.
- It is inappropriate to use student ratings to compare teachers across subjects or levels, or to rank individual teachers within a department or the university according to the raw ratings they receive.

In actual practice, however, it seems that the recommendations have not been generally heeded. Meech (1976) comments that teaching evaluation is a “stateless art”. Recently, Theall (2000) admits that “the state of evaluation practice is pretty sad.” It appears that despite the large amount of research carried out in relation to the evaluation of teaching and in particular, student evaluations, there has been very little improvement in the way that teacher evaluations are done in universities over the last twenty-five years. Weimer (1997: 415) comments that the policy and practice of evaluating instruction today is at least 20 years behind the current research, because most faculty and administrators “continue to be largely ignorant of this research.” There is evidence that many of those who need to use student evaluation data are unable to interpret them accurately (Franklin & Theall, 1989). There is clearly an urgent need to find effective ways to disseminate the research findings to university lecturers and administrators, enabling them to become more informed users of the evaluation data.

9.9 Conclusion

Teaching is a complex social activity. Any attempts to accurately evaluate it must take note of the multi-goal, multi-facet, and multi-stakeholder nature of the process. Students, as direct recipients of the teaching, have a legitimate role to play in evaluating the quality of teaching. Although student evaluations are not perfectly valid and foolproof, they can, if properly designed and implemented, provide useful first-hand information about students’ learning experiences that, *in conjunction with other evidences*, enable informed judgements to be made about the teaching.

It must be stressed that the usefulness of student feedback should not be taken for granted. A real danger of student feedback systems is that they may become banal

and routine. Furthermore, the information collected may be devalued or totally ignored by teachers or administrators because of their distrust or fear. On the other hand, it can be misinterpreted, abused, or misused. But as Fitz-Gibbon (1997: 316) argues

... misinterpretation of data is always a problem, whether by researchers or by teachers, or, particularly, politicians. But at least when there is data available the debate can be joined on the basis of data, and the error in the data can be readily demonstrated.

Braskamp & Ory (1994) suggest that a new perspective about teacher evaluation is needed. Instead of viewing it as an objective scientific endeavour with absolute truth as the goal, it may be more productive to see it as a form of argument. They contend that arguments “involve discussions, debates, discoveries, dialogue, and deliberations that can lead to decision” (p. 6). Student ratings, in this case, provide useful information on which such arguments can be initiated.

Suggestions for further research

The present study shows that students have differential attitudes toward the evaluation process, adopt different strategies, and make use of different criteria and standards in arriving at their ratings. However, there are still many questions that need to be answered before we can fully understand the mental processes of the students in the evaluation process. For example, what are the proportions of students with the different categories of views about teacher evaluation in universities and in specific classes, and how do they influence the ratings that they give for their teachers? To what extent are students' attitudes influenced by their personalities, goals, or perceptions of their roles in learning? What are the factors influencing students' choices of rating strategies, and how exactly do the strategies determine their ratings? To what extent do students' background variables (e.g., personalities, values, interests, and conceptions of learning) and the course features (e.g., discipline areas, level of study, required or

elective, etc.) influence the criteria and standards that students use in making ratings? To what extent do students' criteria and standards match those of their teachers?

Much more research is also needed for making the evaluations more useful. For example, how can we make students more responsible and serious in the evaluation process? What forms of evaluation will be more meaningful to the students, and motivate them to be more thoughtful in making ratings? How do teachers and administrators actually interpret and use the data for instructional or personnel decisions? How can we help teachers and administrators to interpret and use the data appropriately?

Institutions differ in their systems as well as in their organisational cultures. What has been found to work in one university may not be equally applicable to another. Furthermore, in any social institution, people change and so do their views and thinking. Universities need to examine their local context to see if their system is really working according to design, and to constantly monitor its implementation. In other words, the evaluation system itself must be subject to evaluation periodically.


McKeachie & Kaplan (1996: 9) plead:

Let's Do Better – we are fallible; we are not likely to achieve perfection; but we can do better, and we should. We have an ethical obligation to maximise the value of the time spent by students, faculty and personnel committees.

Given the large amount of resources in terms of class time and effort spent on collecting and analysing the student feedback data, it is a responsibility for those concerned to ensure that the system will yield high quality data, and that this data will be used optimally for improving students' learning. We owe it to the students and the teachers to try our best to achieve this.

Appendix A: The Student Feedback Questionnaire Form

Designed by Wenhoad Computer Consultants Ltd. 90HFKPLUR Printed by TEAMMARKING LTD. TMKZ0309



**THE HONG KONG
POLYTECHNIC UNIVERSITY**
香港理工大學

Staff Number:

STUDENT FEEDBACK QUESTIONNAIRE

Your feedback from this questionnaire will be useful to staff members in improving the quality of teaching. The overall results will also be considered in the annual staff appraisal exercise. **Your responses will remain anonymous and will not affect your assessment grades.** Please give your honest impression of this staff member in this academic year.

Marking Instructions:

1. Use HB pencil or black/blue point pen to fill in the oval completely.
e.g. ● right ○✗ wrong
2. Erase clearly any changes and make no stray marks.
3. Do not staple or fold the form.

Subject Code

1	2	3	4	5
6	7	8	9	0
1	2	3	4	5
6	7	8	9	0
1	2	3	4	5
6	7	8	9	0
1	2	3	4	5
6	7	8	9	0
1	2	3	4	5
6	7	8	9	0

Tutorial Group

1	2
3	4
5	6
7	8
9	0
1	2
3	4
5	6
7	8
9	0

For the questions on this page and the Extra Questions on the next page, please fill the appropriate oval to indicate your attitude to each statement:

= Strongly agree
 = Agree
 = Neutral
 = Disagree
 = Strongly disagree

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
LEARNING OUTCOMES 學習成果					
1. I have understood the subject matter taught by the staff member. 我明白這位教師所教的學科內容。	5	4	3	2	1
2. The staff member's method of teaching has helped my understanding. 這位教師所採用的教學方法能幫助我理解。	5	4	3	2	1
INTERACTION 課上交流					
3. The staff member gave students opportunities to ask questions and discuss ideas. 這位教師讓學生有機會發問和討論。	5	4	3	2	1
4. The staff member encouraged active participation in class. 這位教師鼓勵同學在課上積極參與。	5	4	3	2	1
INDIVIDUAL HELP 個別輔導					
5. The staff member provided appropriate help for students with learning problems. 這位教師會為在學習上遇到困難的同學提供適當的輔導。	5	4	3	2	1
6. Assistance was available from the staff member when necessary. 同學在有需要時能夠向這位教師尋求幫助。	5	4	3	2	1
ORGANISATION & PRESENTATION 組織和表達					
7. The staff member's teaching was well-organised. 這位教師教學時很有條理。	5	4	3	2	1
8. The staff member presented the subject material clearly. 這位教師教學時把內容表達得很清楚。	5	4	3	2	1
MOTIVATION 引發學習動機					
9. The staff member explained the significance of what was taught. 這位教師有解釋所教授的內容的重要性。	5	4	3	2	1
10. The staff member's teaching stimulated my interest in the subject. 這位教師的教學令我對這科目產生興趣。	5	4	3	2	1
FEEDBACK 回饋					
11. The staff member gave me regular feedback on my progress. 這位教師常會讓我了解自己的進度。	5	4	3	2	1
12. The feedback from the staff member was helpful and constructive. 這位教師提供的評語 / 意見有益而有建設性。	5	4	3	2	1

Please turn over

The Student Feedback Questionnaire Form (page 2)

W009598K/PC/08 TMR20030

EXTRA QUESTIONS (Please refer to the Extra Questions Form for the questions)

Strongly agree	Strongly disagree	Strongly agree	Strongly disagree	Strongly agree	Strongly disagree	Strongly agree	Strongly disagree	Strongly agree	Strongly disagree
13. (1)(2)(3)(4)(5)		17. (1)(2)(3)(4)(5)		21. (1)(2)(3)(4)(5)		25. (1)(2)(3)(4)(5)		29. (1)(2)(3)(4)(5)	
14. (1)(2)(3)(4)(5)		18. (1)(2)(3)(4)(5)		22. (1)(2)(3)(4)(5)		26. (1)(2)(3)(4)(5)		30. (1)(2)(3)(4)(5)	
15. (1)(2)(3)(4)(5)		19. (1)(2)(3)(4)(5)		23. (1)(2)(3)(4)(5)		27. (1)(2)(3)(4)(5)		31. (1)(2)(3)(4)(5)	
16. (1)(2)(3)(4)(5)		20. (1)(2)(3)(4)(5)		24. (1)(2)(3)(4)(5)		28. (1)(2)(3)(4)(5)		32. (1)(2)(3)(4)(5)	

OPEN-ENDED QUESTIONS

For the two questions below, write as much as you feel is necessary to give a full answer to the question. Comments made should focus on the quality of teaching and be free from racist/sexist overtones, or any other form of personal attack. If they are not they will be discarded. Please provide critical comment in a constructive way.

1. What aspects of this staff member's teaching were most helpful to your learning?
 這位教師的教學有哪幾方面對你的學習最有用？

.....

.....

.....

.....

.....

.....

2. How would you like the teaching be changed (if at all), to help you learn better in the subject?
 你認為這位教師的教學可以怎樣改變（如需要的話）來幫助你把這科學得更好？

.....

.....

.....

.....

.....

.....

Thank you for completing this questionnaire.

Appendix B: Letter to Chairpersons of Departmental Learning & Teaching Committees



THE HONG KONG
POLYTECHNIC UNIVERSITY
EDUCATIONAL DEVELOPMENT UNIT

31st March, 1999.

Dear

We write to request your assistance and support in a teaching development project we are carrying out funded by the PolyU Learning & Teaching Development Grant Fund. The project focuses on improving the interpretation and use of the Student Feedback Questionnaire (SFQ) data for enhancing teaching. The ultimate aim of the project is to develop a Web-based system which enables departments, staff, and other users of the SFQ data to comprehend and interpret student feedback data in a more appropriate and valid manner.

The Student Feedback Questionnaire (SFQ) has been in place in PolyU for three years. However, we know less than we should about the nature of the student ratings obtained in this exercise. For a more appropriate and meaningful interpretation of the data, we need to understand more about the following three items.

1. How do our students actually go about rating their teachers? For example, do students from different departments view 'good teaching' differently, and to what extent do they differ in their rating behaviours?
2. To what extent are student ratings on the SFQ influenced by the characteristics of the students in class? For example, how and to what extent are students' ratings influenced by students' motivation, conception of good teaching, and approaches to learning, etc.?
3. To which extent are student ratings influenced by the characteristics of the teaching of the subject? For example, to what extent are student ratings influenced by the teaching approaches, and the nature of knowledge of the subject – 'hard' versus 'soft' disciplines, and 'applied' versus 'theoretical', etc.?

As the first part of this project, we plan to conduct a series of individual interviews with students from different disciplinary backgrounds to explore their attitudes, rating behaviours, views about 'good teaching' in their discipline area, and the factors influencing their decision on the specific SFQ ratings they give for a particular staff member. The interviews will be conducted in Cantonese during April and May 1999.

We have identified six departments which broadly represent the major disciplinary areas in PolyU. Your department is one of the six selected for inclusion in the project, and your support would be invaluable in the successful completion of the project. We would be grateful if you could kindly help us by inviting 5 students at random from each of the year groups in a Full-time undergraduate programme offered by your department, for the interviews (total = 15 students). Please list the names and contact number of the students in the reply slip included and return it to us at your earliest convenience, preferably before 16 April, 1999. We shall then approach the students to make interview arrangements. Should you have any queries, please feel free to contact either one of us by phone (John: x 6320, KP: x 6287) or via e-mail.

We believe that the results of the project will contribute to a better understanding of student ratings of university teachers/teaching, and will enable a more meaningful interpretation and use of the SFQ data. Thank you in advance for your kind assistance.

Yours sincerely,

John Jones
Director of Educational Development, EDU

Kam-Por Kwan
Senior Officer, EDU

c.c. Hd (BUSS)
Encl.

Appendix C: The Interview Protocol

IMPROVING THE INTERPRETATION AND USE OF STUDENT FEEDBACK QUESTIONNAIRE (SFQ) DATA FOR ENHANCING TEACHING

Interview questions

1. *Brief introduction:*
 - Thank students for attending the interview
 - Introduce briefly the purposes and significance of the study
 - Ask for consent to participate in the study, and for audio- recording the interview

2. *Demographic information:*
 - Department / major discipline / year of study

3. *Conception about nature of discipline, teaching and learning*
 - Please describe the nature of your discipline area. What kind of ability or skills are needed for studying your course?
你就讀這科是一科怎樣的科目？有什麼特性？需要什麼的能力？
 - Based on your experience, what does it take for you to 'learn well' in the discipline?
根據你的經驗，在你就讀這科目如何才算「學得好」？
 - Based on your experience, what does it mean to 'teach well' in your discipline?
根據你的經驗，在你就讀這科目怎樣才算「教得好」？

4. *Questions about the administration of the feedback questionnaires*
 - When were the feedback questionnaire forms normally administered?
一般來說，這些評課問卷是何時派發的？
 - Who distribute them, and with what instructions before-hand?
誰負責派發？之前有什麼指示？
 - From your observation, have any of your teachers tried to influence your ratings on the forms and if yes, how?
憑你的觀察，你的教師有沒有嘗試去影響你們的評分？如有的話，怎樣去影響？

5. *Questions on rating behaviours:*
 - Do you usually give your ratings in a serious manner? Why?
在進行評課時，你一般會否認真地考慮才決定某位教師的評分？為什麼？
 - From your observation, do you think that your classmates give their ratings seriously? Why?
你的同學又怎樣？為什麼他們會這樣？
 - Do you think that your ratings are usually given in a fair and objective way? Why and how?
一般來說，你認為你對教師的評價是否公正和客觀？為什麼？怎樣去做到？
 - Do you and your classmates discuss the teaching performance of your teachers in daily conversation? When? Where? About what?
你和你的同學平時有沒有談論老師的表現？何時？何地？談些什麼？
 - Do you and your classmates usually have the same view on the teaching performance of your teachers?
一般來說，你們的意見是否一致？
 - Do you and your classmates discuss the teaching performance of your teachers during the evaluation process?
在進行評課時，你和你的同學有沒有談論該名老師的表現？

The Interview Protocol (page 2)

- Do you think that your evaluation of a teacher is affected by the views of your classmates?
你認為同學的看法會否影響你對該名教師的評價？
 - Can you explain how you arrive at a specific rating to give for a particular teacher when you are completing the student feedback forms? (What criteria do you use? To whom or what do you compare the teacher with? etc.)
你可否解釋一下你是如何決定某位教師的評分？（你用什麼的原則？和什麼或誰來比較？）
6. *A rating simulation*
Ask student to recall a teacher they have had this year whose teaching he/she thinks is really good (not the ideal one, nor necessarily perfect) and another whose teaching is unsatisfactory to him/her, then rate the two teachers separately on a rating form provided (with the 6 items drawn from the standardised SFQ and a few extra items).
7. *Actual considerations in deciding ratings on specific items*
For each pair of ratings given to the two teachers on a specific items, ask the following questions:
- Why did you give such a rating for this staff member on this item but that rating for the other?
你為什麼給這位教師這個分數而給那位教師那個分數呢？
 - How did the two teachers differ in their teaching that made you rate in this way?
這兩位教師在教學上有什麼分別令你這樣給分呢？
- Besides the items above, what other things do you think students should be asked to rate on to distinguish between 'good' and 'poor' teachers/teaching?
除上列之問卷題目外，你認為還應包括哪些題目來分辨教師教學的「好」與「壞」呢？
8. *Questions on attitudes and beliefs towards student feedback:*
- Why do you think the University wants to collect student feedback on teaching?
你認為大學方面為什麼要去搜集學生對教學的意見？
 - As a student, do you think you should be given the opportunity to rate the teachers teaching you receive? Why or why not?
你認為同學應否有機會去評課呢？為什麼？
 - Do you think that collecting student feedback will lead to improvements in teaching? Why do you think so?
你認為同學評課的制度會否導致教學的改善？為什麼？
 - Do you think that student feedback should be considered in personnel decisions such as promotions or contract renewals made by the University? Why or why not?
你認為大學在決定一名教授的升遷或續約時，應否考慮學生對該名教師的意見？為什麼？
 - Overall, how do you feel towards giving feedback on your teachers/teaching? Do you like it? Why or why not?
總的來說，你對學生評課這個制度有什麼感受？你是否喜歡它？為什麼？

Appendix D: Questionnaire for Rating of 'Good' and 'Poor' Teachers

A Study of Students' Rating Behaviours in Evaluating University Teachers

Instructions:

Think of a good teacher and a poor teacher that you have encountered this year and rate them separately on each of the items below according to the following scale:

1: Strongly Disagree 2: Disagree 3: Neutral 4: Agree 5: Strongly agree

Your department: _____

Year of study: _____

Please circle as appropriate:

Items	Rate the GOOD teacher for each of the items on the left:					Rate the POOR teacher for each of the items on the left:				
	SD	D	N	A	SA	SD	D	N	A	SA
1. The staff member's method of teaching has helped my understanding. 這位教師所採用的教學方法能幫助我理解。	1	2	3	4	5	1	2	3	4	5
2. The staff member encouraged active participation in class. 這位教師鼓勵同學在課上積極參與。	1	2	3	4	5	1	2	3	4	5
3. The staff member provided appropriate help for students with learning problems. 這位教師會為在學習上遇到困難的同學提供適應的輔導。	1	2	3	4	5	1	2	3	4	5
4. The staff member presented the subject material clearly. 這位教師教學時把內容表達得很清楚。	1	2	3	4	5	1	2	3	4	5
5. The staff member's teaching stimulated my interest in the subject. 這位教師的教學令我對這科目產生興趣。	1	2	3	4	5	1	2	3	4	5
6. The staff member gave me regular feedback on my progress. 這位教師常會讓我知道自己的進度。	1	2	3	4	5	1	2	3	4	5

REFERENCES

- Abbott, R. D., Wulff, D. H., Nyquist, J. D., Ropp, V. A., & 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., d'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology*, 82, 219-231.
- Abrami, P. C., d'Apollonia, S., & Rosenfield, S. (1996). The dimensionality of student ratings of instruction: What we know and what we do not. In: J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research*, Volume 11, 213-264. New York: Agathon Press.
- Abrami, P. C., Leventhal, L., & Perry, R. P. (1982). *Educational seduction. Review of Educational Research*, 52, 446-464.
- Adelman, C. (1997). Judging quality in higher education: Towards bureaucratic conformity. In: K. Watson, C. Modgil & S. Modgil (Eds.) *Reforms in Higher Education. Educational Dilemmas: Debate and Diversity*. Vol. 2. London: Cassell.
- Aleamoni, L. M. (1981). Student ratings of instruction. In: J. Millman (Ed.), *Handbook of Teacher Evaluation*. Beverly Hills, CA: Sage.
- Angelo, T.A., & Cross, K. P. (1993). *Classroom Assessment Techniques: A Handbook for College Teachers*. Second Edition. San Francisco: Jossey-Bass.
- Arreola, R. A. (1987). The role of student government in faculty evaluation. In: L. M. Aleamoni (Ed.) *Techniques for Evaluating and Improving Instruction*, 39-46. (New Directions for Teaching and Learning, Number 13, Fall 1987). San Francisco: Jossey-Bass.
- Ballantyne, C. (1999). Improving university teaching: Responding to feedback from students. In: N. Zepke, M. Knight, & A. Viskovice (Eds.) *Adult Learning Cultures: Challenges and Choices in Times of Change*. Wellington: WP Press.

- Banner, J. M. Jr., & Cannon, H. C. (1997). The 'who' of teaching. *Education Week*, April 16, 1997.
- Bauer, H. H. (1997). In response... [to Trout, 1997]. *Change*, September/October 1997, 26.
- Becker, B. E., & Cardy, R. L. (1986). Influence of halo error on appraisal effectiveness: A conceptual and empirical reconsideration. *Journal of Applied Psychology*, 71, 662-671.
- Benz, C. R., & Blatt, S. J. (1996). Meaning underlying student ratings of faculty. *The Review of Higher Education*, 19 (4), 411-433.
- Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, 57 (3), 195-203.
- Borg, W. R. (1987). *Applying Educational Research: A Practical Guide for Teachers*. Second Edition. London: Longman.
- Borg, W. R., & Gall, M. D. (1979). *Educational Research*. New York & London: Longman.
- Boyer, E. L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate*. Princeton, N.J.: Carnegie Foundation for the Advancement of Teaching.
- Brinko, K. T. (1990). Instructional consultation with feedback in higher education. *Journal of Higher Education*, 61, 65-83.
- Brinko, K.T. (1993). The practice of giving feedback to improve teaching: What is effective? *Journal of Higher Education*, 64 (5), 575-593.
- Braskamp, L. A., & Ory, J. C. (1994). *Assessing Faculty Work: Enhancing Individual and Institutional Performance*. San Francisco: Jossey-Bass
- Cadwell, J., & Jenkins, J. (1985). The effects of semantic similarity of items on student ratings of instruction. *Journal of Educational Psychology*, 77 (4), 383-393.
- Cahn, S. (1987). Faculty members should be evaluated by their peers, not by their students. *Chronicle of Higher Education*, (October 14), B2.

- Candy, P. C. (1989). Alternative paradigms in educational research. *Australian Educational Researcher*, 16 (3), 1-11.
- Cannell, C. F., & Kahn, R. L. (1968). Interviewing. In: G. Lindzey & A. Aronson (Eds.) *The Handbook of Social Psychology, Vol. 2, Research Methods*. New York: Addison Wesley.
- Carr, W. & Kemmis, S. (1983). *Becoming Critical: Knowing Through Action Research*. Highton, Victoria: Deakin University Press.
- Cashin, W. E. (1988). *Student Ratings of Teaching: A Summary of the Research*. IDEA Paper No. 20. Manhattan, KS: Kansas State University, Centre for Faculty Evaluation and Development.
- Cashin, W. E. (1990). *Student Ratings of Teaching: Recommendations for Use*. IDEA Paper No. 22. Kansas State University: Centre for Faculty Evaluation & Development.
- Cashin, W. E. (1995). *Student Ratings of Instruction: The Research Revisited*. IDEA Paper No. 32. Manhattan, KS: Kansas State University, Centre for Faculty Evaluation and Development.
- Cave, M., Hanney, S., Henkel, M. & Kogan, M. (1997). *The Use of Performance Indicators in Higher Education: The Challenges of the Quality Movement*. Third Edition. London: Jessica Kingsley Publishers.
- Centra, J. A. (1993). *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*. San Francisco: Jossey-Bass.
- Chandler, T. A. (1978). The questionable status of student evaluation of teaching. *Teaching of Psychology*, 5 (3), 150-152.
- Cholakian, R. (1994). The value of evaluation. *Academe*, 80 (5), 24-26.
- Coe, R. (1998). Can feedback improve teaching? A review of the social science literature with a view to identifying the conditions under which giving feedback to teachers will result in improved performance. *Research Papers in Education*, 13 (1), 43-66.

- Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis. *Research in Higher Education*, 13, 321-341.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309.
- Cohen, P. A. (1987). *A critical analysis and reanalysis of the multisection validity meta-analysis*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC, April 1987.
- Cohen, L., & Manion, L. (1994). *Research Methods in Education*. Fourth Edition. London & New York: Routledge.
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90, 218-244.
- Crittenden, K. S., & Norr, J. L. (1973). Student Values and Teacher Evaluation: A Problem in Person Perception. *Sociometry*, 36 (2) 143-151.
- Crutcher, R. J. (1994). Telling what we know: The use of verbal reports methodologies in psychological research. *Psychological Science*, 5 (5), 241-244.
- d'Apollonia, S., & Abrami, P. (1996). *Variables moderating the validity of student ratings of instruction: A meta-analysis*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA, April 1996.
- d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52 (11), 1198-1208.
- Dean, J., & Whyte, W. F. (1958). How do you know if the informant is telling the truth? *Human Organization*, 17 (2), 34-38.
- Dowell, D. A., & Neal, J. A. (1982). A selective review of the validity of student ratings of teaching. *Journal of Higher Education*, 53, 51-62.
- Dunkin, M. J. (1986). Research on teaching in higher education. In: M. C. Wittrock (Ed.). *Handbook of Research on Teaching*, 3rd Edition. New York: MacMillan Publishing Co.

- Elton, L. (1984). Evaluating teaching and assessing teachers in universities. *Assessment and Evaluation in Higher Education*, 9 (2), 97-115.
- Entwistle, N. J., & Tait, H. (1994). Approaches to studying and preferences for teaching in higher education: Implications for student ratings. *Instructional Evaluation and Faculty Development*, 14 (1-2), 2-9.
- Ericsson, K. A., & Simon, H. A. (Eds.) (1993) *Protocol Analysis: Verbal Reports as Data*. Revised Edition. Cambridge, MA: MIT Press.
- Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers – what we know and what we don't. *Research in Higher Education*, 9 (3), 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), 45-116.
- Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. *Research in Higher Education*, 24 (2), 139-213.
- Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatches priorities. *Research in Higher Education*, 26, 227-298.
- Feldman, K. A. (1989a). 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), 583-645.
- Feldman, K. A. (1989b). Instructor 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. (1993). College students' views of male and female college teachers. Part II – Evidence from student evaluations of their classroom teachers. *Research in Higher Education*, 34 (2), 151-211.
- Feldman, K. A. (1998). Reflections on the study of effective college teaching and student ratings: One continuing quest and two unresolved issues. In: J. C. Smart (Ed.). *Higher Education: Handbook of Theory and Research*, Volume 13, 35-74. New York: Agathon Press.
- Fitz-Gibbon, C. T. (1997). Monitoring with feedback: The democratization of data: A response to Lindsay Paterson. In: K. Watson, C. Modgil, & S. Modgil. (Eds.) *Educational Dilemmas: Debate and Diversity, Volume 4, Quality in Education*, 313-321. London: Cassell.
- Franklin, J., & Theall, M. (1989). *Who Read Ratings: Knowledge, Attitude and Practice of Users of Student Ratings of Instruction*. Paper presented at the 70th annual meeting of the American Educational Research Association, April 1989, San Francisco.
- Gaski, J. F. (1987). On construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology*, 79 (3), 326-330.
- Gay, L. R. (1976). *Educational Research: Competencies for Analysis and Application*. Columbus: Bell and Howell.
- Gilmore, G. M., Kane, M. T., & Naccarato, R. W. (1978). The generalizability of student ratings of instruction: Estimates of teacher and course components. *Journal of Educational Measurement*, 15, 1-13.
- Goodwin, L. D., & Stevens, E.A. (1993). The influence of gender on university faculty members' perceptions of 'good' teaching. *Journal of Higher Education*, 64 (2), 166-185.
- Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52 (11), 1182-1186.

- Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52 (11), 1209-1217.
- Guba, E. G., & Lincoln, Y. S. (1984). Epistemological and methodological bases of naturalistic inquiry. In: G. F. Madaus *et.al.* (Eds.). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer-Nijhoff Publishing.
- Harrison, P. D., Ryan, J. M., & Moore, P. S. (1996). College students' self-insights and common implicit theories in ratings of teaching effectiveness. *Journal of Educational Psychology*, 88 (4), 775-782.
- Haskell, R. E. (1997). Academic freedom, tenure, and student evaluation of faculty: galloping polls in the 21st century. *Educational Policy Analysis Archives*, 5 (6). Available: <http://olam.ed.asu.edu/epaa/v5n6.html>
- Hativa, N. (1996). University instructors' ratings profiles: Stability over time and disciplinary differences. *Research in Higher Education*, 37 (3), 341-365.
- Hitchcock, H., & Hughes, D. (1995). *Researcher and the Teacher: A Qualitative Introduction to School-Based Research*. 2nd Edition. London and New York: Routledge.
- Ho, A. S. P. (2000). A conceptual change approach to staff development: A model for programme design. *The International Journal for Academic Development*, 5 (1), 30-41.
- Hoyt, D. P., & Cashin, W.E. (1977). *Development of the IDEA System* (IDEA Technical Report Number 1). Manhattan: Kansas State University. Center for Faculty Development and Development.
- Hoyt, D. P., & Pallett, W. H. (1999). *Appraising Teaching Effectiveness: Beyond Student Ratings*. IDEA Paper No. 36. IDEA Centre, Kansas State University.
- Husbands, C. T. (1998). Implications for the assessment of the teaching competence of staff in higher education of some correlates of students' evaluations of different teaching styles. *Assessment & Evaluation in Higher Education*, 23 (2), 117-139.

- Husbands, C. T., & Fosh, P. (1993). Students' evaluation of teaching in higher education: Experience from four European countries and some implications of the practice. *Assessment & Evaluation in Higher Education*, 18 (2), 95-114.
- Ingrim, L. C. (1979). Broken frames and emergent meaning: the case of student evaluation of teaching. *Human Relations*, 32 (9), 803-818.
- Jacobs, L. C. (1987). *University faculty and students' opinions of student ratings*. Indiana Studies in Higher Education, No. 55. Bloomington, Indiana: Indiana University, Bureau of Evaluative Studies and Testing.
- Johnes, J. & Taylor, J. (1990). *Performance Indicators in Higher Education: UK Universities*. Buckingham: SRHE & Open University.
- Keeves, J. P. (1988). Towards a unified approach. In: J. P. Keeves (Ed.) *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford: Pergamon Press.
- Kelly, G. A. (1955). *The Psychology of Personal Constructs*. New York: Norton.
- Kelly, G. A. (1969). *Clinical Psychology and Personality: the Selected Papers of George Kelly*. Edited by B. A. Maher. New York: John Wiley & Sons.
- Kember, D., & Wong, A. (2000). Implications for evaluation from a study of students' perceptions of good and poor teaching. *Higher Education*, 40, 69-97.
- Kerridge, J. R., & Mathews, B. P. (1998). Student ratings of courses in HE: Further challenges and opportunities. *Assessment & Evaluation in Higher Education*, 23 (1), 71-82.
- Kishor, N. (1995). The effect of implicit theories on raters' inference in performance judgment: Consequences for the validity of student ratings of instruction. *Research in Higher Education*, 36 (2), 177-195.
- Kolitch, E., & Dean, A. V. (1999). Student ratings of instruction in the USA: Hidden assumptions and missing conceptions about 'good' teaching. *Studies in Higher Education*, 24 (1), 27-42.

- Kulik, J. A., & McKeachie, W. J. (1975). The evaluation of teachers in higher education. In: F. Kerlinger (Ed.). *Review of Research in Education*, Volume 3, 210-240. Itasca, IL: Peacock.
- Kwan, K. P. (1999). How fair are student ratings in assessing the teaching performance of university teachers? *Assessment & Evaluation in Higher Education*, 24 (2), 181-195.
- Landy, F. J., & Farr, J. L. (1980). Performance ratings. *Psychological Bulletin*, 87, 72-107.
- LeCompte, M. D., Preissle, J., & Tesch, R. (1993). *Ethnography and Qualitative Design in Educational Research*. Second Edition. San Diego, California: Academic Press Ltd.
- L'Hommedieu, R., Menges, R.J., & Brinko, K. T. (1990). Methodological explanations for the modest effects of feedback from student ratings. *Journal of Educational Psychology*, 82 (2), 232-241.
- Lin, W., Watkins, D., & Meng, Q. (1995). Student evaluations of university teaching: A Chinese perspective. *Higher Education Research and Development*, 14 (1), 61-74.
- Lo, T., Wong, W., Barrett, J., & Wong, C. (1999). *An Evaluation Sourcebook for Higher Education in Hong Kong*. Manuscript published by the Evaluation of Student Experience Project funded by the University Grants Committee, Hong Kong.
- Lodzinski, A. (1991). *The systematic distortion hypothesis and diagnostically meaningful student ratings of university instructor effectiveness*. Unpublished doctoral dissertation, University of Windsor, Windsor, Ontario.
- Magner, D. K. (1994). Report to focus on standards for assessing what professors do. *Chronicle of Higher Education*, 40 (23), A22. 9 February 1994.
- Marchese, T. (1997). Student evaluation of teaching. *Change*, September/October 1997, 4.

- Marincovich, M. (1998). *Ending the Disconnect Between the Student Evaluation of Teaching and the Improvement of Teaching: A Faculty Developer's Plea*. Stanford University, CA: National Centre for Postsecondary Education.
- Marlin, J. W. Jr. (1987). Student perception of end-of course evaluations. *Journal of Higher Education*, 58, 704-716.
- Marsh, H. W. (1981). The use of path analysis to estimate teacher and course effects in student ratings of instructional effectiveness. *Applied Psychological Measurement*, 6, 47-60.
- Marsh, H. W. (1984). Students' evaluation of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76 (5), 707-754.
- Marsh, H. W. (1987). Students' evaluation of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11 (3), 253-388.
- Marsh, H. W., & Dunkin, M. (1992). Students' evaluation of university teaching: A multidimensional perspective. In: J. C. Smart (Ed.). *Higher Education: Handbook on Theory and Research*, Volume 8, 143-234, New York: Agathon Press.
- Marsh, H. W., & Hocevar, D. (1991). Student evaluations of teaching effectiveness: The stability of mean ratings of the same teachers over a 13-year period. *Teaching and Teacher Education*, 7, 303-314.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52 (11), 1187-1197.
- Marsh, H. W., & Ware, J. E. (1982). Effects of expressiveness, content coverage, and incentive on multi-dimensional student rating scales: New interpretations of the Dr. Fox effect. *Journal of Educational Psychology*, 74, 126-134.
- McCutcheon, G. (1981). On the interpretation of classroom observation. *Educational Researcher*, 19 (5), 5-10.

- McGaghie, W. C. (1991). Professional competence evaluation. *Educational Researcher*, 20 (1), 3-9.
- McKeachie, W. J. (1997a). Student ratings: The validity of use. *American Psychologist*, 52 (11), 1218-1225.
- McKeachie, W. J. (1997b). Good teaching makes a difference – and we know what it is. In: R. P. Perry & J. C. Smart (Eds.) *Effective Teaching in Higher Education: Research and Practice*, 396-408. New York: Agathon Press.
- McKeachie, W. J., & Kaplan, M. (1996). Persistent problems in evaluating college teaching. *AAHE Bulletin*, February 1996, 5-8.
- Meech, L. R. (1976). The stateless art of teaching evaluation. Reprinted in D. DeZure (2000) (Ed.) *Learning from Change: Landmarks in Teaching and Learning in Higher Education from Change Magazine 1969-1999*, 347-349, Sterling, Virginia: Stylus Publishing, LLC.
- Miller, P. V., & Cannell, C. F. (1988). Interviews in sample surveys. In: J. P. Keeves (Ed.). *Educational Research, Methodology and Measurement: An International Handbook*. Oxford, New York: Pergamon.
- Murray, H. G. (1985). Classroom teaching behaviours related to college teaching effectiveness. In: J. G. Donald & A. M. Sullivan (Eds.) *Using Research to Improve Teaching Effectiveness*, 21-35. (New Directions for Teaching and Learning, No. 23). San Francisco: Jossey-Bass.
- Murray, H. G. (1987). *Impact of student instructional ratings on quality of teaching in higher education*. Paper presented at the 1987 annual meeting of the American Educational Research Association, Washington, D. C. (ERIC Document Reproduction Service No. ED 013 371).
- Murray, H. G. (1991). Effective teaching behaviors in the college classroom. In: J. C. Smart (Ed.). *Higher Education: Handbook of Theory and Research*, Volume 7. New York: Agathon Press.
- Naftulin, D. H., Ware, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 48, 630-635.

- Newport, J. F. (1996). Rating teaching in the USA: Probing the qualifications of student raters and novice teachers. *Assessment & Evaluation in Higher Education*, 21 (1), 17-21.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Novice, R. (1996). Actual schools, possible practices: New directions in professional development. *Educational Policy Analysis Archives*, 4 (14). Available: <http://olam.ed.asu/epaa/v4n14.html>
- Overall, J. U., & Marsh, H. W. (1980). Midterm feedback from students: Its relationship to instructional improvements and students' cognitive and affective outcomes. *Journal of Educational Psychology*, 71, 856-865.
- Owens, R. (1982). Methodological rigor in naturalistic inquiry: Some issues and answers. *Educational Administration Quarterly*, 18 (2), 1-12.
- Palmer, P. J. (1999). The heart of a teacher: identity and integrity in teaching. In; B. A. Pescosolido & R. Aminzade (Eds.) *The Social Worlds of Higher Education: Handbook for Teaching in a New Century*, 504-506. Thousand Oaks, California: Pine Forge Press.
- Payne, J. W. (1994). Thinking aloud: insights into information processing. *Psychological Science*, 5 (5), 241-248.
- Piccinin, S., Cristi, C., & McCoy, M. (1999). The impact of individual consultation on student ratings of teaching. *International Journal for Academic Development*, 4(2), 75-88.
- Pidgeon, N. (1996). Grounded theory: Theoretical background. In J. T. E. Richardson (Ed.), *Handbook of Qualitative Research Methods for Psychology and the Social Sciences*. Leicester, U.K.: BPS Books.
- Pratt, D. D., Kelly, M., & Wong, W. S. S. (1999). Chinese conceptions of 'effective teaching' in Hong Kong: Towards culturally sensitive evaluation of teaching. *International Journal of Lifelong Education*, 18 (4), 241-258.

- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The course experience questionnaire. *Studies in Higher Education*, 16, 129-150.
- Ramsden, P., & Dodds, A. (1989). *Improving Teaching and Course: A Guide to Evaluation*. Parkville, Victoria: The University of Melbourne, Centre for the Study of Higher Education.
- Rist, R. C. (1977). On the relations among educational research paradigms: from disdain to détente. *Anthropology and Education Quarterly*, 8 (2), 42-49.
- Rudduck, J., & Hopkins, D. (1985). *Research as a Basis for Teaching*. London: Hienemann Educational Books.
- Ryan, J. J., Anderson, J. A., & Birchler, A. B. (1980). Student evaluation: The faculty responds. *Research in Higher Education*, 12 (4), 317-333.
- Sacks, P. (1997). In reponse... [to Trout, 1997]. *Change*, September/October 1997, 29.
- Schmelkin, L. P., Spencer, K. J., & Gellman, E. S. (1997). Faculty perspectives on course and teacher evaluations. *Research in Higher Education*, 38 (5), 575-592.
- Schon, P. A. (1987). *Educating the Reflective Practitioners: Toward a New Design for Teaching and learning in the Profession*. San Francisco: Jossey-Bass.
- Scriven, M. (1981). Summative teacher evaluation, In: J. Millan (Ed.) *Handbook of Teacher Evaluation*, 244-271. Beverly Hills: Sage Publication.
- Scriven, M. (1988). The validity of student ratings. *Instructional Evaluation*, 9, 5-18.
- Scriven, M. (1995). *Student Ratings Offer Useful Input to Teacher Evaluation*. ERIC/AE Digest. Available:
http://www.ed.gov/databases/ERIC_Digest/ed398240.html
- Seldin, P. (1998). How colleges evaluate teaching 1988 vs. 1998. *AAHE Bulletin*, March 1998, 3-7.

- Smith, H. W. (1975). *Strategies of Social Research: The Methodological Imagination*. London: Prentice-Hall.
- Smith, M. C., & Carney, R. N. (1990). *Students' perceptions of teaching evaluation process*. Paper presented at the annual meeting of the American Educational Research Association.
- Spencer, K. J., & Schmelkin, L. P. (1995). *Student perspectives on teaching and its evaluation*. Paper presented at the annual meeting of the American Educational Research Association.
- Stake, J. E. (1997). Response to Haskell: "Academic Freedom, Tenure, and Student Evaluation of Faculty". *Education Policy Analysis Archives*, 5 (8). Available: <http://olam.ed.asu.edu/epaa/v5n8.html>
- Stone, J. E. (1995, June). Inflated grades, inflated enrollment, and inflated budgets: An analysis and call for review at the state level. *Education Policy Analysis Archives*, 3 (11), 1-33. Available: <http://olam.ed.asu.edu/epaa/v3n11.html>
- Tagomori, H. T., & Bishop, L. A. (1999). *Student Evaluation of Teaching: Flaws In the Instruments*. Available: <http://www.ivygreen.ctc.edu/InstEval/stu-eval.htm>
- Talley, K. D., & Timmer, D. A. (1992). A qualitative methods exercise: Student and faculty interpretations of classroom teaching. *Teaching Sociology*, 20, 75-79.
- Tang, T. L., & Tang, T. L. (1987). A correlational study of students' evaluation of faculty performance and their self-ratings in an instructional setting. *College Student Journal*, 21 (Spring), pp. 90-97.
- Taylor, J. S., & Ricketts, M. S. (1982). Student attitudes about participation in instructor and course evaluations. *Journal of Instructional Psychology*, 9 (1), 18-24.
- Theall, M. (2000). Message sent to the POD Network (Professional and Organizational Development in Higher Education) mailing list at pod@catfish.valdosta.edu, 19 April 2000.

- Theall, M., & Franklin, J. (1990). *Student Ratings of Instruction: Issues for Improving Practice*. (New Directions for Teaching and Learning No. 43). San Francisco: Jossey-Bass.
- Trout, P. A. (1997). What the numbers means: Providing a context for numerical student evaluations of courses. *Change*, September/October 1997, 24-30.
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23 (2), 191-211.
- Walker, J. C., & Evers, C. W. (1999). Research in education: Epistemological issues. In: J. P. Keeves & G. Lakomski (Eds.) *Issues in Educational Research*. Oxford: Pergamon/ Elsevier Science Ltd.
- Ware, J. E., & Williams, R. G. (1975). The Dr. Fox effect: A study of lecturer expressiveness and ratings of instruction. *Journal of Medical Education*, 5, 149-156.
- Ware, J. E., & Williams, R. G. (1977). Discriminant analysis of student ratings as a means of identifying lecturers who differ in enthusiasm or information giving. *Educational and Psychological Measurement*, 37, 627-639.
- Weimer, M. (1997). Exploring the implications: From research to practice. In: R. P. Perry & J. C. Smart (Eds.) *Effective Teaching in Higher Education: Research and Practice*, 411-435. New York: Agathon Press.
- Whitely, S. E., & Doyle, K. O. (1976). Implicit theories in student ratings. *American Journal of Educational Research*, 13 (4), 241-253.
- Williams, R. G., & Ware, J. E. (1976). Validity of student ratings of instruction under different incentive conditions: A further study of the Dr. Fox effect. *Journal of Educational Psychology*, 68, 48-56.
- Williams, R. G., & Ware, J. E. (1977). An extended visit with Dr. Fox: Validity of student ratings of instruction after repeated exposures to a lecturer. *American Educational Research Journal*, 14, 449-457.

- Williams, W. M., & Ceci, S. J. (1997). How'm I doing? Problems with student ratings of instructors and courses. *Change*, September/October 1997, 13-23.
- Wilson, T. D. (1994). The proper protocol: validity and completeness of verbal reports. *Psychological Science*, 5 (5), 249-252.
- Wragg, T. (1993). Foreword. In: H. Green (Ed.) *The School Management Handbook*. London: Kogan Page.
- Wulff, D. H., Staton-Spicer, A. Q., Hess, C. W., & Nyquist, J. D. (1985). The student perspective on evaluating teaching effectiveness. *ACA Bulletin*, 53, 39-47.

