USING STUDENT EVALUATIONS FOR TEACHER DEVELOPMENT IN NINE URBAN SECONDARY SCHOOLS

STEPHEN EARL IRVING

1996

Abstract

Student evaluations are becoming more common in education, particularly in the tertiary sector. They have been used by students and by the authorities to make judgments about teachers and courses. Their use for helping teachers to improve their teaching is a recent phenomenon which has barely touched secondary schooling, let alone primary schools.

Teacher evaluation data collected for summative purposes have had little effect on teaching performance as there is usually little or no feedback designed to help the teacher to improve. Teachers need to know not only what to change but also how to change their behaviours in a desired direction. This research set out to develop a questionnaire and methodology that could be used by secondary teachers to evaluate their teaching using the students as the source of information, and then use that information to help the teacher to improve their teaching. The methodology draws heavily on the work of Wilson (1986), and Marsh and Roche (1993, 1994) who constructed a process that supplemented feedback with a collegial consultation to help the teacher interpret the data in a meaningful way and then act on it. This methodology has been shown to be the best practice in this field, even though the results of carefully researched studies are modest.

A questionnaire appropriate to the New Zealand secondary school environment was constructed and administered in nine urban secondary schools to 344 students. The subject teachers they evaluated were from a wide cross-section of curriculum areas. Most were experienced teachers. At the same time, the teachers completed a self evaluation using the same questionnaire. The teachers received the results of the evaluation with notes on how to interpret the tables and graphs. This was followed by a consultation with the researcher, using a methodology developed from appraisal interviewing techniques.

An action plan was devised during this consultation. The teacher then put this into action, and the students were re-surveyed after approximately thirteen weeks. The results of the two surveys were compared to see whether this process was beneficial in improving teaching, as perceived by the students.

Overall, the results showed a rather modest improvement across the board. There was a noticeable difference between two groups of teachers dependent on the difference between their own self-evaluation and the average student response. Teachers whose self-evaluation was similar to the student evaluation, or whose self-evaluation was worse than the student evaluations changed little between the two administrations of the questionnaire. On the other hand, if the self-evaluation was better than the average student evaluation, then there were significant improvements in the student evaluations on the second administration. This finding is in keeping with the theory of cognitive dissonance, first espoused by Festinger (1957). When the teacher has a positive self-evaluation but the students rate that teacher poorly, then the teacher is motivated to change their teaching behaviours so that the next student evaluation is favourable.

As part of a teacher development programme, students and teachers similarly felt that this form of evaluation is valuable and has a place in appraisal schemes designed to help teachers improve their teaching. There is still considerable reluctance on the part of teachers for this type of evaluation tool to be used for the purposes of promotion, tenure and reward. In light of the requirements for schools to implement performance appraisal schemes, and the need for appraisals to be based on "objective" data, student evaluations can provide the desired information.

Acknowledgments

Contents

List of Tables

List of Appendices

Chapter 1   Introduction

This thesis is based on the principles of empirical research, as an investigation of student evaluations in secondary schools together with collegial consultation as a means of enhancing a teacher's professional development. The approach used was to provide a comprehensive trial and review of the use of student evaluations in actual secondary school environments, and to seek teacher and student opinions with a view to proposing an alternative method of obtaining data for teacher improvement and using that data to maximise its effect.

A comprehensive review of the literature on student evaluations was undertaken to determine an appropriate methodology which was valid, reliable and generalisable, and effective in helping teachers to improve. Of necessity, this review was of literature written about research in tertiary institutions - worldwide, very little research has been conducted in secondary schools (and even less in primary schools). Therefore, "best practice" conclusions are drawn from the literature in an effort to use what information is available. The findings of this review are contained in Chapter Two.

Sixteen classes and teachers in nine urban secondary schools were involved in the research. Both quantitative and qualitative research data were obtained by student evaluation questionnaires, teacher self-ratings, the consultations with the teachers, and questionnaires from both students and teachers seeking their feelings about the methodology and its effects. Detailed information about the methodology is contained in Chapter Three, Methodology.

The data were analysed by statistical analysis in the case of the quantitative data, with factor analysis as the basis for drawing conclusions from the data set. Content analysis was used for the qualitative data, using the student evaluation questionnaire as the framework. Throughout, the two sets of data were compared for agreement and also for divergent findings or observations. The findings of the research are presented in Chapter Four, Results.

The findings are discussed in light of their ability to ascertain whether teachers, as perceived by their students, can improve using this method. Related questions are

also addressed, and implications for teachers and further research are outlined. Chapter Five, Discussion, synthesises the findings and observations, and points the way that student evaluations could be used in secondary schools to help develop better teachers.

This research grew out of a required reading for a first year paper of the Master of Educational Administration degree at Massey University. Haller and Strike (1986: 286-324) discuss a case of teacher incompetence, and traverse the legal and ethical issues that surround this matter. One of the issues raised was the use of student evaluations for determining teacher competence (pp299-303). This sparked a determination to find out whether student evaluations could be positively used in secondary schools to assist teachers in their professional development rather than as tools to make decisions affecting their continued employment. A key statement was that it did not matter if the teachers "taught to the test" - if the items are known to improve student learning, then teaching to the test is exactly what is desired (Haller and Strike, 1986: 302).

Research Objectives

From reading the literature, the researcher made a number of assumptions which underpin the use of student evaluations for this purpose. They are:

1 Student evaluations are a valid, reliable, stable, useful, and cost-effective means of gathering data about what happens in the classroom.

2 Students have experienced a wide range of teachers and teaching and have built up a clear understanding of what they regard as effective teaching.

3 Students can communicate this understanding when asked in an appropriate way.

4 Students are willing to provide their opinions about the teaching and learning in their classrooms if they are asked.

5 Teachers are committed to students and their learning, and intend to teach to the best of their ability.

6 Teachers know the subjects they teach and the students they teach, and are best placed to determine how to teach those subjects to those students.

7 Teachers will use data from students to reflect on their teaching and to devise action plans that seek to improve some aspects of their teaching.

8 Teachers and students are partners in the teaching-learning relationship in the classroom and can benefit from regular and systematic two-way feedback.

9   Student evaluations are a partial indicator of teacher effectiveness and consequently are just one of the means available for obtaining the information necessary for a teacher development programme.

10   Teachers will use student evaluations for teacher development purposes when the purpose is purely formative and uncontaminated by summative purposes, and where there is a climate free from bureaucratic control.

11   Teacher development is a continuous process.

Based on these assumptions, the main research question to be answered was developed:

Are student evaluations of teachers, coupled with a consultation conference to develop an action plan for change, an effective and useful means of bringing about teacher development, particularly improvement?

Subsidiary research questions were:

Do teachers accept student evaluations as a valid, reliable, credible and useful means of gathering data for development, and are they therefore prepared to act upon them?

Do teachers see this form of evaluation as part of a developmental (formative) appraisal process, as part of a performance (summative) appraisal process, or as part of both? How best can student evaluations be incorporated in an appraisal system, if at all?

Are the time requirements realistic and manageable for teachers in gathering and analysing data, and developing and implementing an action plan for improvement?

In addition, as this type of teacher evaluation was rare in secondary schools especially in New Zealand, this research provided the ideal opportunity to go on a "fishing expedition" and see what students in New Zealand secondary schools thought about their teachers and learning. Therefore there was an additional general research question:

What can we learn about the way that secondary students view their teachers and their learning experiences?

In <u>Making It Happen</u>, Stewart and Prebble (1985) outlined the four essential stages of a school development model - data gathering, increasing collaboration, structural change, and improving the quality of teaching and learning. Their emphasis was on a process rather than a product. In a later work, Stewart and Prebble (1993) developed a model for appraising teachers based on a conceptual job description which encouraged them to reflect on what they are doing. Throughout, the primary emphasis was on "helping students to acquire new knowledge and skills" (Stewart and Prebble, 1993: 206). Borrowing from these models and applying them to teacher development, this thesis investigates one means of gathering data about a teacher's performance and using that data in a collaborative and reflective way to make the changes which the teacher believes will improve the quality of teaching and learning.

To apply this approach to the Stewart and Prebble models, the researcher examined the literature on student evaluations and ratings to establish a justification for and validation of the methodology used. This is explored in the next chapter, A Review of the Literature.

Chapter 2   A Review of the Literature

This research is about the use of student evaluations of teaching performance coupled with a consultation approach to feedback for the purpose of teacher development within the wider context of teacher evaluation and teacher appraisal. This focus on a teacher development strategy (which uses student evaluations to provide the data that is a fundamental requirement of a good teacher development process) was prompted by two statements, one thirty years ago:

> Whether the student's judgement is correct is largely beside the point. The real point is that his attitude toward the instructor is a vital factor in the total learning situation. ... Nor has the teacher any choice as to whether he will be "rated" by his students. Such rating goes on in every classroom everywhere. The only real choice the instructor has is whether he wants to know what these ratings are. If he chooses to get this knowledge, he is in a position to profit thereby. (Remmers and Weisbrodt, 1965 in Page 1974: 17)

and later:

> the most important purpose of evaluation is not to prove but to improve (Stufflebeam et al, 1971: i).

In Chapter One, eleven assumptions were outlined regarding the use of student evaluations for teacher development purposes. These assumptions are based on the researcher's understanding and interpretation of the existing literature. Teachers have very real reservations about asking students to evaluate their teaching performance. Therefore, teachers want to be convinced that their fears are largely unfounded especially when the purpose is to help them to improve their teaching. This chapter examines these assumptions and the practice that follows from them, with a view to addressing the common issues that are inevitably raised concerning the validity, reliability and use of student evaluations. Until this is done, teachers will give preference to other methods of teacher evaluation (whether for summative or formative purposes) for which the same standards of evaluation performance are not demanded. This dilemma is summed up by one commentator who said that

better teacher evaluation will only occur when educators attend to the issues of reliability and validity that have so far only been addressed in any substantial way in the research on student ratings of instruction (Aubrecht, 1984: 89).

The same opinion was also expressed by Dowell and Neal (1982).

Therefore, the first part of this review will call on the extensive research conducted in the tertiary education sector and endeavour to address these concerns by covering general issues about student rating forms before moving on to consider issues related to their use in a teacher development environment.

Part One: Student Evaluation of Teacher Performance

Student evaluations of teaching performance have a long history within the overall context of teacher evaluation. Indeed, students have probably been evaluating their teachers ever since the first teacher-student relationship was established, but teachers may not have always realised this, nor taken any notice of what their students have had to say about their performance as a teacher. This history has been marked by regular controversy, during which a series of persistent questions about the validity of student evaluations have been raised that have become the focus of research into their use. Without addressing these questions, teachers in schools will be left with doubts as to the usefulness and validity of the evaluations they collect from students.

Historical Context

From antiquity, teacher evaluations can be found in the writings of Plato and Xenophon who wrote of their admiration for their master Socrates after he was executed in 399BC for allegedly corrupting the minds of his students by his teachings. References in the modern teacher evaluation literature can be found as far back as 1896, when Kratz completed a study entitled Characteristics of the Best Teachers as Recognized by Children. In the first two decades of the twentieth century, rating scales became very popular, but these tended to be completed by administrators interested in making a judgement of the worth and efficiency of a teacher. As Kratz's title shows, they sought to find those traits that made a good teacher, and to determine the characteristics that made good or bad teachers. Also illustrative of this

is an article entitled Can students discriminate traits associated with success in teaching? (Stalnaker and Remmers, 1928)

Elliott (1915) produced one of the first teacher rating instruments which he described as a tentative scale for the measure of teaching efficiency. In 1924, a group of students at Harvard released a collection of course and teacher ratings so that other students could make some informed decisions when selecting courses in subsequent years. The "anti-calendars" of the 1960s and 1970s were a continuation of this trend. These publications were dismissed by university authorities as irrelevant and useless until 1965, when Cornell University conducted its own thorough review of teaching at the under-graduate level. It found that grossly inadequate teaching occurred far more frequently than was tolerable and the student dissatisfaction expressed in these "anti-calendars" had a basis in fact (Rayder, 1968: 77).

The University of Washington began to collect student evaluations on a campus-wide basis from 1925, and Purdue University and the University of Texas followed soon thereafter. In 1927, the literature of student ratings of teacher performance gained momentum with an article by Remmers and Brandenburg about experimental data from the Purdue Rating Scale, and Remmers followed with a series of articles from the late 1920s through to the early 1950s.

By 1951 Mueller found that 37% of US colleges used student evaluations and planned to continue using them, or were planning to use them. Gustad (1967) suggested a substantial decline in the frequency of use because of the lack of convincing validity data. This was not substantiated by another study a year later which showed that approximately half of the colleges were using student evaluations, and a further 13% had experimented with them but not persisted (Bryan, 1968). They are now widely used in tertiary institutions and have generated thousands of studies (Marsh, 1994), but the research on them is not always conclusive and consistent.

Arguments For and Against Student Evaluations

Arguments in favour of student evaluations are simple and appealing - the students see the teacher every day, and are ideally placed to comment on particular aspects of the teacher-student relationship and the learning environment in the classroom. Furthermore, if teachers are clear about the effect that they want to have on the

students whom they teach, then the students can provide exactly the kind of feedback that will determine whether that has occurred or not.

There are many supporters of student ratings in addition to Remmers and Weisbrodt quoted earlier. Guthrie (1949) observed that "students when called on to judge a teacher, have sat through many hours of his course". Fraser (1986) reflected on the Fifteen Thousand Hours that Rutter (1979) described as the lifetime of a school pupil and commented that

> Students therefore have a large stake in what happens to them at school, and students' reaction to and perceptions of their school experiences are significant (Fraser, 1986: 1).

On the first page of a review of the state of student ratings, one writer discussed teacher evaluation by noting

> the intrusion of a foreign body [evaluator] into the classroom would be necessary, and many teachers might feel that this would be bound to upset the delicate balance of personal interaction which was to be evaluated. An academic Heisenberg effect would interfere seriously with the measurement of this fundamental process. ... But do we need a foreign body? We already have one body of assessors formally present with the teacher: a body which is continually forming opinions on his performance. ... The students are inescapably there (Page, 1974: 1)

Another commentary compared the courts with universities and colleges by asking whether the courts would accept hearsay evidence (peer evaluations) as opposed to that of the eye-witnesses (the students) (Kulik and McKeachie, 1975).

Yet another observer took a very early stand for consumerism:

> The students are the consumers of teaching, and they know what they can and cannot consume, even if they are foggy about the reasons. ... It is much easier to fool one's colleagues than one's students ... Students are no fools (Cole, 1940)

As for one's colleagues, another commented:

Superintendents, supervisors, principals and colleagues tended to rate good teachers low and poor teachers high .....The only persons in the school system who were found to be professionally competent to judge the worth (as measured by gains in achievement) of teachers were their pupils (McCall and Krause, 1959: 73).

Similarly, there is no shortage of appeals which condemn the use of student evaluations. Hildebrand (1972) put forward 23 common criticisms of the use of student evaluations and answered each of them perceptively, but unfortunately without specific reference to the literature. When a system of student evaluations was proposed for Lincoln University, the Education Centre (Lincoln University, nd) was forced to publish a response to the myths about student ratings that had circulated amongst the faculty. In an eloquent piece of writing, Bryant (see Page, 1974: 25-27) fulminated against student ratings and all of the evils he could see in them, in particular the need for teachers to lower their sights and expectations of students and cater to the lowest common denominator in order to obtain good ratings from the students. In one comment pointed at these critics, it has been noted that teachers who condemn student ratings in their own institution seem to have no difficulty talking to others about the views of their own children about their teachers (Scriven, 1988).

Validity

Validity means that the student evaluation instrument measures what it sets out to measure, and does not in fact measure some other dimension. There are many concepts of validity - construct, content, face, internal and external, discriminant and convergent - and these, in turn, lead to different ways of estimating validity. Essentially, student rating validity can be explored and explained in two ways - one, if they reflect accurately students' opinions about the quality of the teaching they receive, regardless of whether the ratings reflect what students learn; or second, if they accurately reflect instructional effectiveness. Teachers interested in improving their teaching would want student evaluations to meet both of these criteria.

Do Student Evaluations accurately reflect student opinions?

Consider the first case - student evaluations are valid if they accurately reflect student opinions about the quality of the teaching the students receive. For the teacher, this means that the information that is provided by students is worth knowing for its own

sake because it is an accurate reflection of their views. Student evaluations enable us to reflect the beliefs of the raters but that does not test the validity of those beliefs. (Soar, Medley and Coker, 1983)

To assess the accuracy of the data obtained from student ratings, cross-validation studies using different methods of obtaining student opinions can be used. Ory, Braskamp and Pieper (1980) looked at three methods of obtaining data from students to determine the degree of congruence between the selected methods. They collected standard objective questionnaire items, written student comments to open-ended questions and conducted group interviews. The correlations across the three data sources were all significant and ranged from 0.81 to 0.94. In a similar later study Tiberius, Sackin and Cappe (1987) compared the information received from a student evaluation questionnaire with the results of a discussion with the same class. They found no difference in what the feedback actually covered, but noted that teachers preferred the feedback from the discussion method because of the 'depth' of the information they received.

Aubrecht, Hanna and Hoyt (1986) worked in secondary schools and used multitrait-multimethod (MTMM) matrix analysis and factor analytic techniques to compare student evaluations with teacher self-evaluations. Three factors emerged consistently from the analysis, and the corrected convergent validity coefficients almost perfectly matched those of Marsh (1975, 1979) with median figures between 0.45 and 0.49. In a variation on the usual validation design, Drews, Burroughs and Nokovich (1987) also used teacher self-ratings as the validity criterion for student evaluations. In this study, they took daily evaluations for 15 days from both the faculty and the students. The ratings were significantly correlated in all three areas under consideration: material covered, instructor performance and overall impressions of the success of the class.

In a consumer appraisal model, teachers who are interested in their relationship with a class want to know what the student (customer) thinks of that relationship with a view to making the process better for them (the student). However, this view of the usefulness of student evaluations merely tends to remove the "irritants" from the classroom environment, and does not necessarily lead to better teaching and learning (Ker, 1994). Using student evaluations to assess what the students feel about their learning experiences in the classroom is a measure of the processes of teaching and learning, without necessarily making any connection with the outcomes or products of that teaching and learning. Therefore the link between student evaluations and

learning outcomes is seen as being very important, and is used as a criterion for establishing whether teachers who receive high ratings are also teachers whose classes score highly on some measure of effective learning.

Do Student Evaluations accurately reflect instructional effectiveness?

The second view of student evaluation validity studies asks whether they accurately reflect effective teaching, and hence connect with the products of teaching. That is, do students do a good job of distinguishing between teachers on the basis of what they have learned?

This begs the question of what is meant by "effective teaching". This elusive question and the multitude of answers that educationists give to it make the task of evaluating effective teaching difficult as there is no consensus (Anderson and Burns, 1989). Instead one could legitimately ask whether students have a concept of effective teaching, especially as it this concept that will be reflected in the ratings they give their teachers. The answer to this seems to be clearer - yes, they do. Surveys of high school students using different techniques have found that students have a remarkably consistent concept of effective teaching (Batten, 1989, 1994; Clark, 1987; Wragg and Wood, 1984) just as college students do (Greenwood et al, 1973). In fact, Greenwood et al's findings indicate that if students do not know what effective or good teaching is, then teachers have even less of an idea!

Rating questionnaires usually consist of low inference items that reflect the nature of the implicit theories that students have. These implicit theories of teaching are brought to the task of evaluating teachers, and are powerful determinants of the way in which students differentially rate their teachers for each of the factors found in evaluation questionnaires (Whitely and Doyle, 1976). Some writers suggest that it is more realistic to expect a consistent and valid assessment of a teacher's performance from the students than can be obtained from other educators (Webb and Nolan, 1955; Centra, 1975; Blackburn and Clark, 1975; Doyle and Crichton, 1978). However, there is a very real concern that student ratings are becoming *de facto* criteria of effective teaching (Doyle and Whitely, 1974).

Multi-section Studies

The most common means of validation is to use student learning as the criterion for effective teaching and this is usually done by a multi-section validation study. In the

ideal multi-section study, a large number of students taking the same course would be randomly assigned to various sections (classes) being taught that course. The teaching programme, required reading and textbooks, assigned work and assessments, course objectives, and final examination would all be identical, with the final examination marked by an external examiner - in effect, the only difference between the sections would be the teacher. In addition, the sections would be pre-tested, and the findings used as a covariate. Multi-section designs are particularly strong in internal validity. In spite of this apparent strength, it has been noted that even using the final examination results as the criterion for student learning presents difficulties because students may well make up for the teacher's deficiencies through extra studying and tuition thereby washing out the variation in teaching quality (Cohen, 1983).

Centra (1977) conducted a classical multi-section validity study and found significant correlations with student learning on two factors, strong to moderate correlations on a further six factors, and weak or inconsistent correlations for two factors using the Student Instructional Report (SIR) rating scale. Frey et al (1975) conducted a study in three universities with students in two subject disciplines using the Endeavour Instructional Rating Form. Three factors (presentation and clarity, organisation and planning, and student accomplishment) showed a fairly strong positive correlation (0.51 to 0.59) with the final examination score, while one factor (students' perception of workload) was the only factor to have a negative correlation.

One common feature of the instruments used in these and other validity studies is the inclusion of overall instructor and/or course ratings. A comparison of the correlations between these general rating items and end of term examinations is instructive of the controversial nature of student ratings. Statistically significant but moderate correlations between instructor/course ratings and student achievement have been found by Braskamp, Caulley and Costin (1979), Centra (1977), Doyle and Whitely (1974), Marsh and Overall (1980), McKeachie et al (1971), and Sullivan and Skanes (1974). However, one multi-section study is frequently quoted as running counter to the pattern of these studies. Rodin and Rodin (1972) investigated eleven teaching assistants working in a tutorial situation with a large undergraduate calculus class. They reported a correlation of -0.75 between the mean student rating of the tutorial instructor and the mean examination score of the students in the different tutorial sections of the course, and claimed that students rated highest those teachers from whom they learned the least. This claim was startling but there are some serious methodological flaws in the design. A year later, a similar situation was investigated

with a more carefully controlled design, and this calculated correlations of 0.91 and 0.60 between the student achievement in the final examination and the factor for teacher presentation on the two courses studied (Frey, 1973).

Meta-analysis

Multi-section study results are not consistent enough for teachers to feel entirely comfortable about the validity of student evaluations. Meta-analysis (Glass, 1978) provides an opportunity to bring together the results to present a composite picture. Major meta-analyses have been conducted by Cohen (1981, 1982, 1983), Dowell and Neal (1982), Abrami (1984) McCallum (1984), and, Abrami, Cohen and d'Apollonia (1988). Cohen (1982) paints a compelling picture of the validity of student rating forms. In a meta-analysis of 16 studies covering 21 separate courses, he found an average correlation between student achievement and instructor ratings of 0.44, and when only full-time staff were studied the average correlation increased to 0.65. His 1981 study, which included the Rodin and Rodin study, found correlations between student achievement and student rating factors of Skill (0.50), Student Progress (0.47), Structure (0.47), Overall Course (0.47), and Overall Instructor (0.43).

Abrami, Cohen and d'Apollonia (1988) conducted an analysis of the six meta-analyses they could find on this subject, and found that even the meta-analyses could not agree. They found that the meta-analyses could only reach 47% agreement in calculating individual validity coefficients, with the disagreements in some cases quite large: 0.90 to 0.136 for one study and 0.87 to -0.19 for another. For the global instructor ratings Abrami et al calculated means of 0.43 (Cohen, 1981), 0.44 (Cohen, 1982) and 0.32 (McCallum).

Even though there are differences between the reviews as a result of the inclusion criteria used and the method used to calculate the study outcomes, these results reinforce the perception that there is a moderate (as defined operationally for the social sciences by J Cohen, 1977) and positive correlation between student ratings and the performance of the students in the end of course examination.

Generalisability

To talk of student ratings as generalisable means to consider the extent to which the sampling of student opinion can be done under any circumstance without regard to student, teacher or other characteristics. Studies have investigated the use of student

evaluation rating forms in a wide variety of contexts. Individual validation studies control the variables in the study, so it is not possible to find a single study which covers a range of teachers teaching a range of students in a range of subjects at a range of institutions in different countries.

Doyle and Whitely (1974) in their study of ratings taken from a beginners' French course concluded that

> student opinion about the general and specific course and instructor characteristics can apparently be sampled without attention to student characteristics like year in school, sex, and ability (Doyle and Whitely, 1974: 270).

Most of the specific ratings could also be sampled without regard to students' liking for the subject and attitude about the utility of the ratings. The Centra (1977) study covered 72 sections in seven different introductory courses, while the Frey et al study (1975) covered three universities and two different subjects. Sullivan and Skanes (1974) conducted their research in ten different first year courses. In a review of the research, Kulik and McKeachie (1975) conclude that some of the variability in teacher ratings is predictable from student characteristics but this effect is very small. They illustrate this by reference to Rayder's (1968) study which showed that variability resulting from student characteristics was less than 2% of the total variability in the ratings.

Do teacher characteristics have a significant influence on the student ratings? Some characteristics such as sex of the teacher, age, and rank of the teacher have been shown to have no or little influence (see Marsh 1984 for a discussion of this), while others do. A New Zealand study of the relationship between teacher personality as perceived by students and student ratings of teaching quality found that they are significantly correlated (Jones, 1989). This lends support to one of the most common objections to student ratings (that is, student evaluations are mere personality contests), but as Jones points out

> the activity of teaching is essentially one of human interaction, and as such it is arguable that it *should* be inextricably linked with the personal qualities which characterise the situation, if optimal use is to be made of it .... student ratings of teacher competence *do* depend upon their perceptions of teacher personality. This is valid [Original emphasis] (Jones, 1989: 558)

The applicability of university findings to the primary school level has been tested by Fox et al (1983). Teachers at sixth grade level were evaluated by their students for two consecutive years and observed by trained observers on multiple occasions. The class average ratings were stable over the two years (r=0.68) providing further evidence of the generalisability of student evaluations of teaching, and certainly counter the argument that school students are too immature and capricious to act as a source of teacher evaluation data. Validation studies have been conducted at university level in Australia, New Zealand, Papua New Guinea, Spain, India, Nepal, Nigeria, Philippines, and Hong Kong (see Marsh and Roche, 1994) using either the Students' Evaluation of Educational Quality (SEEQ) or Endeavour instruments that are widely used in the USA, and these reinforce their credibility in a cross cultural context.

Educational Seduction

The interesting case of Dr Fox and educational seduction has had good mileage in the literature. The original Dr Fox study (Naftulin, Ware and Donnelly, 1973) seemed to be a triumph of style over substance. A professional actor delivered a lecture to a group of educators and graduate students in an expressive and enthusiastic manner. Despite the content of the lecture being designed to have little substantive value, it was rated by the audience very favourably. It was concluded that an expressive lecturer or teacher could seduce favourable ratings from their students even if the content level of the lecture is low. However, the original Dr Fox lecture suffered from a number of methodological weaknesses, leading Ware and Williams to develop a new, standard Dr Fox design. In a series of experiments between 1975 and 1977, they studied this phenomenon, and in later reviews (1979, 1980) of those and other experiments concluded that teacher expressiveness does contribute to the variance in student ratings. Marsh and Ware (1982) in reanalysing the Dr Fox data found that a high degree of expressiveness only affected the rating for Instructor Enthusiasm, which is the factor most logically related to expressiveness, while the low content level affected Instructor Knowledge and Organisation/Clarity which are the two factors most logically related to content. As Jones (1989) said, "We should be concerned if it were otherwise."

Student ratings and the dimensionality of teaching.

The question arises as to whether it is possible to evaluate a teacher using a single question or scale that places the teacher somewhere on the scale between good and poor. A single item student rating questionnaire would certainly be very cost effective. On the other hand, do students view and rate their teachers differently for different aspects of their teaching, and if so what are these different dimensions?

To answer these questions, factor analysis is usually applied and a clear pattern has emerged. Kulik and McKeachie (1975) summarised nine previous studies and compared the factors that emerged with the factors found in another survey (Isaacson et al, 1964). Four factors were fairly consistent across all studies - Teacher Skill, Rapport, Structure and Overload. In perhaps the most persistent long-term study of student rating factors, Marsh (1982, 1983, 1984, 1987, 1991; Marsh and Hocevar, 1991) has identified nine factors in the Students' Evaluations of Educational Quality (SEEQ) questionnaire that are stable across a wide variety of conditions - Learning/Value, Instructor Enthusiasm, Organisation/Clarity, Individual Rapport, Group Interaction, Breadth of Coverage, Examinations/Grading, Assignments/Reading, and Workload/Difficulty.

Not all writers take this view however. Abrami and d'Apollonia (1991) acknowledge that teaching is a multi-dimensional activity but go further to retrieve a global factor from the analysis of student ratings. They claim that the research of Marsh and others is methodologically limited and that if these problems are ironed out a global rating factor useful for summative decision making about teaching can be found. In reply, Marsh concludes that

> while a few teachers may be consistently poor or consistently good across all specific SETE [Student Evaluation of Teaching Effectiveness] dimensions, most will have a systematic profile with some strengths and some weaknesses (Marsh,1991b: 421).

In a later article (Marsh and Hocevar, 1991a), Marsh suggests that a weighted average of the different components could be used to obtain a single rating score, where the weights could be arrived at as a function of logical and empirical analyses. Unfortunately, this work has yet to be done, and caution should be always be exercised when combining any data (Mehrens, 1990).

Reliability

Reliability refers to the extent to which the information is "free from biases due to sampling of students, courses and time of administration" (Braskamp et al, 1984: 42), and is generally thought of in two ways - agreement and stability. Agreement means the extent to which students in a class agree with each other, and also whether they are consistent with the ratings that are given by other student evaluation techniques - for instance, structured interviews with students about their teachers, or descriptive questionnaires. Alternatively, comparisons can be made with ratings performed by sources other than current students - ratings taken from alumni after graduation, administrator ratings, peer ratings, the ratings of trained external observers, or even teacher self-ratings provided caution is used with these.

Reliability of student evaluations can be calculated from the results of item analyses, and from studies of inter-rater agreement. There appears to be high internal consistency between items, with correlations reported by Costin et al (1971) between 0.69 and 0.85. In a New Zealand study at tertiary level, Watkins et al (1987) found that the reliability coefficients of the SEEQ and Endeavour scales ranged from 0.83 to 0.97. However, as Marsh (1984) notes, these high internal reliability figures need to be treated with caution as they may in fact give an inflated picture because they ignore the substantial portion of error resulting from a lack of agreement between different students. This is supported by Hepworth and Oviatt (1985).

Inter-rater reliability figures are also consistent depending on the number of students in the class. Costin et al indicate correlations of between 0.77 and 0.94 between two randomly paired students in the same class, while Marsh (1987) estimates the reliability for factors on his rating scale (SEEQ) at about 0.95 for the average response from a class of 50 students reducing to 0.60 for a class of five students.

Studies which investigate the relationship between student ratings and the ratings of other observers and raters have generally shown greater correlation between student ratings and one other source (for example between student ratings and colleague ratings) than they have between any two other sources (say, between colleague ratings and teacher self-ratings). Blackburn and Clark (1975) used a number of studies and found a correlation between student ratings and colleague ratings of 0.62, 0.78, and 0.69, while self-ratings and colleague ratings had correlations of 0.28 and

0.17. In each of the inter-correlations, those involving student ratings were the strongest.

As part of a study to determine the influence of the teacher or the course on student ratings, Hogan (1973) obtained correlations between three sets of data collected in three different ways. He looked at the relationship between two sections of the same course rating the same instructor, two different courses rating the same instructor, and two different instructors taking the same course over the four factors that Kulik and Kulik (1974) found to be fairly consistent in the literature. The results are shown in Table 2.1:

Table 2.1

Correlation between Teacher/Course and the Factors of

Skill, Rapport, Structure and Difficulty

Correlation/Factor	Skill	Rapport	Structure	Difficulty
A Same teacher, same course	0.67	0.74	0.36*	0.69
B Same teacher, different course	0.39	0.47	0.42	0.14
C Two teachers, same course	0.18	0.50	0.08	0.52

* Noted by the author as almost certainly an under-estimated value.

These results give a clear indication of the reliability of the student ratings when the same teacher takes the same course, and of the importance of the teacher in the actual rating. The size of the correlations for A indicate a substantial degree of agreement between students judging the same teacher teaching the same course. The size of C relative to A is a rough index of the importance of the teacher. If a course received similar high ratings irrespective of who teaches it, then the influence of the teacher is relatively minor. Comparing A and B gives a rough indication of the course influence.

Stability

A number of studies have been conducted to investigate the long term stability of student evaluations. Part of the reason for this interest is the frequently expressed view that students can be capricious, and do not have the maturity and experience to fairly judge the worth of a teacher or course until they have graduated and had the opportunity to put their learning into practice. There have been cross-sectional studies comparing the ratings of existing and past students of the same teacher (Marsh, 1977; Centra, 1977) which have shown good correlations between these two groups. In addition longitudinal studies (Overall and Marsh, 1980; Marsh and Hocevar, 1991b) have shown the long term stability of student evaluations. The first of these compared the ratings of the same students at the end of their course and again one year after graduation. There was a large and statistically significant correlation between the two ratings and very small differences between the two sets of data. The second study used data collected over thirteen years and after many statistical comparisons were made (including whether students ratings were changing over the years) concluded that the ratings on the nine content-specific dimensions were stable over that period. Using an entirely different approach, Kohlan (1973) had the students rate their teachers in the second and last hour of a course. There was a correlation of 0.58 between the two ratings, suggesting that students start to form a very definite impression of the teacher even in the first hour of a course and retain that throughout the course.

In summary, student evaluation ratings are stable over time. This is not entirely surprising for as Amidon and Flanders (1971: 72) suggest:

> As a result of participating in classroom activities, pupils soon develop shared expectations about how the teacher will act and what kind of person he is and feelings about how they like the class. These expectations and feelings color all aspects of classroom behaviour, creating a social atmosphere or climate that appears to be fairly constant, once established.

Utility of Student Evaluations.

There are three main uses for student evaluations - for personnel decisions (covering tenure, merit pay, promotion, and dismissal), for student course selection purposes, and as a source of information for teacher improvement strategies. The first two are

forms of summative evaluation - that is, they are used to make judgements about the teacher or the course. Summative student evaluations frequently meet with vociferous antagonism from teachers, in spite of the fact that student evaluations are typically skewed in favour of the teacher (Centra, 1973; Haller and Strike, 1986). The third purpose is for formative evaluation, designed to assist with improvement. The summative-formative dilemma in teacher evaluation, and the implications for teachers, run prominently through the teacher evaluation literature, especially in the student ratings literature. There is a strong argument for separating the two purposes (Darling-Hammond et al, 1983; Ker, 1994) while there are others who argue that one person can serve both masters (Edwards, 1992a).

This research project is formative. A number of ways have been successfully used to provide information from students for this purpose - teacher improvement. Apart from the common standardised student rating forms, researchers have used structured individual interviews, class group interviews (for example the Small Group Instructional Diagnosis technique introduced at Purdue University by Dr Mark Redmond in 1982, initially as a form of course evaluation) and open-ended questionnaires as ways of eliciting information from students.

In each of these approaches, feedback from the evaluations has been the operative part of the thrust for improvement. Feedback from student rating type evaluations has had a mixed degree of success depending on the feedback mechanism, the way that it has been used and supplemented, and the attitudes of the teachers and students to the form that the evaluation has taken. This has also impacted on whether there has been any positive change in teaching behaviours. Part Two of this chapter investigates the impact that student evaluation feedback has had in improving teaching, and will show that the best practice devised so far involves a face-to-face consultation to discuss the data collected.

Part Two: Using Student Evaluation Feedback for Teacher Improvement

There is a certain appeal to the belief that if a teacher knows how students feel about their teaching, then this will bring about changes and improvements in their classroom behaviours. This belief is not quite as straight forward as it might seem, and this can be partially explained when considering three variables which were theorised to impact on the effectiveness of feedback (McKeachie et al, 1980: 168).

Does the feedback convey information that the teacher does not know?

Does the feedback increase or decrease motivation?

Does the performer know what to do to improve?

Simple feedback research

Attempts to use the evaluation results alone as a means of feedback for improving teaching have met varying degrees of success. Two early studies illustrate this. Braunstein et al (1973) reported highly significant change in the median response from students for an experimental group of college professors. These professors received feedback in the form of results while a control group did not receive the results of the initial evaluation. As change was defined as a shift in the median in either direction of one scale point (on a 5 point Likert scale), these results are extremely remarkable. Miller (1971) investigated teaching assistants and found no significant difference between student ratings in a pre-test and post-test design after the teachers had (or had not) received feedback from the ratings. In fact, most subsequent studies agree with Miller's findings (Carter, 1974; Centra, 1973; Gage et al, 1963; Larsgaard, 1971; Pambookian, 1972, 1974; Ramsey, 1980; Scott, 1976; Siagin, 1983; Smith D, 1977) with only a few in accord with Braunstein (Smith W, 1975; Tuckman and Yates, 1980). Miller went further than this and also investigated whether the initial attitude of the instructors towards the value of student ratings had an effect. Again, no significant difference was found between instructors who were favourably or unfavourably disposed towards student ratings.

One of the earliest studies to investigate the connection between ratings and feedback demonstrated a vital point in the debate about the usefulness of student ratings, and has important implications for teachers and administrators concerning the source of information used in teacher evaluation. Tuckman and Oliver (1968) subjected 286 vocational teachers at high school or technical institute level to one of four conditions: feedback from students only; feedback from supervisors only; feedback from both students and supervisors; and, no feedback. Student feedback was found to lead to a (very small) positive change as measured by pre-test/post-test gain scores. The addition of supervisory feedback added nothing to this effect, but when taken alone supervisory feedback produced a change in a direction opposite to that of student feedback. In addition, the researchers found that less experienced teachers showed greater receptivity to student feedback than their more experienced colleagues while the reverse held true for receptivity to supervisor feedback. This study was a replication of an earlier study by Bryan (1963) and clearly shows that teachers react negatively to feedback from their supervisors whereas they do accept

feedback from their students. The implications are clear - student evaluations are a credible and useful source of information for teacher improvement, but the studies also point to one of the reasons why teachers do not readily accept summative evaluation and appraisal by their superiors.

Why does a method designed to improve teaching and which has such intuitive appeal not consistently produce the goods? This question has been addressed in a number of reviews (Rotem and Glasman, 1979; Levinson-Rose and Menges, 1981; Menges and Brinko, 1986; Menges, 1988; L'Hommedieu, Menges and Brinko, 1990), and one review concluded that

> feedback from student ratings (as elicited and presented to teachers in the studies reviewed) do not seem to be effective for the purpose of improving university teachers (Rotem and Glasman,1979: 507).

Levinson-Rose and Menges (1981) state that most of the studies in their analysis appear to support feedback intervention, but they temper this with the need for cautious interpretation. The findings from the meta-analysis of Menges and Brinko (1986) are not entirely encouraging regarding the ability of student rating information to act as a powerful stimulus to improved teaching. They investigated 30 studies involving feedback from student ratings, of which 10 found significant differences between groups (with all comparisons favouring the feedback group), one study had mixed results, and 19 studies found no significant differences. 23 of those 30 studies compared feedback from student ratings alone with a no feedback condition, and 70% of them found no significant difference. Their meta-analysis showed that feedback of this type produced a very small effect, raising subsequent ratings by one-fifth of a standard deviation. Expressed another way, after receiving student ratings feedback, the average instructor was rated "higher than 59 percent of control group teachers" (Menges and Brinko, 1986: 5). This may not seem like much of an improvement, but teachers in the classroom may see any improvement (no matter how small) as worth pursuing.

These analytical studies highlight certain problems in the research designs that seem to emerge. First, the paradigm that most researchers use is the pre-test/post-test control group design. Feedback is obtained from the pre-test, provided to the teachers in an experimental group (and withheld from the teachers in the control group), and the effects of the feedback measured by comparing the results of the post-test with the pre-test results. This design has the disadvantage that it cannot

control the reactive effect of the pre-test on the student raters - that is, it is not possible to tell whether the results on the post-test are directly the result of the feedback and any changes the teacher may have made, or whether the results on the post-test could have resulted from a knowledge of the items used in the questionnaire and the students reacting to them rather than any discernible changes made by the teacher. Secondly, most of the studies use an instrument that employs a Likert scale, usually with a five-point scale. This causes two difficulties. There is an artificial ceiling to the rating that any teacher can receive thereby distorting any possible measurable change that the scale can reflect, especially as the pre-test ratings are typically above the mid-point (Marsh and Roche, 1993) and leave little room for improvement. In the view of L'Hommedieu et al (1990), this is the single most important factor limiting the measurement of student rating effects. Furthermore, the very fact that the ratings are reliable in terms of their test-retest stability means that there are questions about the sensitivity of students to any change that may occur and their ability to record that change on a five-point scale (Rotem and Glasman, 1979). A third concern is the size of the samples used and the power of the statistical analysis to detect any significant effect. Cell sizes in the analyses have frequently been less than 20, which reduces the power of the analysis to detect change. At least 30 teachers in each treatment group is regarded as an acceptable minimum. One survey noted that the median cell size in all of the studies they analysed was 16 (L'Hommedieu et al, 1990) with some groups as small as 6. Fourth, the use of gain scores is fraught with difficulties as a measure of change primarily because the gains may be systematically related to any random error of measurement. Finally, the length of time between ratings may not be enough for any effect to show in the ratings. In some studies the interval between ratings was as little as four to six weeks - initial ratings were completed after about the first eight weeks of the semester (sufficient time for the students and teachers to become familiar with each other) with the re-test at the end of the semester about six weeks later.

In response to the three questions posed at the start of this section, simple feedback loops may provide information for the teacher which may motivate the teacher to improve, but the obvious disadvantage of simple feedback is that it does not provide the teacher with information about what they need to do to improve, and hence does not motivate them to improve.

Augmented feedback studies

If student ratings are to have a beneficial effect on the way that teachers behave, it is not sufficient for a teacher to passively receive ratings information - teachers need to be geared towards using the results. Four approaches to supplementing the information have been tried - providing a booklet of advice with the results, using a third party to facilitate discussion and consideration of the results (the consultation approach), discussing the results with the class, and a hybrid of the first two. Three studies have used booklets as the means of assisting the teacher to interpret and use the results (Toney, 1973; Cohen and Herr, 1979, 1982). This approach is suggested for use where institutions cannot afford the high cost of face-to-face consultation. In their 1979 study, Cohen and Herr developed a rating instrument (FACT) and a programmed approach using a booklet which lead the teacher through the student evaluation data to the stage of designing improvement strategies. They spell out some of the difficulties that emerge from the literature (teachers need enough time to improve; teachers need to know how to change; and global ratings are not enough to enable teachers to diagnose their teaching weaknesses), and argue that their approach overcomes these. Their 1982 study set out to compare this interactive method with a traditional feedback only condition and a control group who received no feedback at all. While the interactive feedback group consistently rated higher on the second evaluation than the other two groups, there was no significant difference between the two feedback groups. Augmenting the results of student evaluations by providing programmed booklets is not a necessary or sufficient condition for teachers to improve their teaching as measured by later student ratings. Using Toney's methodology (1973), the teacher's role was rather passive in dealing with the data. A manual was provided to assist the teacher to interpret the results, but no suggestions for planning an improvement strategy were incorporated in it. This approach was found to be more beneficial than merely receiving the results, but it stands alone as a methodology and is untested elsewhere.

One study focussed on completing the feedback loop as a way of supplementing student rating results. Friedlander (1978) was interested to know whether a discussion between the teacher and the class after the first evaluation would improve the ratings that the students gave the teacher in a subsequent survey. He conducted a student evaluation using an instrument which contained 17 specific rating items, 2 global rating items, and two open-ended questions that invited the students to (a) comment on positive and negative aspects of the course and (b) suggest methods to improve the course. Teachers were advised to discuss the results of the initial survey

with their class. In the subsequent survey, students rated the teachers who had a "meaningful discussion" with their class significantly higher than those teachers who had little or no discussion with their classes. The design however is flawed in that the teachers who had meaningful discussions with their classes were self-selecting, and the comparison groups were not randomly assigned. However this may indicate another factor in the usefulness of student ratings for teacher improvement - teachers motivated to use student rating information effect greater improvement than teachers who are not motivated to do so. In addition, there is no control for the possibility that the discussion influenced the subsequent results with no actual change in the teachers' classroom behaviours. As a means of augmenting information from student ratings, discussion with the class needs further investigation.

Feedback with consultation

The one methodology that seems to bear further consideration is to supplement the results with a face-to-face conference involving an educational consultant, and this research is based on this methodology. In one of the first experiments involving consultation conferences, Aleamoni (1974) as the researcher met with the teachers in the experimental group for a period of 15-20 minutes to discuss the data from the first student evaluation. This data consisted of course item means, institutional course item means and deciles, subscale means, and other normative data (categorised by rank of teacher, level of course, institution, and all institutions) for comparison purposes. The control group received all of the data, but did not participate in the conferences. The consultation conferences involved close scrutiny of the results to see if any problems existed, and

> if a source of difficulty was identified then the discussion shifted to possible ways of trying to resolve the difficulty. If, on the other hand, the source of difficulty was not identified ... then other procedures (such as the use of optional items, classroom visitation, video-taping, etc) were explored (Aleamoni, 1974: 5).

The results indicate that the experimental group significantly improved their student ratings on two out of five dimensions - Course Content, and Instructor - and Aleamoni argued that his inability to randomly assign teachers to the experimental and control groups may account for the moderate results on the other scales. A second study (Trank, 1968) investigated the use of a faculty adviser as a consultant. Teachers (and their classes) were assigned to one of three groups - the first

completed the student questionnaire at the end of term; the second and third groups both completed mid-term and end of term questionnaires. The second group of teachers received printed results from the mid-term questionnaire while the third group of teachers had the printed results plus a conference with the faculty adviser. The third group also had to discuss the results with their class. The results show that there was no significant difference on any of the 28 items in the questionnaire between the two feedback groups, nor between the two feedback groups and the control group. However, it should be noted that the significance level for this study was set at an extremely demanding $p < 0.002$.

A later study (Erickson and Erickson, 1979) looked at consultation as part of a faculty development programme, with the teachers meeting for about ten hours with the instructional consultant to review data and plan improvement strategies between the two administrations of the student questionnaire. This was a major time commitment on the part of the teacher, unmatched by any other descriptions in the literature. Their study was in two parts because the first part had an expectancy factor that could have been a confounding variable in the interpretation of the results. The instrument used (Teaching Analysis by Students - TABS) has three skill component groups and on all three components the results were significant for both the student ratings and the teacher self-ratings. In addition, the estimated ω (omega) square statistics indicate that the treatment (consultation) accounts for about 30% of the variance in the ratings, which is quite a considerable impact given the known sources of variance in student ratings (Kulik and McKeachie, 1975). However, the expectancy effects arise from the students in the experimental groups knowing that the purpose of the experiment was to improve teaching, and therefore they expected things to improve while the control students did not know this. The second study sought to overcome this expectancy by using students in subsequent years to rate the same teachers who were involved in the first study - that is whether the improvements that had been made would be apparent to subsequent students who were unaware of the previous purpose and teacher improvement project. For the twenty teachers involved, the differences between the two ratings were statistically significant for eleven teachers, and thirty of the thirty-five comparisons were more positive in the second rating than they had been in the first rating. As this second study was conducted from one to four semesters after those of the first study, these results indicate that the improvements are durable and relatively long-lived.

That same year, another study found that feedback from mid-term student ratings, coupled with a frank discussion of their implications with an external consultant,

resulted in more favourable ratings at the end of term ratings (on five out of seven factors), better final examination scores and more favourable affective outcomes (Overall and Marsh, 1979). The first outcome supports earlier findings, but the second and third outcomes are important. They suggest that feedback from student ratings used in this way can result in a

> climate more conducive to cognitive growth ... and enhance the supportiveness of the instructional process for learning (Overall and Marsh, 1979: 863).

In this study the consultant's role was to give a detailed outline of the data to the individual teacher and suggest possible approaches for improvement, but take a passive role with respect to the development of any improvement strategies, which were left to the individual teacher. The limited role of the consultant leaves room for further improvement in the process, and therefore in the outcomes, argue the authors. Further support for the use of student rating information together with consultation is provided by McKeachie et al (1980) in considering subsequent student ratings and affective outcomes.

In a ten year follow-up on a group of teachers who had been involved in a prior study which used student rating feedback with expert consultation, Stevens and Aleamoni (1984) found that these teachers received higher ratings throughout that ten year period, and also used student ratings and instructional services more frequently during the follow-up period than did teachers who had not been involved. As this was a follow-up study, experimental control over the assignment of teachers was not possible and the authors advise caution in interpreting their substantive conclusions. What is suggested by the results is that the teachers who had been exposed to a general strategy for effecting instructional change now had the knowledge and ability to analyse feedback information, and the confidence to seek support when needed.

The most comprehensive study of student rating feedback plus consultation for teacher development was conducted by Wilson (1986). It adds an extra phase to the consultation which picks up on a variant of the work of Cohen and Herr described above. Wilson, working at Berkeley, wanted the consultation phase to tap into the best practices of the best teachers so that other teachers could apply these practices to specific aspects of their teaching, even though they could not personally consult the 'best' teachers. Therefore, the first stage was to get ideas from excellent teachers. Graduating students were asked to name one faculty member whom they regarded as

the best teacher they had, and also to name the teacher from whom they had learned the most during their years at Berkeley. Thirty nine of these excellent teachers were interviewed with the interview focussed on the particular questionnaire items that students rated as most descriptive of that teacher's behaviour. These teachers were asked to think of anything that they did that would lead students to describe them that way - for example, encourages class discussion, or explains clearly. The ideas generated from this discussion were compiled in a "idea" notebook. This idea notebook was used as a source of suggestions in the consultation phase.

The teaching improvement process used was quite elaborate, and incorporated nine steps designed to motivate teachers to consider ways they can improve. The first step was the initial student assessment and teacher self-assessment. The results of this were posted to the teacher in the form of a computer printout and a preliminary feedback session arranged. The second step was the planned preliminary session of about half an hour at which the teachers were asked to identify the questionnaire items with which they would like help. The third step involved the consultant choosing the three highest scoring items, and between three and five of the lowest scoring items. The open-ended responses were also studied to find any information concerning strengths and weaknesses. The consultant would then examine the ideas notebook to sort out appropriate suggestions (for the items where the teacher wanted help) to make at the main consultation. The main consultation, the fourth step, took place shortly before the teacher taught that course again, sometimes up to a year later. At this meeting, to start the consultation on a positive note, the teacher was asked to contribute ideas for the booklet for their three highest scoring items. Ideas that seemed useful were added to the book. The consultation then moved on to the items identified for action. The suggestions that had been sorted out were discussed, and the three that the teacher felt able to implement were explored in more detail, with photocopies of the relevant page from the ideas book made available. The next three stages involved a follow-up letter which summarised the consultation, friendly follow-up telephone calls with offers of assistance if required, and additional consultations or observations or even further student evaluations. For most teachers, the next phase was the last - the second student evaluation to provide post-test data followed by a meeting to discuss the changes that had been made by the teachers and observed by the students. A few teachers continued the cycle as a final step in the process. This process clearly fits the collaborative model described by Brinko (1990).

Of the ten items that were suggested most frequently to teachers, six had in excess of 50% of the teachers show a statistically important positive change. One item had

90% of the teachers change in a positive direction, while the smallest proportion of staff to positively change on any given item was 39%. On the overall teacher effectiveness item, twenty-four of the forty-six classes (52%) showed statistically important changes.

Wilson used tests of statistical importance on changes in the mean ratings between the first and second evaluations. Statistical importance is defined as a shift of one-tenth of a standard deviation in either direction, a definition that cannot be found in the quoted reference (Cohen, 1977). Indeed, this shift seems to be rather small and within the standard error of measurement for items on a five-point scale. There is not enough data for the reader to reconstruct an acceptable statistical test of significance.

While this lapse is unfortunate, the methodology has been adopted and strengthened by Marsh and Roche (1993) who provided a cogent comment on the weaknesses of the Wilson design and interpretation. This study used a similar intervention for two groups, one of which received feedback and consultation mid semester (MT) and the other receiving their feedback and consultation at the end of the semester (ET). A control group received no feedback. For both the ET and MT groups, targeted items improved more than non-targeted items, and for all groups the ratings improved over time. Interestingly, only the ET group improved significantly more than the control group, suggesting that the consultation and suggested improvements take time to bed in to the teacher's repertoire sufficiently for the students to register this as an improvement. This study was part of a larger project undertaken to improve teaching at the university level in Australia, and the complete procedure including the "ideas notebook" were published for general use (Marsh and Roche, 1994).

The line of research involving feedback coupled with consultation seems to have merit, and meta-analysis confirms this conclusion. Menges and Brinko (1986) investigated five studies (all earlier than the Wilson consultation design) that coupled consultation with student rating information, and found that four of the five studies reported significant differences favouring the group that had feedback augmented with consultation. On average:

> feedback of this type raised the subsequent ratings by more than one standard deviation ... the average instructor who received student ratings feedback and consultation was subsequently rated higher than 86 percent of teachers in the control group (Menges and Brinko, 1986: 5).

This is an impressive size effect, and must lead teachers to give very serious consideration to this type of intervention in teacher development processes.

Summary

At the start of this chapter, some assumptions were made regarding the use of student evaluations for teacher development. The first of these was a critical one in terms of answering the concerns of teachers who could feel threatened by this form of evaluation. Are student evaluations a valid, reliable, stable, and useful means of gathering data about what happens in classrooms? While the literature is almost exclusively tertiary based, it points to the appropriateness of this assumption. What little research there is in secondary schools tends to mirror the findings from the tertiary sector. This review has closely examined the research on validity, reliability, stability, and usefulness and found that student evaluations match, if not better, other forms of teacher evaluation. Those who would argue against student ratings and in favour of some other means of teacher evaluation need to apply the same standards to their preferred method. The evidence at present suggests that their preference will be found wanting.

The literature also shows that students do have a clear picture of effective teaching and can therefore be called on to provide data for teachers in their quest for better teaching. Student rating evaluations are the most common means of generating that data, and this data has been effectively used by teachers to further that quest. It is clear that simply providing the teacher with this data as feedback has minimal positive effect. Of the various feedback methods, feedback with consultation provides the most effective means of using student evaluations to improve teaching.

Chapter 3 Methodology

The purpose of the research was to determine whether student evaluations could be used by teachers to help them improve their teaching, and to evaluate the effectiveness of this process in the view of both the student and teacher participants.

The main research question was

> Are student evaluations of teachers coupled with a consultation conference to develop an action plan for change, an effective and useful means of bringing about teacher development, particularly improvement?

Subsidiary research questions were:

> Do teachers accept student evaluations as a valid, reliable, credible and useful means of gathering data for development, and are they therefore prepared to act upon them?
>
> Do teachers see this form of evaluation as part of a developmental (formative) appraisal process, as part of a performance (summative) appraisal process, or as part of both? How best can student evaluations be incorporated in an appraisal system, if at all?
>
> Are the time requirements realistic and manageable for teachers in gathering and analysing data, and developing and implementing an action plan for improvement?
>
> What can we learn about the way that secondary students view their teachers and their learning experiences?

The research methodology involved:

> preparing a student evaluation of teaching questionnaire,
> administering the questionnaire twice - a simple pre-test/post-test design,
> obtaining a teacher self evaluation at the time of the student pre-test,
> conducting a consultation conference with the teacher to discuss the results and write a teacher development action plan,

obtaining written teacher evaluation of the process at the time of the student post-test,

obtaining a written student evaluation of the process at the time of the post-test.

This methodology draws on many of the practices described in the literature on student evaluations coupled with consultation, and in particular the designs of Wilson (1986) and Marsh and Roche (1994). The use of an ideas booklet was not possible to implement as there was insufficient time to fully develop this as described by Marsh and Roche. Instead, the consultations had to draw on the resources of the teacher concerned and on other people available to them. Another important deviation from the usual design for this type of research was the absence of a control group. Details for this are outlined in the next section, Procedural Overview.

This design was intended to obtain data from students in the pre-test (the Time One administration), which could then be communicated to the teachers. The data were then used at a teacher development consultation conference during which an action plan was written to address areas that the teacher wanted to improve. The teacher then took the responsibility of implementing the recommendations in the plan. The post-test (the Time Two administration) was designed to determine the extent to which the action plan had been effective in bringing about change and, in particular, improvement.

The post-test process evaluations from both the students and the teachers were to allow the participants to comment on any aspect of the student evaluation process they experienced, and provide an added dimension to the research which would enable general conclusions to be made about the possible effectiveness of this method of data collection and analysis, and their usefulness for teacher development programmes in secondary schools.

Procedural Overview.

A letter was written to the Principal of all secondary schools in a major metropolitan area seeking permission to conduct the research in their school. Using proportional sampling, ten schools were selected to participate so that the sample covered all types of secondary school in the region. Staff from each school were asked to volunteer, and from the list of volunteers two staff were randomly selected to participate. At this stage, one of the schools was unable to find two volunteers, and efforts to replace

that school with another from the same strata were unsuccessful. Therefore, only nine schools were in the final sample. Also, at this time, one of the selected staff withdrew leaving seventeen teachers for whom the first (Time One) questionnaire was completed. One of these seventeen teachers withdrew just prior to the second (Time Two) administration of the questionnaire, so that a total of nine schools and sixteen teachers participated in all aspects of the research.

The researcher met with the two teachers in each school and fully described the project, the time requirements, their part in it and covered all of the Human Ethics matters relating to the conduct of the research. They also received written details of the project. Once informed consent was obtained, one class was selected from their timetabled classes. The teachers were able to nominate one class that they did not wish to involve in the project, and one class was then randomly selected from their remaining classes. The researcher visited the selected class and outlined the project and their role in it. A letter outlining these details was issued, and informed consent was obtained. In the case of students under sixteen years of age, written parental consent was also obtained. Students were guaranteed confidentiality - they were told that the teacher would receive only response frequencies and the class mean for each item. Open-ended question responses would be typed up verbatim before the teacher received them. The students were also told that the purpose of the research was to provide the teacher with information that would be used to help them to work on improving their teaching.

The project used a simple pre-test/post-test design with a treatment (feedback of results plus consultation) in between. This design is not particularly strong in controlling for threats to experimental validity, and is not, therefore, a true or quasi-experimental design. To strengthen the design from a statistical standpoint, a control group would be required to complete the pre-test and post-test but not receive the treatment. This option was rejected because all of the teachers were volunteers and it was felt that it was unethical for any of them to be denied the purported benefits of the treatment. As an alternative method of strengthening the design, a post-test only group could be used to overcome the sensitisation of students to the questionnaire and treatment that a simple pre-test/post-test design is susceptible to. This too was not possible, but there were a number of students in each class who did not complete the first questionnaire but did complete the second administration of it. In all, 37 students were in this category and consideration was given to using this group to provide some measure of control over the sensitisation issue. As a control group, these students would have to be treated with considerable caution. They are a self-

selecting group, based on their attendance (in fact, their non-attendance) at the Time One Administration. They are not a randomly selected control group, and indeed the reason for their participation in only the second administration could make them a very biased group in their views on school, teachers and learning. Therefore, the decision was made to reject this group as a post-test only group.

These statistical methods of controlling the threats to experimental validity can be partially overcome by collecting both qualitative and quantitative data, and comparing them. This was done at each stage of the research, from both sets of participants. In this way, an experimental design that would otherwise be weak in explaining the outcomes of the research can be strengthened, and the results can be used with more confidence. By combining the collection of both qualitative and quantitative data at each stage, it was possible to conduct the research in a more naturalistic setting than is often the case for true experimental designs. As Cross (1987: 499) notes:

> In the search for general laws of learning, researchers deliberately hold constant or rule out the specific conditions in any particular classroom. But what each classroom teacher wants to know is, What is happening in *my* classroom, given my students and my subject matter? [Original emphasis]

The Time One student ratings were completed in the first half of Term Two (June and July). A set of standardised instructions were read aloud to the students at the start of the administration. Each administration was timed. Times were noted when the first person finished, the tenth person finished, and when all of the students had finished. Any students who did not have the required consent were supervised by their teacher in another room. During this time the teacher completed a self-rating using the same form as the students.

The results were processed by the researcher using a spreadsheet, and tabulated. The teacher received a printout containing the verbatim responses of the students to the open-ended questions, plus two versions of the ratings data. Both versions consisted of the response frequencies, the mean, the teacher's self-rating for each rating item, and the difference between the mean and self-rating . The first version gave these details in the order that the items appeared on the questionnaire. The second version was presented with the items grouped together according to factors corresponding to a modification of Marsh's Students' Evaluation of Educational Quality (SEEQ) analysis. This *a priori* grouping was used as some teachers received their results before all of the classes had completed the Time One administrations, so that there

was insufficient data for the questionnaire to have been factor analysed at that stage. In addition, the response frequencies for each item were graphed and included with the printouts. Similar printouts were provided for subsets of the class (breakdowns by gender, ethnicity, or ability) where the number of students in the subset was greater than five. These printouts were posted to the teacher within a week of the questionnaire administration, together with a guide on reading and interpreting the printouts. The printouts contained a shorthand descriptor which was used for each item in the questionnaire (for example, 01 Pride stands for Item 01 "I am usually proud of the work I do in this class"), and these descriptors are used throughout this report. In addition, negatively worded items are indicated by an asterisk. (Examples of a printout and the instructions are attached in Appendix 3.1)

The consultation conferences were scheduled within the next week. The researcher took the role of the consultant, and the meetings took place in the teacher's school at a time that suited them. The conferences typically took one and a half hours. The teachers were asked to prepare for the conference by marking their first version printout to indicate three things: items they were pleased with; items that they felt required explanation for any reason but did not require any action; and items that they were disappointed with or concerned about, and from which they would target three or four items in their desire to improve.

At the start of the consultation conference, discussion focussed on the results on the first version printout. As the consultation progressed, information noted on the first version was transferred to the second version printout which had been organised in to factor groups. This was designed to draw the teacher's attention to items which taken separately may have little meaning, but when collected by commonality indicate a pattern of behaviour or student response.

At the end of the consultation an action plan was drawn up for the teacher to follow. The teacher provided the ideas for change as the teacher was regarded as the expert on their subject and their students, not the researcher. Therefore all of the suggestions for change were actions that the teacher felt it was possible to implement in their classrooms, and were not imposed. The action plan included a timetable for achieving each action.

The conferences were audio-taped and the researcher also made notes. These notes consisted of the teacher's general impression of the results, the items that they were pleased with, the comments that they made, the items that they were displeased with,

the items that they wished to focus on for improvement, the action plan and the timetable. These notes were typed up immediately after the conference and posted to the teacher as a record of the conference together with a letter of encouragement to see the action plan through. The audio-tapes were used to supplement the notes, then archived for reference.

The individual teachers took the responsibility for implementing the action plan.

The second (Time Two) survey was to provide information on the effect of the intervention by comparing the gain scores of the teachers on targeted items and related items within factors. The Time Two student responses were processed and printouts produced as before, along with some statistics comparing the class responses on the two administrations, and posted to the teacher for reference. The teachers also received the response frequencies of the students to the additional eight statements regarding student participation in the research. During the Time Two student survey, the teachers completed a comprehensive questionnaire covering all aspects of the research project, their involvement in it, and their views about this type of evaluation and its place within an overall appraisal system. They did not complete the student questionnaire on this occasion. The Time Two survey was conducted an average of 11 weeks after the Time One survey, with a minimum of eight weeks occurring between the two administrations and a maximum of 14 weeks.

Human Ethics considerations.

The research involved the delicate relationship between teachers and students, and there was always the potential for difficulties to arise between these two parties as the result of the students being asked for their opinion about the teaching behaviours of their teachers. Therefore, an application for Human Ethics committee approval was submitted in February to ensure that the research met the requirements of the University and contained the protections necessary for all parties. Approval was obtained from the committee at the start of April.

All participants (teachers and students) received written details of the research, an oral briefing and had the opportunity to ask questions before giving their written consent. Students aged 16 years and under also had to obtain their parent/caregiver's written consent.

In particular, all participants were informed that they could withdraw from the research at any time without having to offer any reason and without penalty of any sort. This option was exercised by two of the teachers, and the final number of teachers involved was sixteen.

Student confidentiality was maintained in two ways. For the 52 rating items, the teacher received a listing of the number of students in the class who gave a particular rating plus some relevant descriptive statistics. For the two free response items, the students comments were typed up verbatim and presented to the teacher in that form. To further ensure confidentiality, ratings data from the Sex, Ethnic and Ability demographic groups were not supplied to the teacher unless there were at least five students in the subset.

For the consultation conference, teachers had the option of inviting a third party to act as a support person. This option was made available so that the researcher did not have to act as a counsellor should any unintended consequences arise as a result of the information received. None of the teachers took up this option.

The Participants:
The Schools

In an ideal research situation, the selection of schools, and teachers and classes within those schools would be truly random. In this way, the research methodology would provide acceptable controls against problems of bias and of validity, and enable the results to be generalised across all schools in the population. In practice, this ideal is not practicable for a variety of reasons: for example, some class sizes are too small for practical research; students may have already been assigned to classes for timetable or other reasons; some teachers already have supervision of their teaching because they are beginning teachers or subject to competency procedures; lack of consent on the part of the teacher or student or school.

All 75 secondary schools in the metropolitan area were approached early in the 1995 school year. In a letter to the Principal, the research project was outlined and the school invited to participate. Written replies were received from 46 schools, 28 agreeing to participate and 18 declining. The remaining schools were contacted by telephone and replies received from a further 16, of which 10 agreed to participate. The last 13 schools did not respond despite two further attempts.

To ensure that the schools selected would represent a cross section of all types of secondary school, the selection was done on a proportional basis, using the following categories: state co-educational, state boys, state girls, integrated, independent boys, and independent girls. Ten schools were selected, with five from the first category and one from each of the other five categories.

Table 3.1

Categories of All Schools and the Selected Schools in the Region

Category	Number in Region	Regional	Agreeable	Number Selected	Selected from N2	Selected from N1
	N1	%	N2	N3	%	%
State co-educational	42	56	23	5	22	12
State boys	4	5.3	3	1*	33	25
State girls	4	5.3	2	1	50	25
Integrated	18	24	3	1	33	6
Independent boys	3	4	3	1	33	33
Independent girls	3	4	3	1	33	33
Independent co-ed	1	1.3	1	0	0	0
Totals	75	99.9	38	10		

The selected state boys school (School H) withdrew soon after being selected, and a second school was randomly selected from the two remaining schools in the category. This school was also unable to proceed, and by that stage it was too late to commence the project with the one remaining school and have a suitable time-frame to obtain meaningful results. Therefore, the actual sample contained just nine schools.

In 1994, the Ministry of Education provided a socio-economic status (SES) decile for each state school, and this has been used in the descriptions below. To obtain the decile rating, the Ministry took a random sample of students from the school files, and compared their home address with 1991 Census mesh block data. These deciles were used by the Ministry to determine each school's eligibility for Targeted Funding for Educational Achievement (TFEA, formerly known as Equity Funding). SES deciles of 10 indicate a very high proportion of children attending the school come from high income/professional families. SES deciles of 1 indicate that the school has the greatest proportion of its students from families with the lowest income levels. The Ministry does not collect or provide SES ratings for independent schools as these schools do not receive TFEA from the state coffers. It would be reasonable to presume that the independent girls school (School J) would have an SES decile of 10, but no assumptions could reasonably be made about the independent boys school (School I) because of the unique enrolment policy of that school.

School A

State co-educational, Forms 3-7. The school draws on a broad cross section of the local area, and has an SES decile of 6. A high proportion of senior students go on to tertiary education.

School B

State co-educational, Forms 3-7. This school is one of the four Technology Development schools, has a Special Education Unit attached as well as a bi-lingual unit. A child care centre and kohanga reo are also on site. The school is a multicultural community school with an SES decile of 2.

School C

State co-educational, Forms 3-7. This school regards itself as a high achieving suburban secondary school with a strong emphasis on guidance systems. The 1994 SES decile for the school was 10.

School D

State girls school, Forms 3-7. This is an inner city girls school with over a century of tradition, and considerable ethnic diversity. There is a day care centre on campus for the staff and the community. In 1994, the SES decile for this school was 3.

School E

State co-educational, Forms 3-7. There are three attached units - Special Needs Unit, Experience Unit, and an Activity Centre. A multicultural school, founded in 1931, the school offers an extensive range of courses to students. This school has an SES decile of 1.

School F

State co-educational, Forms 3-7. An IHC Unit and three Experience Units are attached to this school. The SES decile is 3.

School G

Integrated co-educational school, Forms 1-7. The school is a Catholic school, with a multicultural student body. The SES decile for 1994 was 3.

School H

State boys school. This school withdrew from the project and was not able to be replaced.

School I

Independent boys school, Standard 3 - Form 7. This school has just celebrated its centenary. It is funded entirely from an endowment trust, and all students are boarders. The school has two campuses, one for the primary school and another for the secondary school.

School J

Independent girls school, New Entrants-Form 7. The school also has attached hostel accommodation for girls from ages 13-18.

Table 3.2 shows the 1995 roll numbers, gender and ethnic percentages, and the number of teaching staff in each of the selected schools.

Table 3.2

Roll Numbers, Student Characteristics and Number of Teaching Staff in Selected Schools

School	Roll	Girls	Boys	European	Maori	Pacific Island	Asian	Other	Teaching Staff
	N	%	%	%	%	%	%	%	N
A	709	48	52	64	12	4	20	0	43
B	1099	48	52	21	24	35	15	5	65
C	1728	54	46	67	2	1	30	0	100
D	1144	100	0	23	18	40	16	3	65
E	1304	49	51	12	20	55	13	0	82
F	1968	52	48	60	7	18	15	0	115
G	487	44	56	40	8	36	15	1	28
I	445	0	100	85	12	2	1	0	31
J	1201	100	0	90	1	2	7	0	83

Table 3.3 shows the breakdown of the school by type.

Table 3.3

Selected Schools by Type.

	Coeducational	Boys	Girls	Total
State	5	0*	1	6
Integrated	1	0	0	1
Independent	0	1	1	2
Total	6	1	2	9

*   The school(s) selected for this category withdrew from the research as noted above.

The Teachers.

In ideal conditions, the selection of teachers to participate would have involved making a random selection from a complete staff list. Under the Human Ethics requirements for this research, all participating teachers would have to give informed consent. Furthermore, this research was going to involve teachers in a process that could be threatening for some who would therefore prefer not to participate. This point was reinforced in the comments of some of the students in their evaluation at the end of the project in which they could clearly see that some of their teachers would not be brave enough to participate in research involving students evaluations:

> I also wondered about the fact that only teachers confident about their ability as teachers would volunteer - I know a couple of my other teachers would not be too happy to have evaluations done on them and they would never volunteer! (J20)

> Perhaps teachers should be selected and not asked to do this as questionable teachers probably won't consent (J20)

Consequently, the process of selecting the teachers was not truly random.

To select the teachers, a contact person in each school briefly outlined the nature of the research to the staff and asked for teachers to indicate whether they would be willing to participate. Beginning teachers (that is, teachers in their first two years of teaching) were not to be included as they have a special induction programme in their

schools leading to teacher registration. From the list of willing teachers that was submitted, two staff were randomly chosen. Most schools submitted lists with from two names (one school) to ten names (three schools) to choose from. Reasons for volunteering can be summarised as:

a desire to improve teaching effectiveness (B03, C06, D07, D08, E09, E10, F11, F12, G14, I17, J19)

to satisfy curiosity about what the students thought of the teacher (A01, C05, F11, F12)

because I believe in students having a say in how I teach them (B03, C06, D08, E10, F12, G14)

to satisfy my curiosity about the project (A01, A02, E09, J20)

to compare this research with a similar process I was using (D08)

to monitor a pilot programme I had set up (D08)

to support a good cause (B03, D08, I17)

because of an overall interest in this form of appraisal (B03, J20)

to see how it could fit into overall professional development (J19)

It was at this stage that one school was unable to find the minimum of two staff who would participate. A second school was selected from the same category, and it too was unable to find two volunteers. This resulted in this category (state boys' schools) being deleted from the research as there was insufficient time left to include another school.

In the following brief description of each teacher, the letter corresponds with the school coding above, and the two numerals are simply a code number to ensure confidentiality. The school codes and the teacher codes were randomly allocated.

Teacher A01

Male with twelve years teaching experience. He is an assistant teacher of English and Form 7 Dean. This year he teaches English (Forms 3-6) and in the past has taught English to Form 7, as well as Social Studies and Physical Education to junior classes. He is a member of the Professional Development Committee and also runs the Student Council in his school. Selected class is Form 4 English.

Teacher A02

Female with eighteen years teaching experience. She is an assistant teacher of Mathematics and PR1 Dean of Form 3. This year, she is teaching Mathematics to

students in Forms 3 to 7, and has taught Science (Forms 3-5), Biology (Form 5-7), Physics (Form 5) and Chemistry (Form 5). The selected class is Form 5 Mathematics.

Teacher B03

Male who has taught for 19 years. He is the Head of the Social Sciences department and this year teaches Geography (Levels 2 and 3) and Social Studies (Years 10 and 11). In the past, he has also taught Year 11 History and Year 9 Social Studies. The class selected for the research project is his Year 10 (Form 4) Social Studies.

Teacher B04

This teacher withdrew from the research at the start of Term Three prior to the Time Two student survey.

Teacher C05

Female with twenty seven years teaching experience. This year she is teaching English to Forms 3 to 7, and in the past has taught Social Studies, Science (up to Form 5) and Forms 5 and 6 History. This year, she is implementing a polytechnic course for the alternative Form 7 English class. Her selected class is Form 6 English.

Teacher C06

Female with thirteen years experience as a teacher, most of it in South Africa where she had some experience of student evaluations in a previous school. She is an assistant teacher who teaches English (Form 5), Drama (Form 4 and 5) and Social Studies (Forms 3 and 4) this year. Has had experience in teaching Sociology to Form 6 and all levels of English. The class selected for this research is her Form 4 Drama class.

Teacher D07

Female with eighteen years teaching experience. She is the Tertiary Studies Co-ordinator in her school and this year teaches two classes in computer applications (Forms 6 and 7). These classes rotate each term, so the research was restricted to just one term with this teacher/class combination. She has experience in teaching Accounting (Forms 3 to 6) and Economics (Forms 3 to 5). The selected class is Year 12 Database which includes students from Forms 5, 6 and 7.

Teacher D08

Female with eleven years teaching experience. She holds the position of PR1 Integrated Studies and this year teaches English (Forms 4 and 7), Integrated English

and Social Studies (Form 3) and Drama (Forms 6 and 7). Teacher D08 has previous experience teaching English to all levels and Drama to classes from Forms 4 to 7. The class selected is her Form 4 English class.

Teacher E09

Male with three years teaching experience. Currently he holds a Long Term Relieving assistant teacher position, teaching Science (Forms 3-5) and Form 6 Landscape. He has taught Mathematics and Science to all levels from Form 1 to 5, Form 4 Social Studies, Form 5 Biology, Human Biology, Agriculture and Horticulture, Form 6 Chemistry and Physics, as well as Form 6 and 7 Accounting and Economics. A Form 4 Science class was selected for this project.

Teacher E10

Female who has taught for twenty years. This year Teacher E10 teaches Geography (Forms 5-7), and Social Studies (Forms 3 and 4). She has previously taught English (Forms 3-7) and Transition (Forms 5-7). The selected class is her Form 6 Geography group.

Teacher F11

Female with six years teaching experience. Teacher F11 is the Dean of Form 5 and Teacher in charge of Form 5 Mathematics. This year she teaches Mathematics with Calculus (Form 7), Form 5 Mathematics and Form 5 English. In previous years, she has taught all levels of both English and Mathematics as well as junior classes for Health Education. The selected class is Form 5 Mathematics.

Teacher F12

Male who has taught for seventeen years after many years in industrial science. He retired from teaching at the end of the 1995 school year. An assistant teacher, he currently teaches 3 classes of Form 5 Science, as well as one class each of Form 3 and 4 Science. He has taught Mathematics to Form 5, Form 6 Chemistry, Automotive Engineering (Form 6) and Metalwork and Technical Drawing to Form 3. The class selected is one of his Form 5 Science classes.

Teacher G13

This teacher withdrew from the research before the administration of the Time One questionnaire.

Teacher G14

Male with fourteen years teaching experience. He is a Long Term Relieving assistant teacher of English. This year he is teaching English (Forms 6-7), and Reading/Extra Language (Forms 3-4), as well as having responsibility for the Library. Has taught Form 3 and 4 French, and been a Lecturer at the Catholic Teachers' College for eight years. The selected class is Form 7 English.

School H never entered the research.

Teacher I17

Male with fourteen years teaching. He is the Head of the Social Studies department, and a House Master in one of the boarding hostels. This year he teaches History (Forms 5-7) and Forms 3 and 4 Social Studies. He has taught Form 5 Geography. His Form 3 Social Studies class was selected for this project.

Teacher I18

Male who has had six years teaching experience. He is the Dean of Form Four, and a House Master of one of the boarding houses. This year, he is teaching Mathematics to Forms 3 and 4, and Classical Studies to Forms 6 and 7. He has previous teaching experience in Computer Studies. His selected class is Form 4 Mathematics, and they are studying for the School Certificate Mathematics examination.

Teacher J19

Female who started teaching twenty one years ago with spells of part-time teaching when her children were born. She has been teaching full-time for the past seven years. She holds the position of Head of the Languages department, and teaches Forms 3-6 Latin, and Form 6 Classical Studies. She has had experience teaching German (Forms 1-7), French (Forms 3 and 4), and Form 7 Classical Studies. Selected class is her Form 5 Latin class.

Teacher J20

Female with eight years teaching experience. Now Head of Department Geography and Social Studies, this year she teaches Geography (Forms 5-7) and Form 4 Social Studies. She has also had experience teaching History. Her Form 7 Geography class was selected for participation in this project.

To summarise, the teachers in the study have taught for an average of fourteen years (with a range from three to twenty-seven years), have taught a wide range of subjects

in the school curriculum, and ten of the sixteen currently hold positions of responsibility (either Head of Department or Dean) in their schools. They were experienced teachers.

The Classes

Again, the selection of classes was not done in a truly random manner which would have provided for some control of variables. To select the class, the teacher provided the researcher with a list of all the classes that they taught. Classes with very small numbers were deleted from consideration for two reasons: firstly, the results would be unlikely to be statistically useful (a class size of about 25 was desirable), and secondly because it was not practically useful - a teacher engages in a different relationship with a class of 6 compared with the relationship they have with a class of 26. This often means that they do not have the same need for the kind of feedback that this procedure was going to provide. Each teacher was then given the choice of eliminating one class from the remaining list, so that they felt that they were not being assessed by their least preferred class. From the classes that remained on the list, one class was randomly selected to participate.

The classes covered all form levels from Form 3 to Form 7, and a range of subjects across the curriculum as shown in Table 3.4

Table 3.4

Summary of Classes Selected by Form Level and Subject.

Form Level	3	4	5	6	7	Total
Number of Classes	1	6	4	3	2	16

Subject	Number of Classes
English	4
Mathematics	3
Sciences	2
Social Studies	2
Geography	2
Drama	1
Information Technology	1
Latin	1
Total	**16**

The Students

As is typical in cases involving classes in schools, the researcher was unable to randomly assign students to classes. Therefore, although the schools were randomly chosen from the schools that volunteered, the selection of the teachers from the list of volunteers was random, and the selection of the classes was randomly done, in reality the overall sampling was anything but random.

Altogether 344 students in sixteen classes ranging from Form Three to Form Seven completed either one or both of the administrations of the questionnaire. This is shown in Table 3.5

Table 3.5

Number of Students Completing the First, Second or Both Administrations of the Student Questionnaire.

Number completing the first administration	307
Number completing the second administration	306
Number completing the first administration only	38
Number completing the second administration only	37
Number completing both administrations	269

All students were required to provide a written consent after they had been fully briefed about the research. In the case of students under the age of 16, consent was also required from a parent or caregiver. In all, 89.2% of the class totals consented to take part, and of them, 99.1% actually completed one or both of the questionnaires.

The distribution of respondents in each class is shown in Table 3.6 below, while the class levels of the students are shown in Table 3.4 above.

Table 3.6

Classes, showing the Number in the Class, Number who Consented, and Number who Completed the Two Administrations of the Questionnaire.

Class	No in Class	Consented	First	Second	Both
A01	32	26	23	21	18
A02	27	25	22	24	21
B03	25	23	19	19	15
C05	27	26	22	24	20
C06	27	22	17	21	16
D07	18	17	15	14	12
D08	26	18	15	16	13
E09	24	22	18	20	16
E10	21	20	19	17	17
F11	36	26	26	23	23
F12	27	23	22	17	17
G14	14	14	12	13	11
I17	23	23	22	22	21
I18	26	26	24	22	20
J19	21	21	18	19	17
J20	15	15	13	14	12
Totals	389	347	307	306	269

The Questionnaire

In a search of the literature, almost all of the studies involving student rating questionnaires have used instruments that have been designed for use in universities and tertiary colleges. Two studies were found that took place in schools, and in both cases no instruments were attached for reference. Therefore it was necessary to develop an instrument for this research project.

The development of the instrument followed a path suggested by a number of writers (Berk, 1979; Greenwood et al, 1973; Marsh, 1994). In a process of distillation, a large item pool was collected from a literature review (Greenwood, 1973; Frey, Leonard and Beatty, 1975; Cohen and Herr, 1979; Marsh, 1987 which contains the SEEQ, Endeavour, SIRS, SDT and ICES instruments; Kelk, n.d). Duplicated items were rewritten to cover the important teaching behaviours. The items were then grouped together using *a priori* the SEEQ factors that had been well validated and were found to be similar to factors found in other instruments. These factors are: learning/value; organisation/clarity; instructor enthusiasm; group interaction; individual rapport; breadth of coverage; examinations/grading; assignments/reading; and workload/difficulty. Two global rating items were to be included covering an overall teacher evaluation and an overall course rating. In addition, there were to be

two open-ended questions that allowed students to respond in their own words. This also allowed teachers to respond to qualitative data as well as quantitative data.

Two factors (breadth of coverage and assignments/reading) as categorised in the SEEQ were not regarded as being as important for secondary school teaching by an independent group of experienced teachers and were deleted, but two other categories/factors not typically part of a tertiary teacher's job were strongly recommended - homework (similar to and replacing assignments/reading) and classroom control. Items for these two factors were developed and similarly sorted into an order of priority. All items were then subject to checking against the 18 point Checklist for Evaluating Rating Scale Items found in Berk (1979:661) to ensure that they met acceptable standards.

Six experienced teachers from a school that was not included in the study were asked to sort the questions, giving the highest priority to items that would provide useful feedback to the classroom teacher. These teachers worked independently of each other. The items which were included in the final form of the questionnaire, as well as those rejected are shown in Appendix 3.2.

The 50 rating items were then randomly assigned to their position on the final form, followed by the two global rating items then the two open-ended response items. This form was then field tested using a third form class from a non-participating school. Amendments were made to the wording of questions that were ambiguous or beyond the comprehension level of secondary students. Using the Fry Readability Formula (Fry, 1972), the readability was assessed at 11 years, an entirely appropriate level for secondary students. The final form of the questionnaire (Appendix 3.3) was used for the Time One teacher evaluations in June and July.

The Time Two questionnaire was a modified version of the original questionnaire. The same 52 Likert scale statements used in the first questionnaire were repeated. The two open-ended items were removed as they were intended to provide information for the teacher when planning for change. In their place were eight further statements using the same Likert scale as the main questionnaire showing degree of agreement/disagreement. These statements asked the students to indicate whether they felt the project was worthwhile, whether the teacher had thanked the class for their input, whether the teacher had discussed the results with the class, whether they felt the teacher had tried to change, whether the teacher had in fact changed as a result of the first part of the research, whether they would like to

evaluate all of their teachers in this way, and whether they thought any student had suffered (been "picked on") as a result of the research project. They also had one open-ended response opportunity to comment on any aspect of the research project (Appendix 3.4).

The Time Two administrations were completed in late September and early October with one exception. Class D07 was a studying under a trimester system and were to disband at the end of Term Two. This class completed the Time Two survey in the last week of Term Two, at the end of August which was eight weeks after the Time One survey. This was the shortest period between the two administrations for any class.

Important decisions had to be made about the type of questionnaire to be used for collecting the student evaluations. Over the years, the most commonly employed questionnaires have used a Likert scale on which the student indicates their degree of agreement (strongly agree through to strongly disagree) with the item stem. Typically, a five point scale has been used, with 3 as the neither agree nor disagree category. Why is this type of questionnaire so frequently used? The Likert scale instrument is popular because it is easy to construct, is easily understood by both teachers and students, and acceptable levels of reliability and validity can be obtained with instruments using this type of scale (Berk, 1979). Three variants of the Likert scale are commonly used: an intensity scale of more or less; a frequency scale of always to never; and a quantity scale of all or none. The most common choice from these is the intensity scale using the familiar completely agree to completely disagree continuum. This scale has been chosen for this study as it can then draw on the vast body of studies that have previously used this form. A five-point numerical scale (rather than a seven-point or nine-point scale) was used as this did not require the students to make fine distinctions in a teacher's performance, particularly as secondary students have had little or no experience of rating their teachers' performance. Furthermore, one study of the reliability of different numbers of scale points showed that there was a definite levelling off in reliability after five points on the scale (Lissitz and Green, 1975).

For this research, 1 represented Strongly Agree and 5 represented Strongly Disagree. In the case of the items that were expressed negatively, the items were reverse scored so that a low score was usually desirable. Therefore, the absolute minimum average was 1 (all students strongly favouring the teacher's behaviour with regard to that item)

and the absolute maximum average was 5 (all students strongly unfavourable towards the teacher's behaviour on that item).

The five point scale was modified slightly by the addition of a "Not Applicable" category which was scored by the use of a sixth point in the coding of the student evaluations. To enable the results of the research to be generalisable, the instrument was intended to be a common device for use with all classes in the research, and this necessitated the "Not Applicable" option. In particular, there was one question designed to assess whether teachers discriminate between boys and girls in their classes. This question has no meaning or relevance in a single sex school, but to eliminate it from the questionnaire when it could provide interesting information from co-educational classrooms would have been an opportunity wasted. Teachers often do not know whether they do behave differentially towards boys and girls, and the provision of this information was intended to give them a clear indication of how the students perceived their behaviour in this respect. Another set of statements where the "Not Applicable" category was going to be necessary was for the items concerning homework. Not all teaching subjects have a homework component, and this sixth category was going to be needed for these items. In a classroom situation where the teacher was choosing items from an item bank, there would be no need for these items at all. The results for the Time One and Time Two questionnaires used only the five scale points of agreement-disagreement, as the numerical value of six for Not Applicable has no meaning in this context. It is solely a coding feature.

Statements used in the questionnaire were domain-specific rather than factor-specific or behaviour-specific. Single items designed to cover each factor (for example, "This teacher enjoys a good rapport with the class" for the Rapport factor) are too global for the statement to provide any useful information that will assist the teacher to improve. If the students indicate that the teacher does not enjoy a good rapport with the class, is it because the teacher needs to learn the students names, or because the teacher humiliates students in front of other students, or because the teacher is not available to talk to students after class, or for some other reason? On the other hand, statements that are too behaviour-specific may be too detailed as to be almost trivial (for example, "The teacher's use of facial expression is exceptional."). Therefore statements had to be domain-specific adequately representing some domain of teaching so that the teacher could identify their strengths and weaknesses from the student responses. Domain-specific rather than behaviour-specific information enables the teacher to modify their behaviour and maximise the impact of the changes made (Cohen and Herr, 1979). Some of the statements were expressed negatively

(for example, "The teacher is not interested whether I can handle the work in this subject") and required the students to shift their response behaviour to the other end of the scale. In the trial testing, this appeared to cause no difficulty for students, and they did respond in a manner consistent with their responses to other positively stated items.

The Consultation

By itself, student evaluation data alone does not produce the expected improvement in subsequent ratings by students as a measure of improvement in teaching. Menges and Brinko's meta-analysis showed that feedback of this type produced a very small effect, raising subsequent ratings by one-fifth of a standard deviation. That is, "after receiving student ratings feedback, the average instructor was rated higher than 59 percent of control group teachers" (Menges and Brinko, 1986: 5). However, when they investigated feedback studies which included consultation conferences they found that on average, "feedback of this type raised the subsequent ratings by more than one standard deviation" or, teachers who received student rating feedback and consultation subsequently "rated higher than 86 percent of teachers in the control group" (Menges and Brinko, 1986: 5). Therefore, the decision was made to include a consultation conference as an important part of the research design.

Consultancy is closely linked with planned change. One of the pioneers of planned change was Kurt Lewin who developed a model involving three stages: unfreezing which creates the motivation to change; changing which involves developing new ideas, beliefs, attitudes, values and behaviours based on new information; and, refreezing which stabilises and integrates these new beliefs and behaviours into the rest of the system (Schein, 1972). Further development of this model emphasised the importance of a "change agent" in a facilitative role as part of change-oriented problem solving (Lippiatt et al, 1958). Despite general interest in the field of consultancy in organisations, there is a paucity of literature on educational consultancy using feedback from student evaluations (Brinko, 1990). Four models of educational consultancy have been proposed and were investigated in Brinko's study - Product, Prescription, Collaborative/Process, and Affiliative. A year later, she added a fifth - the Confrontational model (Brinko, 1991). Each of these models involve different interactions between the consultant and the client/teacher. Product consultants supply solutions to problems that have been diagnosed by the client; Prescription consultants act like physicians who diagnose the problems and prescribe solutions for the client; Collaborative consultants act more like partners encouraging

clients to diagnose problems and provide solutions to their own problems; and, Affiliative consultants focus on solving the personal problems of the client that may cause problems in the client's work. The Confrontational consultant challenges the client or acts as the devil's advocate. She found that most educational consultants worked along a continuum from prescriptive to collaborative, and that the other two (three) models were by and large absent. Consultants may need to move from one style to another in order to be responsive to the client and she concludes that there is no evidence yet that conclusively shows which strategies and practices make consultation successful. Her results suggest that collaborative consulting may be more effective with "more professionally mature teachers" and that prescriptive consulting may be more effective with "novice teachers or those who do not have the time or interest in collaboration" (Brinko, 1990: 78).

All of the teachers in the project were experienced teachers who had volunteered knowing that the results of the evaluations would be discussed in a consultation conference. Furthermore, insights that you discover for yourself are more likely to be retained and acted on than those other people tell us about (Acheson and Gall, 1992, 189) Therefore, the consultation conference involving the teacher and the researcher was based on a collaborative approach.

This approach to the consultation conference was based on a model after the work of Acheson and Gall (1992), and modified by Aitken (1994) for use by College of Education lecturers when supervising the teaching practice component for teacher trainees. Acheson and Gall outlined the following procedure: the consultant provides the teacher with objective observational data, they analyse the data co-operatively, and then reach agreement on what is happening. The consultant solicits the teacher's reaction to the data (inferences, opinions, feelings) and they consider possible causes and consequences. Together, the teacher and consultant reach decisions about future actions and provide opportunities for practice and comparison.

The conference framework was supplemented with advice from other sources to optimise the value of the feedback. Duke and Stiggins (1990) detail nine characteristics of good feedback that are correlated with the perceived quality and impact of the feedback: quality of ideas for improvement; depth of information; specificity of information; account of information; extent to which the information was descriptive (not judgmental); timing of the feedback to promote attention to the message; extent to which feedback was linked to standards; frequency of formal feedback; and frequency of informal feedback. In describing the key steps involved in

giving good quality feedback, Gibbs et al (1989) suggest that the consultant invite the recipient to speak first, be specific and concrete rather than general, balance positive with negative, and direct the feedback towards behaviour that the recipient can control. Gold and Roth (1993) report similar ideas as well as noting that the consultant should not demand change as feedback does not require the receiver to change their behaviour. The consultant should focus on sharing rather than advising, and also focus on alternatives. Effective praise is an essential component (Good, 1984), and the consultation should be constructed like a "praise sandwich" (Parsons, 1995) - that is, start and end with praise but have a meaty discussion in between. Winder (1996) advises consultants to use four key processes - pose challenging questions; paraphrase responses; seek clarity, elaboration and specificity; and make effective use of silence to allow for thinking. Teachers (who are investing time in a procedure designed to help them improve) want to feel that they have done something and go beyond the mere description and discussion stage to the development of action plans (Finlay, 1992).

The consultant also needs to be aware of the type of language used during the consultation. In an examination of three types of language used in supervisory conferences, Pajak and Glickman (1989) found that teachers gave the highest preference to information with suggestions, then information only, with supervisors who used information with directives as the least preferred communication mode. This is consistent with Deci and Ryan's (1985) theory of information and control which suggests that the more choice individuals are given over their activities, the more productive and satisfied they are. It is also clearly in keeping with the findings of Tuckman and Oliver (1968) who found that teachers react negatively to feedback from supervisors.

In accordance with Richard's advice (1987), teachers were advised to demonstrate to the students the usefulness of the information they had provided by discussing the results with them, and calling their attention to any changes made as a result of their feedback.

The role of the consultant in the conference is critical. The consultant needs to act as a facilitator and as the reflective partner or critical friend of the teacher (Ker, 1995), holding up a mirror which reflects student feelings to the teacher, and helping the teacher to critically examine that information and devise strategies to overcome any concerns. The consultant will ask two key questions of the teacher: "How can we do this better?" but in a more reflective/critical mode must also ask "Should we be doing

this at all?" In practice, in the consultation this latter question was framed as "Is this important for learning?" These two questions borrow from Argyris and Schon's (1974) distinction between single-loop and double-loop learning. This also involves sharing the feedback in a way that raises the commitment to take appropriate action as a consequence (Aspinwall et al, 1992). In the role of facilitator, the start of the consultation would commence with an introduction along these lines borrowed from Marsh and Roche (1993: 228)

> I do not have sufficient background in your area to know what is "best". Instead I will discuss the ratings and work with you to develop some strategies in particular areas selected by you.

Again, this emphasises the teacher's ownership of the process and the planned outcomes.

The researcher-cum-consultant was careful to follow the role prescribed in the Human Ethics application relating to the research. This involved maintaining a professional distance from the subjects, and reminding them that they were entitled to have a support person with them throughout the consultation or to withdraw at any time for any reason without having to provide any explanation.

The framework for the conference is in Appendix 3.5

The Follow-Up Questionnaires

At the Time Two survey, both the students and the teachers completed different questionnaires regarding their feelings about participation in the research project. The student questions have already been outlined above. The teachers were asked to complete a rather fuller questionnaire covering all aspects of their participation and their feelings about issues related to the use of student evaluations in an appraisal context. The main areas covered in this questionnaire (Appendix 3.6) were: the student questionnaire instrument; the written feedback provided; the consultation conference phase; their resulting teacher development; their reactions to the whole research project; teacher appraisal and the place of student evaluations; and, general comments on the research. The questionnaire consisted of a number of items under each heading, with some forced response items, others asking the teacher to identify features that they liked or disliked, together with the opportunity to comment on

every question. Replies were received from all teachers, but some did not respond to every prompt or question. The comments were content analysed for commonalities, and these feature in the results.

Time

A subsidiary consideration was to investigate whether student evaluations are a time consuming means of collecting data for teacher development. Critics of appraisal processes often refer to the considerable time requirements, so a decision was made to check the time requirements for the administration of the questionnaires. The other aspects of a teacher development programme (for example, discussion of the data with a colleague) would require the same time no matter how the data was collected, so the real variable for this project was the administration of the student questionnaires, and the processing and printing of the results.

The administration of the questionnaire followed a standardised procedure, with a prepared set of written instructions that were read to every class. This part of the procedure took between five and ten minutes depending on the amount of help a class required to complete the demographic data at the top of the questionnaire form.

The time each class took to complete the questionnaire was recorded in three steps. Records were kept of the time taken by the first student to finish, the first ten students in the class to finish, and the time taken for all of the class to finish.

The processing of the questionnaires took approximately one hour per class to produce the printouts used in the consultation, and the same amount of time for the final printouts after Time Two.

Summary

The methodology used draws on the literature regarding the use of student evaluations coupled with consultation, and particularly uses the work of Wilson (1986) and Marsh and Roche (1994) for the overall process. A simple pre-test/post-

test design was employed, using questionnaires as the means of collecting data. References on questionnaire construction, consultation procedures, sampling techniques, and other aspects of the process were consulted in the design of the research. The use of volunteers limited the ability of the research to provide suitable statistical controls for threats to aspects of the experimental validity, but both qualitative and quantitative data were obtained from both sets of participants at each stage of the research to overcome this.

Chapter 4 Results

The main research question to be answered was

Are student evaluations of teachers coupled with a consultation conference to develop an action plan for change, an effective and useful means of bringing about teacher development, particularly improvement?

Subsidiary research questions were:

Do teachers accept student evaluations as a valid, reliable, credible and useful means of gathering data for development, and are they therefore prepared to act upon them?

Do teachers see this form of evaluation as part of a developmental (formative) appraisal process, as part of a performance (summative) appraisal process, or as part of both? How best can student evaluations be incorporated in an appraisal system, if at all?

Are the time requirements realistic and manageable for teachers in gathering and analysing data, and developing and implementing an action plan for improvement?

In addition:

What can we learn about the way that secondary students view their teachers and their learning experiences?

Part One: Quantitative Results

Student Evaluations
Descriptive Information

Using the raw data, some impressions about the way that students see their teachers can be gained. The data were examined to see what trends emerged in the way that students responded to the various items in the questionnaire. All items in the

questionnaire were ranked for each of the response categories 1 "Strongly Agree", 2 "Agree", 4 "Disagree", and 5 "Strongly Disagree".

Table 4.1 shows the questionnaire items that the students responded most favourably to over the two administrations, both separately and combined.

Table 4.1

Rank Order for Response Categories 1 and 2, separately and combined.
Time One and Time Two combined.

Rank	Item	1 Str Agree	N	Item	2 Agree	N	Item	Combined	N
1	46	Pronounce	296	01	Pride	315	06	Happy	524
2	38	Inclusiv	222	06	Happy	311	46	Pronounce	511
3	39	Get Job	220	29	Incr Think	293	23	Try Hard	478
4	23	Try hard	217	13	Impt Pts	286	13	Impt Pts	462
5	06	Happy	213	36	Hmk Cpes	283	38	Inclusive	448
6	27	Org/Expct	177	16	Asmt Fair	278	36	Hmk Cpes	444
7	51	Bst Tchr	177	30	Esy Undrst	270	16	Asmt Fair	443
8	13	Impt Pts	176	04	Clarity	266	27	Org/Expct	419
9	40	S Ask Qns	173	02	Discuss.n	265	01	Pride	417
10	47	Cont Nxt Yr	170	23	Try Hard	261	39	Get Job	409

Students are very positive about their teachers because they pronounce their names properly and are inclusive in the language and examples they use. Students believe that they themselves try hard in class, cope with the homework they are given, take a pride in their work, and that success in the surveyed subject will help them to get a job. They also believe that the teacher emphasises the important points to learn, is fair in assessing them, and generally spells out how the course is organised and the expectations they have of the students.

The second table (Table 4.2) shows the questionnaire items that the students disagreed with over the two administrations. As would be expected, the eleven negatively worded items (Items 03, 07, 14, 22, 24, 25, 32, 42, 45, 49, 50, all indicated by an asterisk) occurred within the first fourteen items of the combined Disagreement column (and at the very bottom of the combined Agreement column). This indicates that students were not just using one end of the scale for the whole questionnaire, were thinking carefully about their responses and taking the whole exercise seriously. Students believe that their teachers are in control of their classrooms, are interested in whether students can handle the work, don't leave students who finish early with nothing to do, and that students don't ignore the teacher. In addition, students feel that their teachers do not talk a lot in class, do allow them to work at their own speed

and do not give them too much homework. Disruptive students do not feature prominently in the classroom, at least in the mind of the students on this part of the questionnaire.

Table 4.2

Rank Order for Response Categories 4 and 5, separately and combined.
Time One and Time Two combined.

Rank	Item	4 Disagree	N	Item	5 Str Disag	N	Item	Combined	N
1	03	Handling *	232	07	Control *	257	07	Control *	459
2	49	Fin Early *	228	32	No Hm Hlp *	234	03	Handling *	426
3	25	Ignore Tch *	223	03	Handling *	194	49	Fin Early *	410
4	50	Talk A Lot *	217	49	Fin Early *	182	32	No Hm Hlp *	397
5	28	Too Much	213	42	Difficulty *	177	25	Ignore Tchr *	358
6	45	Speed *	203	25	Ignore Tch *	135	42	Difficulty *	356
7	07	Control *	202	50	Talk A Lot *	131	50	Talk A Lot *	348
8	22	Hmwk Ovld *	191	24	Boys Qns *	124	45	Speed *	305
9	15	Disrupt. Std	190	37	St Disply	103	22	Hmwl Ovld *	292
10	42	Difficulty *	179	45	Speed *	102	15	Disrupt. Std	289

The combined frequencies were calculated for the "Neither Agree or Disagree", and the "Not Applicable" responses. These are shown in Table 4.3.

Table 4.3

Rank Order of Response Categories 3 and 6 for Time One and Time Two combined.

Rank	Item	3 Neutral	N	Item	6 Not Appl.	N
1	14	Strict *	276	24	Boys Qns *	213
2	26	Feel Gd	243	37	St Disply	48
3	10	Related	218	22	Hmwl Ovld *	37
4	22	Hmwk Ovld *	203	36	Hmk Cpes	36
5	28	Too Much	200	32	No Hm Hlp *	34
6	33	Stdy Skls	200	16	Asmt Fair	31
7	09	Approachable	194	43	Cmnts Hlp	27
8	18	Know Wld	185	05	Prompt	25
9	19	Chall/Intrst	185	33	Stdy Skls	21
10	08	Challenge	179	47	Cont Nxt Yr	21

Students may have had difficulty interpreting Item 14 Strict* judging from the high number of neutral (Neither Agree nor Disagree) responses. The wording of the item left students with possibly two ways of viewing their response and this may have caused them to go for the safe territory of the neutral response category. Item 09 Approachable caused some discussion with the teachers, who felt that although students might approach them about matters related to their subject that were on their (the students') minds, students may not approach them about any other matters, and

therefore tended to opt for the neutral category as a way of indicating their ambivalence.

Item 24 Boys Qns* had the highest frequency of Not Applicable responses by a large margin with 213 students opting for this response category. This was to be expected as six of the sixteen classes (127 students) were from single sex schools. Given that they had two opportunities to respond to this item, it is clear that some of the students in the single sex classes did not use this response category on one or both occasions. The three Homework items all occurred in the top five in the Not Applicable ranking. The Form Four Drama class (taught by Teacher C06) had no homework as part of their course, which could account for 27 of the Not Applicable responses to Items 22 and 36, and perhaps Item 32. Although one teacher (E09) expressed considerable concern at the inclusion of the Not Applicable point on the rating scale, it is interesting to note that 34 of the 52 items attracted less than 10 individual student responses over both administrations, and this involved 99 student responses in all.

The rank order for the Mean response rate and the variance was determined. The variance was of interest as the smallest variance indicates the greatest agreement amongst the students. This information is shown in Table 4.4.

Table 4.4

Rank Order of Mean and Standard Deviation for Time One and Time Two combined.

Rank	Item	Mean		Item	Std. Dev.	
1	46	Pronounc	1.78	01	Pride	0.83
2	06	Happy	1.91	13	Impt Pts	0.89
3	23	Try Hard	1.99	06	Happy	0.89
4	07	Control *	2.00	16	Asmt Fair	0.90
5	38	Inclusiv	2.03	29	Incr Thnk	0.93
6	13	Impt Pts	2.06	23	Try Hard	0.93
7	16	Asmt Fair	2.07	46	Pronounc	0.96
8	36	Hmk Cpes	2.09	48	Examples	0.96
9	39	Get Job	2.15	36	Hmk Cpes	0.96
10	32	No Hm Hlp *	2.18	17	Mat Prep	0.97

In the consultations, teachers tended to look at the mean rating for each item for an indication of whether they had performed (and therefore, scored) well on that item. Because all of the negative items were reverse scored, teachers were looking for a low mean score to indicate an item where the students were expressing their pleasure or satisfaction with the performance of the teacher or the class on that item. The mean ranking table effectively merges the two ranking tables above (Tables 4.1 and 4.2) and shows the strength of feeling the students have about the positively and

negatively worded items. These indicate that the students feel very positive about their classes and claim to work hard in them, that their teachers create a positive affective climate in the classroom yet remain firmly in control, and feel that success in the class will improve their chances of obtaining a job in the future.

The standard deviation ranking shows that students expressed some degree of unanimity on six of the items that are in the mean rank order. Over both administrations, twelve items had a standard deviation less than 1.00. At Time One, the minimum standard deviation was 0.85 and the maximum 1.35. Ten items had a standard deviation less than 1.00. At Time Two, the minimum standard deviation was 0.81 and the maximum 1.39 which indicates that the students were making slightly more use of the full extent of the rating scale on the second administration. At the same time, there were 22 items with a standard deviation less than 1.00 indicating that there was less variability in their ratings at Time Two than there had been at Time One. This may be the result of respondent familiarity with the questionnaire, or could indicate that the outcomes of Time One (that is, the changes made by the teacher) had helped to shape the opinions of the students by time of the second administration and narrowed their responses rather more than on the first occasion.

Factor Analysis

Factor analysis is a means of taking a data set and distilling it to help explain what the data means. This is done by looking for commonalities in the way that the responses are grouped together, and using these to find underlying factors. In this research, two approaches have been used to address the matter of factors in the questionnaire. The first of these was to use an *a priori* approach to the items in the questionnaire based on the factors found by Marsh in many analyses of his Students' Evaluation of Educational Quality (SEEQ) instrument (Learning/Value; Group Interaction; Organisation/Clarity; Instructor Enthusiasm; Individual Rapport; Examinations/Grading; and Workload/Difficulty were all included, but Breadth of Coverage; and Assignments/Reading were excluded). These were supplemented by two other aspects of learning and teaching relevant to secondary schooling but not as relevant to tertiary education - Classroom Control, and Homework. This *a priori* approach was useful for the conduct of the research and for the consultation work with the teachers, but was not suitable for the analysis required to extract the best information from the data-set. The questionnaire also included items that did not fit into *a priori* factors based on Marsh's work. These items were designed to help the teacher interpret the results by telling them something about their students and their

attitudes towards the class, subject and the teaching they experienced - these "factors" were called Attitude and Utility.

The questionnaire items were arranged into the following *a priori* factors (the full wording is contained in Appendix 4.1):

Learning/Value
> Items 08, 18, 19, 29, 33, 48, and 50

Organisation/Clarity
> Items 04, 13, 17, 27, 30 and 44

Teacher (instructor) Enthusiasm
> Items 11, 12 and 41

Other Students (group interaction)
> Items 02, 06, 34 and 40

Individual Rapport
> Items 03, 09, 24, 31, 37, 38 and 46

Assessment (examinations/grading)
> Items 05, 16, 20, 21 and 43

Workload/Difficulty
> Items 28, 35, 42, 45 and49

Classroom Control
> Items 07, 14, 15 and 25

Homework
> Items 22, 32 and 36

Overall Evaluations
> Items 51 and 52

Attitude
> Items 01, 23, 26 and 47

Utility
> Items 10 and 39

The second approach was to determine a set of *post facto* factors using factor analysis on the data set from the two administrations of the questionnaire. There were 52 common rating items in the two administrations of the questionnaire but only thirty-six were included in this process.

Fifteen of the items were eliminated for one of three reasons:

a. The item attracted a high non response (or Not Applicable) rate. The questionnaire had been designed to be a generic rating instrument for use in a wide variety of classes, and therefore some of the items in the instrument were not applicable to all class situations. The items on homework and the one item on whether questions were directed to boys in preference to girls attracted high "Not Applicable" response rates. In the statistical process of factor analysis, any student who does not answer all items with a valid rating response is excluded from the analysis. Therefore if the item on boys/girls was left in the factor analysis, all the students in single sex classes would have automatically been omitted, rendering the responses of a significant proportion of the participants as worthless. The overall response rates of all of the items were scanned, and any item with a non response (or Not Applicable) count of 27 students or more was eliminated from the analysis. These items were:

16 Tests and assignments in this subject are fair and cover the work we have been taught.

22 We get too much homework in this subject.

24 Boys are more likely to be asked to answer questions during lessons in this subject.

32 If I needed help with my homework in this subject, I would not be able to get it.

36 I can usually cope with the homework we are given in this subject.

37 Student work is displayed in our classroom and is changed during the year.

43 The teacher's comments on my tests and assignments are constructive and helpful.

b. The item said more about the student than the teacher, or about the subject and therefore was not useful in analysing teacher behaviours. These items had been included to allow the teacher to gauge some sense of how the students in the class felt about the class or subject. In the *a priori* analysis, these items were classified as student attitude and as utility of the subject:

01 I am usually proud of the work I do in this class.

10 The teacher shows us how this subject is related to the other subjects taught at school.

23 In this class I usually try to do as well as I can.

26 The thought of coming to this class makes me feel good.

39 If I do well in this subject, it will help me to get a job.

47 I want to keep doing this subject after this year because I am enjoying it.

c. The item was an overall rating item. These two items were:

51 Overall, this teacher is one of the best teachers I have this year.

52 Overall, this subject is one of the best classes I have this year.

One other item (14 My teacher is strict) was eliminated at a later stage as it was a confounding variable in all attempts at factor analysis and stood alone as a single item factor thereby adding nothing to the analysis. This may well have been because the item could be interpreted in one of two ways - the teacher is strict and that is good, or, the teacher is strict and that is not good - leading to the difficulty in factor analysing it.

The remaining thirty-six items meant that 462 of the 635 student responses (73%) were included in the factor analysis (principal components with varimax rotation) which converged in thirteen iterations and revealed seven factors. The seven factors explain 52.2% of the variance which is less than the 60% standard set as desirable by Hair, Anderson et al (1995, 378) for work in the social sciences. However, under the circumstances of this research, the 52.2% is acceptable given that the research was conducted in nine schools, with 16 teachers teaching eight different subjects across all five levels of secondary schooling. In other words, unlike the research outlined in Chapter Two which tended to focus on one subject at one level in one institution, there was a considerably greater source of variation in the data set for this research than in other research on this topic. The Measure of Sampling Adequacy (MSA) was in excess of 0.80 (described by Hair, Anderson et al as "meritorious") for all except three items which were all in excess of 0.70 (described as "middling"). The lowest factor loading was 0.39 (Item 13, Impt Pts on Factor One) indicating that all of the loadings were of some importance. Indeed, factor loadings of 0.30 require a sample size of 350 for significance, so given the sample size of 462, all loadings indicate significant correlations between each of the original items and the factor it related to. (For a detailed discussion of Factor Analysis refer to Hair, Anderson et al, 1995, 364-419).

The table for the rotated factor matrix appears in Appendix 4.2. The seven factors, in the order of their extraction, are described throughout this report as:

Factor One: **General Teaching Skills**
      Items 03, 04, 09, 11, 12, 13, 30, 31, 35, 38, 40 and 48

Factor Two: **Value of Learning**
      Items 08, 18, 19, 20, 29, 33 and 41

Factor Three: **Classroom Management and Control**
      Items 07, 25, 49 and 50

Factor Four: **Teacher Organisation**
      Items 05, 17, 44 and 46

Factor Five: **Workload/Difficulty**
      Items 28, 42 and 45

Factor Six: **Student Cooperation**
      Items 02 and 34

Factor Seven: **Student Contentment**
      Items 06, 15, 21 and 27

The complete wording of each item in each factor is contained in Appendix 4.3. There is some direct overlap between the factors of this research and those found in Marsh's SEEQ.

Did the Teacher Improve?

The main purpose of this research was to determine whether an approach to teacher development involving student evaluations coupled with consultation had a positive effect which would be measured by a change in the rating of teachers between the Time One and Time Two evaluations.

The analysis for this question was done in three ways:

- Measuring the size effect, d
- Calculating the Student t-test for the difference in the means for each item between Time One and Time Two.
- Calculating the Student t-test for the difference in the means for the seven factors between Time One and Time Two.

Size Effect, d
In the first, the teacher received an analysis of the size effect (d) relayed to them with the results of the Time Two evaluation. Size effect (d) is defined as:

$$d = [\mu(b) - \mu(a)] / [\sigma(a)]$$

where $\mu(a)$ = Time One mean

$\mu(b)$ = Time Two mean

$\sigma(a)$ = Time One standard deviation

Size effect measures the relative change for each questionnaire item between the Time One and Time Two surveys, taking into account the spread (standard deviation) of the first set of results. A negative value for d represented an improvement. Absolute changes in the value of d greater than 0.5 (that is, changes greater than 0.5 if you disregard the negative or positive sign) were operationally defined as being educationally important (Cohen, 1977; Wilson, 1986). This gave the teachers a simple intuitive guide to the extent and importance of any changes made as a result of the research. A comparison of size effect and Student t-test statistics was conducted and showed that as a rough rule of thumb, $d>0.5$ represented significance at about the 10% level ($p<0.1$). This is not an acceptable significance level for the purposes of research but does help teachers to gauge progress towards their goal of improved teaching. The size effect (d) for each item for each teacher is shown in the tables of Appendix 4.4 in the Column F.

Each teacher's results were analysed for the total number of beneficial changes (that is, an improvement on that item in the students' eyes as indicated by a reduced mean score on the item between the two administrations), and total number of non-beneficial changes (a higher mean score on the item). In addition, the results were also examined for those items that had an educationally important change (beneficial or non-beneficial) as defined above. This is shown in Table 4.5.

Table 4.5

Number of Beneficial and Non-Beneficial Changes in Item Scores

between Time One and Time Two for All Items

All Changes	A01	A02	B03	C05	C06	D07	D08	E09	E10	F11	F12	G14	I17	I18	J19	J20
Beneficial Change	33	49	37	37	30	30	20	36	27	20	32	32	16	10	29	34
Non-Beneficial Change	18	3	15	15	21	16	31	15	25	31	18	20	36	40	21	16
Educationally Important																
Beneficial Change	7	20	5	10	8	6	3	5	4	1	4	6	0	0	0	6
Non-Beneficial Change	1	0	4	0	5	2	13	1	2	1	0	1	8	4	3	2

Student t-Tests on Items

The second way that change was measured was the Student t-test. All 52 of the items used in both administrations of the questionnaire were subject to t-tests for difference in means between Time One and Time Two administrations. The Student t-test is a statistically acceptable test for change in the mean between two events and this was used as a comparison with the size effect coefficient (d). The results of the t-tests are shown in Appendix 4.5.

From a total of 826 separate tests on the 52 items, 56 (6.8%) reached statistical significance at or beyond the 5% level ($p<0.05$). Of these 56, 17 were highly statistically significant at the 1% level ($p<0.01$). By chance, 41 of the t-tests would have reached the 5% significance level and 8 the 1% level. The chi-square (χ^2) statistic is 14.792 ($p<0.001$) which indicates that there was a substantial effect beyond that expected by chance. However, the experimental design is weak in that it does not allow for a gain score effect which may have occurred naturally, so this result has to be interpreted cautiously.

As part of the above analyses, particular interest was paid to any changes in the targeted items. As the consultation produced an action plan that targeted two or three specific items from the questionnaire, and the teacher had worked to improve those aspects of their teaching, it was assumed that the students would detect the changes and reflect these in their ratings for those items. The table shows the effect on the targeted items in terms of the number of changes made as well as the number of educationally important changes.

Table 4.6

Number of Beneficial and Non-Beneficial Changes in Size Effect d on Item Scores between Time One and Time Two for Targeted Items.

Targeted Items	A01	A02	B03	C05	C06	D07	D08	E09	E10	F11	F12	G14	I17	I18	J19	J20
Beneficial Change	2	4	0	3	2	5	NA	1	0	1	1	2	3	1	2	NA
Non-Beneficial Change	0	1	2	0	0	3	NA	1	2	4	1	1	0	2	2	NA
Educationally Important																
Beneficial Change	0	2	0	0	0	0	NA	0	0	0	0	0	0	0	0	NA
Non-Beneficial Change	0	0	1	0	0	1	NA	0	0	0	0	0	0	0	1	NA

Teachers D08 and J20 did not target any specific items for improvement, hence the Not Applicable (NA) records for those teachers. Just over a half (27) of the 46 changes (58.5%) were beneficial. Five of the 46 changes (10.8%) were educationally

important in terms of the size effect, d, but only two of these five could be described as beneficial.

However, this approach to measuring changes that a teacher may have made is questioned by Marsh and Roche (1993). They note that Wilson's (1986) study had focussed on a comparison of Time One/Time Two gain scores on targeted items to determine the effectiveness of the intervention, and dispute this kind of interpretation on at least three counts. Firstly, targeted items are typically those for which the teacher received a relatively poor rating at Time One. By itself, regression to the mean would result in some positive gains without any intervention. As in Wilson's study, there was no control in this current research for this effect. The second difficulty with Wilson's study centres on the need to demonstrate that gain scores should be larger for targeted items and smaller for non-targeted items. This can be shown by reporting comparisons of gains on targeted items with gains on non targeted items, or by examining gains on the same items by other teachers who did not target those items. While Wilson failed to do this, the information in Appendix 4.5 can be used for this purpose. Thirdly, comparisons between subsets of teachers on the basis of targeted versus non-targeted items involves self-selecting groups which are no longer random and for which control is dubious. Instead, Marsh and Roche suggest that gain scores on overall rating items might provide a stronger basis of inference about the effectiveness of the intervention. In the following table, the size effect, d, for the two overall items is shown. A negative value for d indicates that the students assessed the teacher or class as having improved from Time One to Time Two.

Table 4.7

Size Effect, d, on the Overall Rating Items for Individual Teachers/Classes.

	A01	A02	B03	C05	C06	D07	D08	E09	E10	F11	F12	G14	I17	I18	J19	J20
51 Best Tchr	-0.74	-0.81	-0.34	-0.66	-0.57	0.07	1.06	0.05	-0.13	-0.05	-0.24	-0.17	-0.07	0.06	-0.12	-0.04
52 Best Class	0.08	-0.67	-0.09	-0.41	-0.32	-0.36	0.84	-0.56	0.08	0.11	-0.23	-0.01	-0.03	0.17	-0.07	-0.12

What this table indicates is that the rating of the teacher improved in twelve of the sixteen cases, and the rating of the class improved in eleven of the cases. On average, eight of the sixteen teachers or classes could be expected to improve. For the 51 Best Tchr item, four of these improvements were educationally important, while only one of the non beneficial changes was educationally important. For the 52 Best Class

item, two of the improvements were educationally important and just one non beneficial change was educationally important.

To overcome any objection to the use of size effect, the Student t-test results for these two overall rating items are also shown in Table 4.8.

Table 4.8

Student t-test for the Differences on the Means for the Overall Rating Items for Individual Teachers/Classes.

	A01	A02	B03	C05	C06	D07	D08	E09	E10	F11	F12	G14	I17	I18	J19	J20
51 Best Tchr	-2.39	-2.45	-1.08	-2.24	-2.17	0.18	2.10	-0.24	-0.45	-0.20	-0.80	-0.44	-0.19	0.19	-0.37	-0.09
df	42	43	36	44	34	27	29	35	34	47	37	23	41	44	35	25
Significance	*	*														
52 Best Class	-0.11	-2.10	-0.26	-1.35	-1.41	-1.20	1.80	-1.70	0.23	0.34	-0.76	-0.03	-0.11	0.07	-0.18	-0.32
df	42	43	36	44	34	27	29	36	34	47	37	23	42	44	35	25
Significance		*														

The t-statistic at the five percent level is more sensitive to indicating significant change than d, the size effect statistic.

Student t-Tests on Factors

The third way in which change was measured was by conducting Student t-tests on the factor scores for all teachers and then for each individual teacher. The results of this are shown in Table 4.9.

Table 4.9

Student t-Test Scores for Each Factor Score for Individual Teachers

Teacher	Factor 1			Factor 2			Factor 3			Factor 4			Factor 5			Factor 6			Factor 7		
	t	df		t	df		t	df		t	df		t	df		t	df		t	df	
A01	-1.14	27		-1.04	27		-0.69	27		-0.62	27		0.04	27		-0.57	27		-0.21	27	
A02	-3.33	21	**	-2.18	21	*	-0.44	21		0.59	21		0.06	21		-2.27	21	*	-0.59	21	
B03	0.36	20		-1.29	20		-0.07	20		0.29	20		-0.37	20		0.5	20		-0.99	20	
C05	2.88	41	**	0.43	41		-0.17	41		-1.65	41		0.24	41		0.67	41		0.57	41	
C06	-1.48	2		0.57	2		-0.97	2		0.33	2		2.8	2		0.08	2		2.25	2	
D07	-0.11	20		-0.63	20		0.14	20		-0.32	20		0.33	20		-0.51	20		-0.08	20	
D08	0.5	25		-0.04	25		-1.39	25		-0.4	25		-2.45	25	*	-2.58	25	*	-1.9	25	
E09	-0.53	27		0.31	27		-0.23	27		-0.96	27		-1.07	27		1.54	27		0.54	27	
E10	-0.88	28		0.28	28		0.41	28		1.62	28		-0.45	28		-0.62	28		0.05	28	
F11	0.41	37		-0.28	37		-1.01	37		-2.30	37	*	-1.36	37		-0.06	37		-0.81	37	
F12	0.41	26		-1.38	26		-0.18	26		-0.12	26		0.17	26		-1.58	26		0.52	26	
G14	-1.08	17		0.34	17		1.82	17		0.49	17		-1.52	17		-0.63	17		-1.53	17	
I17	-0.97	29		-0.28	29		-0.92	29		1.48	29		-0.75	29		0.58	29		1.49	29	
I18	-0.69	26		-0.23	26		-1.25	26		1.43	26		0.29	26		0	26		0.51	26	
J19	1.17	33		-0.85	33		0.67	33		-0.57	33		-0.66	33		-1.44	33		0.42	33	
J20	-1.72	22		-0.69	22		-0.24	22		-0.26	22		0.12	22		1.66	22		0.44	22	
	* p<0.05			** p<0.01																	

This type of analysis overcomes the objections that Marsh and Roche (1993) outline because it focuses on the factors. They argue that individual item changes are embedded in the factors , and that subtle changes may not be obvious at the item level but can be reflected in the factor scores.

On a total of 112 factor t-tests, 7 were significant and of those two were highly significant. In this analysis, simply by chance, there would be an expectation of 6 significant results and just one highly significant result. These chance expectations have been exceeded, but only just and not significantly ($\chi^2 = 1.167$, 1 df, p>0.10). Two highly significant results were both obtained in Factor One, the General Teaching Skills factor. The only other factor to obtain two significant results was Factor Six, Student Cooperation. No significant results were obtained on two factors - Factor Three, Classroom Management and Control, and Factor Seven, Student Contentment.

The factor scores were further analysed by conducting one-way analyses of variance (ANOVA) using the student supplied demographic information and the items excluded from the factor analysis as the dependent variable. For instance, if the dependent variable was the student's ethnicity, then this analysis would provide information about whether students from different ethnic backgrounds rate their teacher differently on a particular factor, without necessarily telling us which way the different ethnic groups view their teacher on the factor under consideration. The demographic variables were their gender, their school (specifically the class involved in the research), their ethnicity, and their perception of their own ability in the specific class which was involved in the research. This information is shown in Table 4.10.

Table 4.10

Summary of One-Way Analysis of Variance on Factors as a

Function of Ability, Class, Ethnicity, and Gender.

	Factor 1			Factor 2			Factor 3			Factor 4			Factor 5			Factor 6			Factor 7		
	F	df		F	df		F	df		F	df		F	df		F	df		F	df	
Ability	3.8	2	*	1.6	2		6.5	2	*	7.9	2	**	24.9	2	**	0.2	2		0.11	2	
Class	6.1	16	**	7.4	16	**	14.5	16	**	14.6	16	**	3.8	16	**	16.6	16	**	4.1	16	**
Ethnicity	5.5	4	**	1.3	4		13.4	4	**	2.9	4	*	6.5	4	**	0.4	4		1.0	4	
Gender	2.5	1		0.8	1		2.5	1		6.9	1	**	6.2	1	*	4.0	1	*	1.1	1	

* p<0.05   ** p<0.01

There was a highly significant difference between the way that the students in any one class rated their teacher for each factor when compared with students in the other classes in the research. One should be concerned if this was not the case. Factor 4

(Teacher Organisation) and Factor 5 (Workload/Difficulty) had significant differences on all four variables, and three of the variables were highly significant on these two factors.

On the questionnaire, students were asked to assess their own ability in the class as Top, Middle or Bottom. On four of the factors, the different ability groups rated their teachers in significantly different ways. Student t-tests for the difference between specific pairs of these ability groups revealed highly significant differences between all three groups on Factor 5, and significant differences on Factor 4. That is, the different ability groups in a class will have a very different view of matters concerning the workload and the difficulty of the work they do in that class, and will rate their teachers accordingly on those items. This confirms an expectation that students of high ability will experience the workload and difficulty of the work they do in class quite differently from students of lower ability, and that teachers will be rated differently as a result. Likewise, they will have different views of the teacher's organisation of the class, and this too will be reflected in their ratings.

Different ethnic groups in a class have a different perception of their teacher on matters affecting Factors 1, 3, 4, and 5. Student t-tests between the five separate ethnic groups identified by the questionnaire show that there are significant differences between most groups on Factors 3 and 5 - that is, Classroom Management and Control, and Workload/Difficulty. Students who identified as Asian gave very significantly different ratings from the other ethnic groups to their teachers on Factor 5 - Workload/Difficulty. This would reflect the widespread public belief that Asian students work exceptionally hard compared with other students, and tend to achieve at a higher level. Also of interest in the t-tests was the difference between European/Pakeha students and Pacific Island students - there were highly significant differences on Factors 1 (General Teaching Skills), 3 and 5, and a significant difference on Factor 4 (Teacher Organisation). Indeed, on Factor 3 Pacific Island students had significantly different ratings from all other ethnic groups.

Males and females rated their teachers and classes differently on Factors 4, 5 and 6 (Student Cooperation) with the Factor 4 result being highly significant.

All variables not included in the factor analysis were also used as the dependent variable in a one-way ANOVA on each factor. Marsh and Roche (1993) have argued that the overall rating items may be a better indicator of whether a teacher has improved than any specifically targeted rating items, so the two overall rating items

were of particular interest. Of the other twelve items, only one (32 No Hm Hlp*) revealed significant differences on a large number of factors. The results for these three items are shown in Table 4.11.

Table 4.11

Summary of One-Way Analysis of Variance on Factors as a Function of Best Class, Best Teacher and No Home Help

	Factor 1			Factor 2			Factor 3			Factor 4			Factor 5			Factor 6			Factor 7		
	F	df		F	df		F	df		F	df		F	df		F	df		F	df	
Bst Class	7.9	4	**	32.9	4	**	0.7	4		9.6	4	**	4.1	4	**	0.8	4		3.4	4	**
Bst Tchr	37.7	4	**	20.9	4	**	6.3	4	**	14.9	4	**	1.6	4		0.8	4		0.8	4	
No Hm Hll	12.4	4	**	11.4	4	**	10.4	4	**	3.2	4	*	3.6	4	**	0.8	4		2.7	4	*

* $p < 0.05$ ** $p < 0.01$

For the students in this research, it appears that one of the factors that distinguishes a "best" teacher from a "best" class is the teacher's classroom management and control (Factor 3). Similarly, the workload and difficulty of the work (Factor 5) and the degree of student contentment (Factor 7) help to distinguish a "best" class from a "best" teacher. The third variable was interesting in that apart from the actual school attended (the class variable in the questionnaire), this was possibly the nearest variable to a socio-economic status indicator. It indicated whether the student could count on help and support for their schoolwork from their family at home, and the above statistic shows that students who could count on help at home rated their teachers quite differently from those students who have little or no support with their schoolwork at home. Where the students were able to count on this help and support, they rated their teachers better than those who did not receive this kind of help and support. The only factor not affected by this variable was Factor 6, Student Contentment, which has to do with working in class with other students.

Teachers intending to use student ratings should control for this variable. In preparing a student evaluation questionnaire, they should include such an item for this purpose.

Teacher Self Evaluations

At Time One, all teachers completed the same 52 item questionnaire as the students. They were asked to complete this as if they were a student in the class. In this way it was intended to assess the extent to which the teacher's perception of their performance was in accord with the students' perception. Unfortunately, there was a

much higher incidence of Not Applicable ratings in the teachers responses (12.6% of all teacher responses versus 2.07% of all student responses), and in one case (Teacher D07) the teacher failed to answer half of the items. These incomplete records may be the result of a natural reticence and modesty on the part of the teachers, but they do confound a possible answer to one of the more interesting questions in the literature - what motivates a teacher to change as a result of receiving student rating data? Therefore, the following results may be indicative but any interpretation needs to be treated with caution.

The Time One printed results that the teacher received had their own assessment for each item and the difference between the mean student rating and their own rating. On these results, an absolute difference was shown, without regard to whether the difference was in the positive or negative direction. Directed differences were later calculated in processing the results for the purpose of trying to verify a finding about which teachers are motivated to change their teaching behaviour as a result of receiving student rating feedback.

To investigate this, the relationship between the directed differences between the mean class rating and the teacher's own rating for each item was explored. The Likert scale on the questionnaire used one as the most favourable end of the scale (negatively worded items had been reverse scored to preserve this), so that if the difference between the mean student rating and their own was a positive value, then the teacher had given themselves a more favourable rating than had the students - that is, the teacher thought they were better than the students did. Table 4.12 shows the number of positive differences between the student and teacher ratings as well as the number of significant and highly significant changes (using the Student t-test analysis in Appendix 4.5) on individual questionnaire items made by that teacher between the two administrations of the questionnaire.

Table 4.12

Number of Positive Differences between Time One Student and Teacher Ratings,
Number of Significant and Highly Significant Changes between Time One and Time Two

	A01	A02	B03	C05	C06	D07	D08	E09	E10	F11	F12	G14	I17	I18	J19	J20
Positive Differences	13	40	21	40	24	NA	35	25	29	21	30	28	17	27	25	25
Significant Changes	7	6	2	4	3	2	5	1	1	1	2	1	2	0	0	2
Highly Significant Change	1	8	0	2	1	0	2	1	1	0	0	0	0	1	0	0

Spearman correlation coefficients were calculated between the total number of positive differences for each teacher (with the exception of the one teacher who had omitted to answer half of the questionnaire), against the total number of significant changes on the 52 questionnaire items, the number of changes at the 5% significance level, and the number of changes at the 1% level. There was a significant correlation between the number of positive differences and the number of highly significant changes a teacher made between Time One and Time Two ($r=0.5365$, $df=15$, $p=0.039$). That is, a teacher whose students rate them less favourably than they do themselves is most likely to make major changes that are reflected by highly significantly better student ratings. This result provides support to a number of earlier findings (Braunstein, Klein & Pachla, 1973; Centra, 1973; Erickson and Erickson, 1979; Gage, Runkel & Chatterjee, 1963; Pambookian, 1972, 1974; Smith, 1977) but it must be treated with caution.

The relationship involving the mean difference between each teacher's self rating and the class average rating on each item and the total number of significant changes a teacher made between Time One and Time Two was observed. As well, the relationship between the mean difference and the number of highly significant changes was explored. These are shown in Table 4.13.

Table 4.13

Spearman Correlation Coefficients between Mean Teacher/Class Average Rating Difference and the Total Number of Significant Changes and the Number of Highly Significant Changes.

Changes	Correlation	df	Probability	
Total Number of Significant Changes	0.5401	15	0.038	*
Number of Highly Significant Changes	0.7929	15	0	**

* $p<0.05$ ** $p<0.01$

That is, when on average the difference between the teacher's self rating and the average class rating is large, then there is likely to be a significant number of changes between Time One and Time Two in the student ratings. Also there is very likely to be a highly significant number of changes in the student ratings between the two evaluations. Put another way, if you want to know which teacher is likely to make significant changes between two administrations of the same student rating instrument then you should look for the teacher whose self ratings (on average) differ markedly from the average ratings of their students.

Student Reactions

The final part of the student Time Two questionnaire contained eight items intended to briefly assess student feelings about the whole research project. The eight questions covered whether the students thought that this type of evaluation was worthwhile, whether the teacher had thanked them for and discussed the results of the Time One evaluation, whether they felt good about having been asked to express their opinion about their teacher and the teaching in that class, to what extent they thought the teacher had tried to improve and had improved, whether any student had been "picked on" as a result of the first evaluation, and finally whether they would want to complete an evaluation of each of their teachers every year.

As a body, the students replied very positively to all of these items, with the least positive average rating (2.53) applying to the question about whether they thought their teacher had actually improved. Four of the items had average ratings less than two, and the other four were all less than 2.60. It was very reassuring to see that by and large the students felt that there had been no repercussions as a result of participating in the research. There was almost universal agreement between student and teachers on the three items they answered in common - 56 Discussed, 57 Tried Impr and 58 Did Impr. When asked whether they would like to complete an evaluation for each of their teachers every year, the students were of a more positive mind to do this than were their teachers who could probably see the enormous amount of time such an undertaking would require.

The tabulations for these items are shown in Table 4.14.

Table 4.14

Time Two Student Responses to Project Evaluation

Item	1	2	3	4	5	6	Mn	SD
53 Worthwhile	117	132	51	2	4	0	1.84	0.82
54 Felt Good	129	136	30	8	2	0	1.75	0.79
55 Thanked	66	123	81	20	14	0	2.32	1.03
56 Discussed	111	79	44	40	30	0	2.34	1.35
57 Tried Impr	80	119	69	22	15	0	2.26	1.07
58 Did Impr	48	107	108	26	17	0	2.53	1.03
59 Picked On *	17	15	51	82	140	0	1.97	1.15
60 Do For All	162	72	43	13	15	0	1.84	1.12

Time Considerations

To ensure consistent administration across all classes, a standardised procedure was followed with a prepared set of written instructions that was read to every class. This part of the procedure took between five and ten minutes depending on the amount of help a class required to complete the demographic data at the top of the questionnaire form.

Once the demographic details were completed, the time each class took to complete the rating part of the questionnaire was recorded in three steps. Records were kept of the time taken for the first student to finish, the first ten students in the class to finish, and the time taken for all of the class to finish.. Table 4.15 shows the times taken (in minutes and seconds) by the classes to complete the questionnaire. The Time Two results for Classes B04 and G14 are missing because of technical failure.

Table 4.15

Time Taken to Complete Time One and Time Two Questionnaires as a Function of First Student Finished, First Ten Students Finished, and All Students Finished.

Teacher		Time 1			Time 2	
	First Fin.	First Ten	All Fin.	First Fin.	First Ten	All Fin.
A01	6:57	10:53	19:56	6:14	9:06	13:21
A02	8:51	10:34	15:40	7:45	9:23	14:49
B03	6:23	10:35	16:13	05:57 *	7:55	15:27
C05	5:20	7:50	12:15	3:26	6:37	12:09
C06	8:35	9:55	12:19	5:03	7:52	10:39
D07	7:56	10:25				
D08	5:44			5:26	8:01	10:51
E09	7:45	11:57	16:50	6:06	7:52	14:15
E10	9:12	10:27	16:46	6:04	8:36	14:05
F11	6:07	11:12	18:23	4:38	9:44	14:11
F12	5:50	8:19	18:19	03:52 *	8:32	12:06
G14	7:49	13:08	13:24			
I17	7:41	9:35	14:57	7:07	8:38	17:10
I18	6:26	8:00	15:52	5:44	7:58	13:36
J19	7:16	9:37	13:20	5:29	6:50	9:11
J20	7:21	10:34		5:22	8:31	12:02
Average	7:03	10:25	16:03	5:35	8:15	13:08

* Denotes the first student to finish started before the completion of the administration procedure.

It should be noted that the questionnaires used in the two administrations were slightly different. The first administration had 52 Likert-scale rating items, with two free response items asking for comments on what the student liked about the class, and what suggestions they would make to improve the class. On average across all of the classes, 98% of the students completed these two free response items. The second administration had the same 52 Likert-scale items, plus a brief questionnaire about the research project. This consisted of eight Likert-scale items, plus one free response item seeking comment on any aspect of the project that the student wished to make. Approximately one-third (35%) of the students did not make any written comments on the second questionnaire, which would have reduced the amount of time it took them to complete the task. The improved times on the second administration could also be the result of respondent familiarity with the style of the questionnaire. On average, each administration of a 52 item questionnaire plus open-ended questions took less than 30 minutes.

Reliability and Validity

The matter of determining the validity of this process and this instrument was not addressed in this research through any statistical validation studies. Rather, the general question of the validity of student evaluations of teacher performance for development purposes was raised and answered through the literature review. It was argued in Chapter Two that student ratings were at least as valid a means of gathering data for teacher development purposes as any other method of data gathering, if not more valid than most. In addition, the procedures used to construct the questionnaire and administer it were in accord with the best practices outlined in the literature. To provide comprehensive validity data on the questionnaire used was beyond the scope of this research.

The data set was analysed for reliability in a number of ways. A pre-test/post-test correlation analysis of the response frequencies for each category produced the following results.

Table 4.16

Pearson and Spearman Correlations for Each Response Category
between Time One and Time Two

	Str Agree	Agree	Neutral	Disagree	Str Disagr.	Not Applic.
Pearson	0.97	0.96	0.88	0.96	0.97	0.97
Spearman	0.96	0.94	0.82	0.93	0.94	0.76

This gives an overall reliability rating of about 0.91 which is very strong. It is noticeable that from Time One to Time Two students moved slightly away from the Neutral category, and expressed an opinion rather than opt for the no-opinion ground for a second time.

When the two agreement categories were aggregated (that is, the Strongly Agree and the Agree responses were combined), and the two disagreement categories aggregated the correlations between the two administrations were similarly high. There was also a strong negative correlation between the opposite response categories. This indicates that the students were consistently using the appropriate end of the scale for positively and negatively worded questions, and were replying to the questionnaire in a serious way, and not frivolously.

Table 4.17

Correlation for Combined Response Categories between Time One and Time Two

	Agreement	Neutral	Disagreement	Not Applicable
Pearson	0.98	0.88	0.98	0.97
Spearman	0.98	0.82	0.96	0.76

Table 4.18

Correlation for Opposite Response Categories between Time One and Time Two

	Time	Agree v Disagree	Str Agree v Str Disagree
Pearson	1	-0.85	-0.66
	2	-0.82	-0.66
Spearman	1	-0.8	-0.66
	2	-0.77	-0.74

The opposite response category correlations were even stronger when the two agreement categories were combined and the two disagreement categories aggregated.

Table 4.19

Correlation for Combined Opposite Response Categories between Time One and Time Two

	Time	Agreement v Disagreement
Pearson	1	-0.92
	2	-0.94
Spearman	1	-0.95
	2	-0.95

These correlations indicate that the students were very consistent in the way in which they responded to the items in both administrations of the questionnaire. That is, the students were very likely to respond in a similar way to their teachers behaviours on all items on both administrations of the questionnaire, and were not capricious in their ratings.

Part Two: Qualitative Results

Time One Student Comments

The students were invited to write comments to two open-ended questions at the end of the Time One questionnaire. These two questions covered what the students liked about the class and what they thought could be improved about the class.

The student responses to these questions were content analysed to look for trends in their responses. The 52 items of the student questionnaire were used as the basis for coding, taking the form of a frequency tally. Where a student made a comment that referred to the subject of one or more of the items of the questionnaire, that comment was coded (counted) against that item. If the comment was not referring to one of the items but instead could be referring to one of the factors or dimensions, then the comment was coded against the factor. In addition, each comment received a positive (+) or negative (-) coding depending on the nature of the comment. The following three examples give an indication of the way comments were coded.

Example A:
> Q: What do you like about this class?
> A: The organisation and general encouragement

Here the student has made positive reference to the organisation of the class (Item 27+ Organisation/Expectation), as well as to the encouragement they receive from the teacher (Item 20+ Improve and Item 41+ Motivates).

Example B:
> Q: What do you think can be improved in this class?
> A: More team work and less talking. We need to move more quickly.

This response covered three of the items. The student wants more team work (Item 34- Team) and less talking - presumably by the teacher as there would clearly be more student talk if there was an increase in team work (Item 50- Talk a Lot). In addition they are commenting on the pace of the class work - this too refers to Item 50 which states that because the teacher talks too much the students don't get enough time to do their work - this too was coded as a second negative reference to Item 50, as well as the teacher not allowing the student to work at their own speed (Item 45- Speed).

Example C:
  Q: What do you think can be improved in this class?
  A: If we were treated more to our age level.

In this example, the implication is that the teacher has not adapted their teaching to the age level/maturity of the class and this is resented by the student. There is no one item that covers this interpretation of the student's comment even though there is an item that talks of the teacher adapting their teaching and language to the level of students' ability. Rather they are referring to the rapport that the teacher has with the class in a negative way. Therefore this was coded as a negative Rapport comment.

The codings for teachers were collated for each item and also summed according to the *a priori* factors or dimensions, which are presented in Appendix 4.6, while the summation for the *post facto* factors is presented in Appendix 4.7

The largest number of positive comments were coded for item 06 Happy (90 comments). This is supported by their questionnaire ratings, where this item had the greatest number of students agreeing with it. The item with the second highest frequency was 19 Chall/Intrst (74) followed by 52 Bst Class (52). There was a large drop off in the frequency of comments on other items after that. The items with the highest number of negative comments focussed on disruptive students in the class - 15 Disrupt. Std (41) with 19 Chall/Intrst (33), 28 Too Much (23) and 45 Speed* (23) following. It is interesting to note that while the students made the most positive comments about liking and being happy to work with other students in the class, they expressed the most concern about those students in their class who made learning difficult for them. This would suggest that although students like being at school with their friends, teachers would improve the learning environment in their classes if they attend to the problem of disruptive students. Most of the negative comments coded against 19 Chall/Intrst indicated that the students found the class "boring". Given 33

negative comments out of a total of 344 negative comments, this indicates approximately 10% of all students sitting in classrooms find the experience boring.

The *a priori* tables show that students made most positive references to the dimension Other Students (24.6%), then Learning/Value (19.8%), while the most negative comments went to Learning/Value (16.6%) then Workload (16.3%). On the *post facto* tables, Factor 2 Value of Learning (23.8%) and Factor 7 Student Contentment (20.0%) attracted the most positive comments. Negative comments were most common on the same two factors, Factor 2 Value of Learning and Factor 7 Student Contentment both with 16.3%. From both of these analyses, it would appear the Value of Learning factor is critical to the success of a classroom. Students realise and like the importance teachers give to this dimension, but are also concerned about some aspects of it.

To summarise, students like being with their friends in class, find class interesting and challenging and generally approve of their teachers and classes. Their teachers tend to make the subject easy to understand and do indeed have a sense of humour. On the other hand, they do have concerns about what happens in class and disruptive students figure prominently in their comments even though they seem less concerned with this in their numerical ratings. Boredom is a significant concern, and students feel that the workload and the speed at which they are permitted to work need to be tempered. It would appear that teachers are in control of their classrooms, without having to be unduly strict.

Time One Teacher Comments

During the consultation, the teachers were asked to make three lists of questionnaire items in response to the student evaluations results - items they were pleased with, items that required some comment but no further action, and the items that they were concerned about and from which their action plan was likely to emerge. The action plans were individually constructed according to the skills of the teacher, type of class and subject, and other environmental factors often unique to that situation. Therefore the most appropriate analysis of the consultations would be the number of teachers who listed an item in each of the above categories. The total number of teachers who listed an item is shown in Appendices 4.9 and 4.10, which also show the items arranged by *a priori* and *post facto* factors.

Teachers expressed pleasure with items in the Rapport dimension (14.8% of all items they were pleased with) and Learning/Value dimension (14.1%) of the *a priori* analysis. In the *post facto* analysis, items in Factor 1 (26.5%) and items not included in the factor analysis (26.1%) stood out. The single item that they felt most pleased about was 16 Asmt Fair (12 teachers) followed closely by 23 Try Hard (11), 46 Pronounce (10) and 06 Happy(10).

The most frequently mentioned items that the teachers felt worthy of comment but no further action required were 22 Hmwk Ovld* (6 teachers), 05 Prompt (4), 14 Strict* (4), 45 Speed* (4) and 33 Stdy Skls (3).

Items of concern were likely to be the source of an action plan for each teacher. Only eight items did not fall into this category. The rest of the items attracted some response, but no one single item stood out as being a major concern which required action by a large number of teachers. The most frequently noted items, all noted by five teachers, were 04 Clarity, 15 Disrupt. Std, 41 Motivates and 45 Speed*. Of these four items, only one, 45 Speed* figured in the most frequently targeted items. Indeed 45 Speed* was the most frequently targeted item (4 teachers), followed by 05 Prompt, 22 Hmwk Ovld*, 28 Too Much, 43 Cmnts Hlp and 44 On Time all with three teachers who wanted to improve their performance on these indicators.

When translated into factors or dimensions, teachers were most concerned about items in the Workload (14.5%), Assessment (13.7%), Rapport (12.8%) and Learning/Value (12.0%) factors of the *a priori* analysis. On the *post facto* analysis, most of the concern was focussed on Factor 1, the General Teaching Skills factor (25.6%) and items which had not been included in the factor analysis (22.2%).

It is interesting to note that the two overall assessment items, 51 Bst Tchr and 52 Bst Class, had two teachers in each case who felt that they were sufficiently concerned that their rating may need some attention. Although none of the teachers targeted these two items, it should be noted that there are teachers who are satisfied with nothing less than being every student's best teacher, and that they want to teach every student's best class.

Time Two Student Comments

The final part of the Time Two student questionnaire invited the students to make any comments about the research, their participation in it and how they felt being

involved. This meant that the comments did not lend themselves to analysis against the 52 items of the questionnaire. They did however fall into five main areas - the evaluation process, the questionnaire, their teacher, the students themselves, and aspects of the class and classwork. The codes for each class have been omitted from the quotations to ensure confidentiality.

The Evaluation Process

The overwhelming majority of the comments (68.5%) were about the evaluation process itself. Over a half of all comments (55.1%) were complimentary about the process and the perceived potential it has for better teaching and learning. Typical comments in this regard were:

> I feel it is important for teachers to know what the student has to say. The only way this can happen is through surveys such as this. This survey has helped my teacher to improve. All teachers should have this opportunity.

> I feel this is a good project as it lets our teacher know how we think her teaching can be improved, making her a better teacher and allowing us to learn more.

> These sort of evaluations are very easy to do and seem to be effective. I also enjoy doing them.

> I think it was good us being able to give reports to the teachers, not the other way round because sometimes you have a really bad teacher who can't tell that they really need to improve.

> It made a difference to know that our teacher was receiving feedback from us on his performance.

In particular, one of the aspects that was viewed favourably was the ability to contribute anonymously, and the freedom this gave the students to comment honestly. In the power relationship that exists between teachers and students, this was obviously a very important point. Given that parental complaints are otherwise the most common form of student expression about their teachers, and these can too often be traced back to a particular student, students are in a position where they are loathe to make any critical comments about their teacher for fear that their teaching

and results may be adversely affected. The guarantee of anonymity was obviously appreciated:

> I think its good to be able to get some things off my mind without having to go up to him in person.

> I didn't mind participating as we could be totally honest while still being anonymous.

> I think these evaluations can be really good because ...it can be anonymous at the same time - so teacher can't "pick on" who she/he thought might have said something against them.

Not all of the students' comments (3.6%) were as positive. Some of these were directed at the process, part at the time involved, and a third part at the lack of positive outcomes in the class. One bluntly expressed their opinion by noting:

> I don't understand this piece of crap. Its totally useless as it will not affect us.

Another student commented simply:

> Evaluations are too time consuming.

although this feeling was not shared by everyone:

> It doesn't take up too much time.

Mixed feelings were also evident:

> It did not show any improvement in teaching skills - if there were any changes it would have shown, so I think in most evaluations it's a waste of time if they are not going to improve from reading these things they should not bother at all. It was fun evaluating but disappointed with the result.

This last sentiment was echoed in another student's comment:

> Overall this evaluation hasn't changed anything in this class.

and one student went even further to illustrate the point:

> [Teacher's name] said after the first evaluation "I will improve" but that lasted the whole of half a week and then was forgotten. Moral: There's no point doing them if no results or solutions are found and used.

This comment suggests the need to use student evaluations judiciously and with a purpose rather than arbitrarily.

Apart from the compliments, the next most frequently commented on aspects of the evaluation process were the desire for more frequent evaluations (3.3%) and the wish to have had the opportunity to evaluate some of their other teachers (4.9%). Students often took the opportunity to elaborate on this last point quite strongly as evidenced by:

> I think this is a very good way to evaluate teachers. However [teacher's name] was not a good example because he is already a good teacher but this would definitely improve other teachers.

> I think that all teachers should be forced to take student evaluations - as some need a lot of improvement. Not just the ones who want too - as they are likely to be better in the first place.

> I wondered about the fact that only teachers confident about their ability as teachers would volunteer - I know a couple of my other teachers would not be too happy to have evaluations done on them and they would never volunteer.

and on the matter of more frequent evaluations:

> I think that there should be a regular evaluation of our teachers as they also regularly evaluate us.

> We should have an evaluation of all our teachers every year.

while one other student suggested "thrice a year".

A couple of students (0.8%) noted that it was difficult to comment on their teachers, and in one case the student offered an extensive caution to teachers concerning the use of student evaluations:

> I think it is important to realise through a survey such as this, for teacher in particular, that not everyone can be pleased. A teaching style acceptable for one student may not be appropriate for another. If the teacher seems to be teaching to an extremely high standard in the first place, by asking students to do an evaluations like this, they feel obligated to point out the negative aspects of a teacher. These negative aspects can be blown out of all proportion in the teachers mind, which would make the evaluation seem not worthwhile.

The Questionnaire

Only seven students commented specifically on the questionnaire instruments. One made a very generalised positive comment, four made critical comments, while a further two made constructive suggestions.

One student clearly was less than impressed with one question:

> Dumb questions hack me off such as the teacher tries to pronounce our names properly - who cares!

Another critic felt that the questionnaire reflected poorly on teachers and leaped to their defence, expressing their thoughts in this way:

> I feel your questions don't give a wide view, they are either positive or negative, you are persuaded by the questions rather than really thinking. You try and make the teacher sound bad.

The two constructive criticisms were:

> There needed to be a number like "sometimes" and some questions have two points that I could easily answer with two different answers.

> Most of the questions aren't specific enough for this subject. You should include one for each separate subject.

Classwork

In the Time One questionnaire, students were given the opportunity to comment on any aspect of their class that they liked, and to make suggestions to improve the class. This was not offered to them in the Time Two evaluation as there was no planned formal use made by the teacher of the Time Two results. However, eight student comments (3.3%) made suggestions to improve what happens in the class. These ranged from a plea to go easy on the homework, to reminding the teacher to slow down, and asking the teacher to tell the class about what was going to be in the exams!

Of real concern was the student who pointed out that "I was one of them" and drew an arrow to Questionnaire Item 59 ("I think that some students were 'picked on' by the teacher as a result of the first evaluation") to which the student had circled Strongly Agree. Although the student ratings for all classes indicated that this had not occurred in any other observable way, it is clear that this student felt that somehow the teacher had been able to identify a group of students in the class who had made some adverse rating or comment, and had singled them out for unfair and unethical attention. However, the ratings of the remaining members of this same class do not corroborate this comment, at least insofar as the suggestion that there was more than one person picked on by the teacher.

The Teacher

In all, 16.5% of the student comments were about their class teacher and their reaction to the evaluation process. Three quarters of them were brief and complimentary, either about their teaching or because the student had noticed an improvement:

> [Teacher's name] has improved since last term, our class now reckon that he has become more helpful and understanding, he's cool!

> The teacher has improved and she explains work much better.

The other quarter claimed to have noticed no difference, or made unfavourable comments about their teacher:

> Q No 14 My teacher is strict on uniform in which is not his duty but prevents us from working and takes up half of class time.

The Students

The final category of comments focussed on the students themselves (4.1% of all comments). Of the ten comments, six mentioned how the first evaluation had lead to direct benefits to them personally. In one case this was in stark contrast to the student who felt they had been picked on, but it does indicate the difficulty of administering a supposedly anonymous evaluation. One way or another, teachers can sometimes work out who was responsible for a particular comment and act on it, in this case beneficially:

> I was able to go through problems with [teacher's name] and told her what I thought because it was obvious the results I gave. It is good she is aware about how I feel and if this evaluation hadn't happened I couldn't have told her and there probably are lots of kids like that in other subjects.

The other four students wrote comments critical of themselves, such as:

> The only reason I may not cope with homework is not the teachers fault but my own laziness.

> I find this subject interesting and the teacher is helpful but I need to listen harder and try harder.

Teacher Post Analysis.

The final word belongs to the teachers who took part in the research, for the success of the research could be measured in terms of whether the teachers felt that this form of teacher development was useful and effective.

Appendix 4.10 contains the results of the teachers responses to the questions in the post evaluation. The evaluation was in seven parts covering aspects of the research - the questionnaire, written feedback, the consultation, teacher development, their reactions to the research project and their involvement in it, an evaluation of the project, and teacher appraisal and the place of teacher development in appraisal.

The Questionnaire

Teachers felt that by and large the questionnaire covered the tasks involved in teaching, while acknowledging that by the nature of the research it tended to be rather general. There was a hint that teachers would like to have a tailor made questionnaire that more accurately reflected the specific work and nature of the teaching that their subject involved. This was captured by one teacher who commented:

> It was very general - rather than subject/skill specific (due to the nature of research) - I'd also like to know how they feel about a variety of activities in class or how they feel about a particular topic. (J19)

This comment partly refers to student course evaluation which is rather more commonly used than teacher evaluation in schools.

While none of the teachers felt that student evaluation questionnaires were an invalid means of gathering data, there was not a unanimous view on the extent to which the questionnaires are valid. Two concerns caused this ambivalence, as reflected in these two comments:

> Students impressions are not entirely objective (A01).

> Personality and emotional baggage often influences a student's perception (J20).

Many of the features of the questionnaire appealed to the teachers. They mentioned the format, the feedback it provides, the scientific basis, the opportunity for free response comments, the student centredness of it, and its comprehensive nature. However, there were a few features that were not liked. Three teachers mentioned what they thought was the excessive length of the questionnaire, while two mentioned the fact that the students did not have to account for their own attitudes and work habits and how this impacted on the evaluations they gave.

When asked which questions they had found the most useful, almost all questions found their supporters - only seven lacked any support (Items 06, 14, 24, 37, 38, 39 and 46). The most favoured items were 41 Motivates (6), 04 Clarity (5), 19 Chall/Intrst (5), 08 Challenge (4), 11 Enthuses (4), 13 Impt Pts (4), 21 Asmt Crit (4), 27 Org/Expct (4), 40 S Ask Qns (4), and 48 Examples (4). On the other hand, the teachers found fewer questions that they classified as least useful - but these were

more closely aggregated. The least favoured were 39 Get Job (4), 09 Approachable (3), 10 Related (3), 14 Strict * (3), 22 Hmwk Ovld * (3), and 24 Ignore Tch * (3).

Teachers are divided in their opinion as to whether there are certain aspects of teaching that student should not be asked about. One third of them feel that students are competent to rate them and comment on any aspect of teaching, while over a quarter feel that there are certain aspects of teaching that students should not be asked about. Given the example of a teacher's subject knowledge, a typical response was:

They can certainly detect if a teacher does not know the work. (D08)

while another teacher commented on a general concern by saying:

I wonder if there is a certain amount of immaturity on the part of the students in the answering of this type of questionnaire. (A02)

On the whole, the teachers expressed qualified support for the type of questionnaire used in the research.

Written Feedback
All of the teachers regarded the written feedback as useful to some extent. This is hardly surprising given that the data was available and natural curiosity takes over in these matters. The reasons for this perception varied however with "my attitude to the results", "enables you to focus on necessary and useful strategies", "makes one do a little self reflection", and "I had already tried to think thru strategies" amongst the reasons proffered.

One of the teachers felt that student evaluation data had no credibility. However, 56% felt that the data had some credibility and the remaining 38% felt that student evaluation data was extremely credible. One of the reasons given for this echoed the Remmers and Weisbrodt quotation (1965) at the start of Chapter Two by stating:

The important aspect to remember is that the evaluation reflects, whether you like it or not, a student's judgment or perception of your performance (G14).

As regards which classes have more credibility, two-thirds felt that any class was suitable and credible, and the remaining third felt that only senior classes were

appropriate. This response had been expected and was expressed by one teacher as follows:

In some cases seniors take the whole thing more seriously. Juniors tend to love you or hate you and are biased (F11).

The teachers received the data in two forms - numerical ratings with tabulations and graphs, and the verbatim written comments of the students. 44% preferred the written comments, 25% the numerical data and the remaining 31% liked to have both. The teachers' written comments however tended to support the provision of both types of data, with the written comments appealing because:

Written comments could hone in on the specific area being commented on (C06).

and

I prefer information from text personally (F11).

One teacher likened the situation to being:

bit like a parent regarding the comments being more important than the grades (A02).

The teachers did not feel that the students' assessment of their teaching was inaccurate, but could not agree upon the extent of the accuracy of the ratings. In most cases the teachers found at least some of the responses curious and puzzling, but on the whole found the assessments to be fair and accurate.

As part of the consultation, teachers were advised to discuss the results of the evaluation with their class. This was intended to demonstrate that the students' efforts were valued and had been considered and discussed with a view to improvement. Three-quarters of the teachers said that they did this, with the remaining quarter not doing so although they intended to, often for good reason. In one case the teacher was ill for a period of two weeks immediately after the consultation and then suffered a family bereavement, and felt that the impetus and impact had been lost. The teachers' assessment of whether they had discussed the results with the class was not shared by the students in two classes. One teacher who had claimed to have

discussed the results with the class received a rating of 4.01 on a five point scale with a rating of 5.00 indicating that all of the students were strongly of the opinion that the results had not been discussed with them. Another teacher who had claimed not to have discussed the results received a rating of 2.53 which indicates that the students did feel that the teacher had discussed the results but were only lukewarm in their feeling about this. All other teacher and student responses about whether the teacher had discussed the results with the class were in accord.

Those who did discuss the results felt that there were very positive benefits from doing so. Typical responses were:

> It ... lead to more open discussion. Class tone has improved which I am delighted at. As a teacher I feel more comfortable with the class. (C05)

> They enjoyed the discussion ... It made the students feel they could speak more freely to me. (D07)

> Open and honest. It made them aware of the fact that as a person a teacher could find this exercise difficult/alienating and it could make them feel vulnerable (ie the teacher). (J20)

In all, the teachers found that the data did reinforce the way that they taught their classes, and also provided useful information to reflect on regarding their teaching.

The Consultation

All but one teacher felt that the consultation was helpful or very helpful in gaining benefit from the student evaluation data. The two words that came through in the teacher's comments were "perspective" and "focus". The data by itself was not enough. It required a collegial discussion to put matters into perspective and bring a sense of order to it.

The non judgmental nature of the consultation was appreciated by most of the teachers but in one case the teacher felt that this lead to a lack of support:

> Personally while I realised the researcher was being non-judgmental, that actually makes me feel uncomfortable - I think I might have liked a bit more positive reinforcement. (J19)

Receiving the printed results prior to the consultation was very helpful in providing time for the teachers to mull them over. This is important, for as one teacher commented:

I'd hate to have it presented to me at the consultation - I hate crying in public!!

All of the teachers found the written report to be helpful to at least some extent. The reports were adjudged to be objective and detailed, and were useful as a written record to refer to. In one case, it was suggested that this record could be used in discussion with the Head of Department if it was of a more general nature. For one teacher, "I found this the most useful".

As to whether the teachers in the research feel that some form of collegial consultation to complement the student evaluation data is desirable, there was an interesting diversity of opinion. Almost one fifth (19%) said that they did not feel that this was important, and the remaining 81% were almost equally divided between feeling that it was very important or of some importance. Of those who commented on this aspect, the comments tended to reinforce the notion that collegial support was important in providing balance and a variety of feedback on the results.

Teacher Development

The purpose of the research (to improve some aspect of teaching using information collected from students) was lauded as worthwhile by all of the participant teachers. As they said, it made them take stock of their performance, and created time for self evaluation which otherwise keeps getting pushed aside. All but one felt that the student feedback had lead to some improvement in their teaching but not extensively so. From the teachers' comments on this the major constraining factor was time. A typical response was:

There was not enough time. It has given me lots of ideas though. (I18)

However, when asked what factors lead to changes in their teaching practice the majority verdict was the student feedback. While this seems at odds with the previous question, it would appear that teachers have access to very little information on which to base considered changes, and the student feedback fills that gap.

Two key questions in the teacher questionnaire related to the extent to which the teachers felt that they had changed/improved their teaching. The first question asked

whether they had tried to improve the identified aspects of their teaching. All of them had tried to some extent, with 39% claiming to have tried very hard. Two of them cited reasons for not having been whole-hearted in their efforts and both of them referred to lack of time and the interference exams and holidays caused. Two others admitted to having concentrated on only one of the identified areas, but expressed the desire to deal with the others later. With reference to the second question, all teachers felt that they had actually improved, with 23% of the opinion that they had made considerable improvements. The two teachers (both of whom were deans) who had problems getting to class on time confessed to still having difficulty with this aspect of their teaching, as captured by:

> I am still late - I don't seem to be as late but I still don't make it on time very often. (F11)

The student reaction to these same two questions is interesting. With one exception, all classes felt that the teachers had attempted to make some improvement in their teaching. The one exception was a class that rated their teacher's attempts to improve at 3.31, with 1 being a unanimously strong feeling of attempting to improve and 5 a unanimously strong feeling that no attempt at improvement had been made. The teacher in this case had assessed that they had made somewhere between a "considerable attempt" and "some attempt" to improve. When the class average ratings on this item are arranged in order from the most positive rating to the lowest (from 1.39 to 3.31), the five teachers who were emphatic in their belief that they had attempted to improve were all in the top six. That is, the teacher said that they had made an attempt to improve and this had certainly registered with the class.

On the second question, fourteen of the sixteen classes felt that their teacher had actually improved during the period of the research, but the students were not quite as sure about whether improvement had actually occurred as they were about whether the teacher had tried to improve - mean ratings ranged from 1.89 to 3.31. The three teachers who were sure that they had improved all occurred in the students' top eight ratings.

The final question in this section asked the teachers to assess the extent to which student learning had been enhanced as a result of the student feedback process. One teacher felt that student learning had been greatly enhanced, while two teachers felt that the impact on student learning was non existent. The remainder felt that there

had been some positive effect. A sense of resignation and cynicism about the likely effect on student learning is typified by this teacher's comment:

> I'm not sure that it makes an awful lot of difference to them. (J20)

The abiding impression is that this process of teacher development does have some positive benefits, but does not provide the answer to all of the teachers' concerns nor lead to immediate and lasting improvement.

The Teacher Reactions

The reaction of the teachers to participation in this project was a mixture of pleasure and concern. On the one hand the adjectives interested, positive, worthwhile, useful, pleased, enjoyed, and appreciated conveyed the delight of most of the teachers in their participation. On the other hand, four of the teachers used adjectives like dismay, sadness, blase, frustrated, and nervous to describe their feelings. Two of these teachers however feel that they can see the worth of the process. In their comments, two teachers referred to the fact that this research had given them the opportunity to do something that they might not have otherwise for lack of time or lack of courage:

> I think it is worthwhile and without it I would not have known it was an area of concern to me but other things would have taken priority. (D07)

and

> Enjoyed the experience. Mainly appreciated the student feedback - I might not have had the courage to do it on my own initiative. (J19)

Receiving the results was a nervous time for some of the teachers but at least ten of the teachers were pleased and affirmed by the results - it was not as bad as they had perhaps led themselves to believe. One of the teachers however felt dismay, horror and sadness as:

> I thought I had a much better rapport with the class than the results suggest. (A02)

This particular teacher felt that the effect of this information was detrimental to job satisfaction and morale, but otherwise the teachers' response was one of affirmation of their already good job satisfaction. As one teacher put it:

I am quite happy and confident in my job and the results were fine. (F11)

Overall, the responses of the teachers to participating in the research and this type of evaluation exercise were positive.

Project Evaluation

The benefits that the teachers saw in this type of development process were many. Among the various points made, they appreciated the positive and affirming feedback that they received, the comprehensive coverage of the teaching function, the pinpointing of weaknesses that needed attention, the ideas for improvement that emerged from the process, the candidness of the students, the ability to see just what the students really did think of their teaching and the positive outcomes in terms of a closer relationship with their class. For one of the teachers it provided the opportunity as a Head of Department to gain an insight into a viable professional development option. The disadvantages can be summed up in two words - time and fear. Eight teachers commented on perceived disadvantages and five of these referred to the amount of time required. The other three referred to the possible effect upon morale and their fear of receiving the results. Therefore the few suggestions made to improve the process used in the research mainly focussed on the need for more time. The only other suggestions called for more written feedback from the students (one teacher), more direction with improvement strategies (two teachers), and in response to a later question the addition of a training component to the whole process.

On the matter of whether a process such as this should be included in a school's teacher development programme, there was overwhelming support from 78% of the teachers, and some support from the rest. However, there was a considerable cynicism as to whether it was possible to gain teacher support for such a programme. This was founded on several factors:

Some teachers could feel threatened with close examination of their teacher performance ... (C05)

There is little incentive for becoming a better teacher. (D08)

I think that some teachers would have no interest in what students think and some would find it quite threatening. (F11)

It would be hard to get it. (I18)

It could be a shock to find out what ever some pupils think ... (J19)

Teacher Appraisal

The final section of the post-evaluation by the teachers asked whether student evaluations of teaching should (or should not) be included in the appraisal of teachers. With schools now obliged to have in place a teacher appraisal programme, is there a place for this type of evaluation in those programmes? To establish a baseline, the teachers were asked to give their opinions about teacher appraisal. Exactly one half thought that teacher appraisal should be for formative (teacher development) purposes, while the other half thought that appraisal could serve both formative and summative (teacher evaluation and personnel decision making) purposes. None of the teachers thought that it should be for summative purposes alone. These comments captured the prevailing opinion:

> This sort of questionnaire, I think, solely has value as teacher development - I would not put myself at risk with teaching strategies if it were going to make a difference to my prospects of a bonus or promotion - could destroy important collegial aspect of teaching. (J19)

> Could be dynamite in the wrong hands, especially if used for the evaluation of teachers! I have had dealings with senior personnel that I would hesitate to trust with an evaluation of this nature. (F12)

Another reason for this could well lie with the expressed concern of another teacher:

> Kids have some ideas but not enough to have this sort of power. (F11)

However, these feelings have not put most of the teachers off the possibility of including student evaluations in a teacher appraisal scheme - 87% would include student evaluations but there are some reservations, with one teacher emphatically rejecting any further involvement with student evaluations. The teachers would not like to see this methodology used as the sole means of appraisal, but it could be offered as one of a series of options for teachers to use. One teacher offered a reason for this when summarising their views on appraisal:

I believe that teachers should be involved in appraisal whether it be peer or student. I tend to support peer with some student because if it was to be done across a staff then students would lose a lot of teaching time if it was all classes. (D07)

Just over two thirds of the schools have an appraisal scheme, and participation is compulsory in 82% of those schools. In 92% of the schools having an appraisal scheme, the purpose of the scheme is solely formative, with appraisal serving both summative and formative purposes in the other school. Therefore, although they expressed the common concerns of teachers about the use of appraisal for summative purposes, this group of teachers are more accepting of appraisal serving both purposes than are their schools. Given that all of the teachers were volunteers who presumably were at ease with students evaluating their teaching, this is not entirely surprising. This is reinforced by their responses to a question which asked them whether student evaluation data should be made available to course supervisors/HODs/senior management. Three of the teachers were happy for their superiors to have free access to this information, another three felt that their supervisors should never have access to the data, while the remaining ten teachers felt that their own discretion should determine whether supervisors had access to the student evaluation data. The following reason:

It would certainly help HOD if staff member having a real problem. (J20)

was supported by one other.

Comparing this method of teacher evaluation and development, the teachers liked the provision of feedback in a non-threatening positive manner; the personal, one-to-one nature of the interaction; the student focus; and they believe that the information has more validity and credibility than data from other observers. This latter opinion was expressed by two teachers who commented:

I think it is correct that only students see what you are really like. Teachers are different around other adults. (I18)

Much better than being observed where rapport with class never the same. (J19)

One of the positive features mentioned by one teacher was that this method was less time consuming than other methods they were familiar with. This was borne out by the 80% who thought that the time demands of this form of teacher evaluation were reasonable given the potential outcomes, while none of them felt that the time required was unreasonable. They suggested that by cutting down the size of the questionnaire, and the scope of the exercise the time requirements could be quite manageable. Other time management controls would be to have only a sample of classes (or even just one class) perform an evaluation on any one teacher, and to have the evaluation performed only once a year (64%). Two teachers thought that student evaluations should be collected and acted on every term, while three others felt that every two years was adequate.

Given a free rein to implement a teacher appraisal scheme which included student evaluations, the teachers were asked to list three features they would include and three features they would exclude. Many suggestions were made by a single teacher (for example, delete this specific question, keep the thorough briefing of the students, more open-ended questions, have an assessor in your subject area) but three themes emerged. The first essential feature to retain was the confidential nature of the student responses, and the second was the overall design viz student questionnaire, feedback/consultation, action plan. The third theme was strongly expressed to exclude any use of the process for summative purposes. Typically they objected to:

> Principal having right to information.
> Use for salary or hiring and firing.
> Being used by superiors to judge the merit of a teacher.

The message was clear in this respect. Teachers are fearful of summative evaluation and do not want student ratings used for this purpose.

In summarising their participation, seven expressed the belief that this was a worthwhile and valuable project to be involved in, and another three took the opportunity to comment again on the time pressures that teacher evaluation takes and the need for schools to budget support, time and funds for this purpose if they want it to be done properly. Finally, 75% agreed that student feedback is an important source of feedback for improving teaching.

Summary

Data were collected from both students and teachers throughout this study. Comparisons of the data have been made to triangulate the responses and verify them or highlight differences. The reactions of the participants have also been taken into account, and these used to consider the merits of a teacher development scheme which uses student evaluations as an integral part.

The raw numerical data were treated in three ways to see if the teachers improved in their teaching as a result of the feedback/consultation intervention. This involved analysing the size effect coefficient d; Student t-tests on the gain scores for each item between Time One and Time Two; and, Student t-tests on the factor scores between Time One and Time Two. In each case, improvement was noted, although the results of the factor t-tests were not as strong as might have been expected (Marsh and Roche, 1993).

The teachers' self-evaluations were also compared with the ratings that the students gave to their teachers. These comparisons were particularly interesting for they revealed the group of teachers who made the most dramatic improvements over the period of the study. Where teachers gave themselves good evaluations but the students in their class did not, the teacher made the greatest improvement.

Quantitative data from both the students and the teachers were also compared, and these supported the conclusions drawn from the numerical data. Students and teachers can see the value of using student evaluations of teacher performance to aid teachers in their quest for better teaching and learning in the classroom.

Chapter 5   Discussion

This research was modelled on Wilson (1986) and the many works of Marsh and others (particularly Marsh and Roche, 1993 and 1994) and was designed to show that student evaluations coupled with feedback and consultation could be used in a teacher development programme for the purpose of improving teaching and learning. Wilson (1986) developed the feedback/consultation model and showed that important changes in overall teaching effectiveness occurred for most of the teachers in the study, and that the greatest number of changes occurred where suggestions from the consultation were concrete, specific and behavioural. He was convinced that the key element was the personal face-to-face contact that the consultation involved, as he was able to show that a similar process that used printed suggestions for improvement did not have the same effect.

Marsh and Roche (1993) stated that their study showed varying degrees of support for the contention that feedback from an Australian version of the SEEQ instrument coupled with Wilson's (1986) feedback/consultation intervention provides an effective means of improving university teaching. They especially noted that this effect was greatest for the teachers who were initially least effective, and that the greatest improvement was shown for those areas which the teacher targeted as the focus for intervention. In addition, they also noted that feedback was more effective after a longer rather than a shorter period of time.

In examining the reasons for apparently modest effects, they speculate that the novelty of student evaluations and the lack of familiarity with the processes may have detracted from the effectiveness of the intervention. L'Hommedieu, Menges and Brinko (1990) identified a "John Henry" effect that could be at work in student evaluation programmes. This could work in two opposing ways to moderate the expected effect - the control group could become a "compensatory rival" to the experimental group, or they could become "resentful and demoralised". In the first case, the control group who are not involved in the intervention try to compensate and improve without the intervention, thus ameliorating the effect. On the other hand, they could become demoralised and resentful because they are denied the attention that the experimental group receives, but get a shock when they receive their

student ratings and make marked changes. This has the effect of again moderating the overall effects.

The main finding of this research in secondary schools provides modest support for the findings of Marsh and Roche, and of Wilson. As observed and recorded by students, teachers do improve their teaching when feedback from student evaluations coupled with consultation is used for teacher development . It is an effective means of improving classroom teaching. The qualitative indicators from students and teachers show that the process was useful and effective, and both groups felt that improvement did occur as a result. In each of these areas the students and the teachers are in agreement in their belief that this has occurred, and the statistical analyses generally support this contention.

The Student t-tests of the individual item gain scores show that there was a significantly greater improvement than could be expected by chance. T-tests on the *post facto* factor scores indicate that improvement occurred beyond chance expectation but not significantly so. Contrary to Marsh and Roche's assertion, it would seem that the factor scores were not as sensitive as the individual item scores in measuring any changes that occurred during the research. Neither did the overall rating items demonstrate any significant positive change. However, further support is provided by the size effect coefficient, d, which was used in reporting the results to the teachers after the Time Two administration. These coefficients show that there were substantial changes in both directions as measured by the student ratings, with desirable changes outweighing undesirable changes. Changes on the targeted items that teachers chose to focus their energies on indicate that there was also some change in the desired direction, but this was not as strong as might have been expected given that the teachers were concentrating their efforts on those items. This is shown by both the t-test results and the size effect coefficients, d.

Together, these three quantitative measures indicate that there was change and that the change was in the desired direction, but it was not as substantial as might have been expected from the literature. There are a number of reasons for this, some related to the design and others to the problems of this type of analysis. Perhaps most significantly, the research involved sixteen classes and teachers in nine different schools, and covered eight different curriculum areas over five different class levels. This meant that there was already a substantial amount of built-in variation which could affect the impact of the research procedures, and washout the measurable effects that were intended. The fact that improvement was achieved under these very

general conditions could be regarded as an important outcome. Future study involving less variability in the research subjects would provide a more definitive answer to the impact of this type of intervention, but that approach would lessen the general application of the findings which this study provides.

Non quantitative methods were also used to support the contention that improvements were made by the teachers. Both the teachers and the students felt that there had been an improvement in the teaching behaviours of the teacher that would lead to better teaching and learning. When asked whether they thought their teachers had tried to improve, the students mean rating was 2.26, with a substantial majority (65%) of the students feeling that their teacher had indeed tried to improve, and in response to whether they thought their teacher had actually improved the mean rating was 2.53, with 51% expressing affirmation of this. All of the teachers felt that they had improved to some extent, and all felt that they had made some effort to improve.

Some of the measured changes were subtle. There were more "not applicable" responses by students on the Time One questionnaire than on the Time Two questionnaire. Some of the questions would naturally attract this response from students - for example, the item regarding teachers asking boys to respond to questions and the items specifically related to homework. For 30 of the remaining 46 items, the "not applicable" responses decreased by an average of almost 3 students per statement. This could indicate that teachers have been alerted to some aspect of their teaching that was not specifically targeted in their action plans, but for which they have made some small improvements in their classes. An alternative explanation is that some aspects of classroom teaching develop during the year and do not become apparent to students until then. This decrease in "not applicable" responses is in accord with the trend observed by Marsh and Roche (1993: 248) when conducting a similar research design involving feedback supplemented by consultation. Similarly, the reduced variability from Time One to Time Two of student responses as indicated by the standard deviation of items could indicate that teachers had made changes that were evident to students and hence narrowed their responses to individual items.

Another important finding is to note the group of teachers who made the most significant improvement, and to offer an explanation for this consistent with a theoretical psychological foundation. It was hypothesised that where students gave their teacher a less favourable rating than the teachers gave themselves (that is, the teacher thinks they are better than they are according to their students or the "Gee, I'm really not as good as I thought I was" situation), the teachers would be more

motivated to make changes in their teaching and classroom behaviour than those teachers whose ratings closely agree with their students or who had given themselves less favourable ratings than their students (that is, teachers who think they are not as good as the students give them credit for, or the "Wow, I'm really better than I thought I was" situation). Where there is close agreement, the teacher sees no need to change - the students' ratings will be accepted as credible and accurate because they are in accord with the teacher's own perceptions. Where the students rate the teacher more favourably than the teacher's self rating, the teacher will believe that matters are a lot better than they themselves had expected, so again there is no need to change. However in the case where student ratings are less favourable than the teacher's own self evaluation, the teacher will be more motivated to change their behaviour so that the student ratings match their own, or the teacher will endeavour to persuade the students to change their perceptions (and therefore their ratings) without actually changing their teaching behaviour, or the teacher will dismiss the results as irrelevant or inaccurate thereby avoiding the need to do anything about them. These findings are premised on the theory of cognitive dissonance (Festinger, 1957) and developed later as equilibrium theory. That is, the teacher tries one way or another to get the two mismatched perceptions (the students' and their own) into equilibrium.

The results of this research are in accord with this hypothesis. Those teachers who rated themselves better than the students did on the Time One evaluation were the ones who made the most significant favourable shifts in later student ratings. These teachers worked very hard to change aspects of their teaching to ensure that the student ratings were more favourable on the Time Two evaluation, rather than trying to convince the students that their assessments were wrong or misguided. The third option of dismissing the results as not credible did not appear to occur amongst the teacher post evaluation responses, except for one teacher who did not fit into this category anyway. This result provides support to a number of earlier findings (Braunstein, Klein & Pachla, 1973; Centra, 1973; Erickson and Erickson, 1979; Gage, Runkel & Chatterjee, 1963; Pambookian, 1972, 1974; Smith, 1977) but it must be treated with caution because the set of teacher self-ratings was not as complete as the set of student ratings. A very profitable area of future research would be to explore this hypothesis in other secondary schools.

Student characteristics do not contribute in a significant way to the ratings that students give their teachers, apart from those aspects of teaching that are directly related to that characteristic, according to the results obtained in this study. For

instance, student ability does affect the ratings given for workload and difficulty, and teacher organisation. One student characteristic, however, was quite significant. The ability for students to obtain help with their schoolwork at home was a significant factor in the way that students rated their teachers. This characteristic affected all but one of the factors found in this study, and that was Factor Six, Student Cooperation. This is not surprising as that is the factor that describes the student's relationship with other students in the class which clearly does not impinge on homework. Students who could obtain support from their family to complete work set for home rated their teachers and classes quite differently from those who could not count on such support. No other student characteristic had the same impact on ratings. Student characteristics and their effect on the ratings that the students give their teachers have been well explored in tertiary settings and would make an interesting area for further investigation in secondary schools.

The Research Design and Student Questionnaire

Earlier analytical studies mentioned in Chapter Two highlighted certain problems in student evaluation research designs that seem to consistently emerge. These problems affect this research too. First, the paradigm used was a pre-test/post-test design. Feedback was obtained from the pre-test, provided to the teachers in an experimental group and the effects of the feedback measured by comparing the results of the post-test with the pre-test results. This design has the disadvantage that it cannot control the reactive effect of the pre-test on the student raters - that is, it is not possible to tell whether the results on the post-test are directly the result of the feedback and any changes the teacher may have made, or whether the results on the post-test could have resulted from a knowledge of the items used in the questionnaire and the students reacting to them rather than any discernible changes made by the teacher.

Secondly, this study used an instrument that employs a Likert scale with a five-point scale. This causes two difficulties. There is an artificial ceiling to the rating that any teacher can receive thereby distorting any possible measurable change that the scale can reflect, especially as the pre-test ratings are typically above the mid-point (Marsh and Roche, 1993) and leave little room for improvement. This problem of the artificial ceiling was particularly true for Teacher J20, whom the students had found hard to fault, and the students did indeed rate their teachers above the mid-point on the pre-test. In the view of L'Hommedieu et al (1990), this is the single most important factor limiting the measurement of student rating effects. Furthermore, the

very fact that the ratings are reliable in terms of their test-retest stability means that there are questions about the sensitivity of students to any change that may occur and their ability to record that change on a five-point scale (Rotem and Glasman, 1979). While this may be true, there was sufficient measurable change for the researcher to conclude that students used the scale sensitively. A third concern was the size of the samples used and the power of the statistical analysis to detect any significant effect. Cell sizes in the analyses have frequently been less than 20, which reduces the power of the analysis to detect change. At least 30 teachers in each treatment group is regarded as an acceptable minimum. One survey noted that the median cell size in all of the studies they analysed was 16 (L'Hommedieu et al, 1990) with some groups as small as 6. Fourth, the use of gain scores was fraught with difficulties as a measure of change primarily because the gains may be systematically related to any random error of measurement, particularly when considering the targeted items. Typically, targeted items are the items that teachers received poor ratings for. Regression to the mean would account for some of the desired changes that occurred, and chance would also account for some of these changes. The research design used is not particularly strong for countering this difficulty. It is also important to recall that with a Likert-type scale, it can be difficult to measure changes particularly on items that have scored well in the first place. Indeed, if the teacher is already exceptionally good and is scoring near the most favoured end of the scale, the scale does not have much tolerance left to reflect any possible improvement. This was particularly true in the case of Teacher J20, whom the students found hard to fault. Finally, the length of time between ratings may not be enough for any effect to show in the ratings. In this study the interval between ratings was between eight and thirteen weeks. This relatively short time span could have lead to a problem identified by Wilson (1986). Teachers developed an action plan through a consultation phase but the outcome was merely knowledge of better ways to do things rather than the ability to do or practice them systematically enough for any planned improvement to take effect.

To standardise the research, a single questionnaire was designed and used. This meant that it had to be sufficiently generic for it to work in (almost) all possible classroom situations. In doing so, it was not possible to include specific aspects of teaching that are peculiar to one or two teaching areas - a point noticed by at least one of the students in their comments. For example, it did not include any items about the management of laboratory or workshop type situations. Teachers in schools would not be advised to use this questionnaire in its totality - rather it would be advisable for them to select from an item bank sufficient items to make a questionnaire suited to the purpose of the evaluation. For example, if the purpose of

the appraisal was to develop communication skills, then a large number of items could be included from the item bank on this aspect along with others from the general pool. It may be that certain items could be used in all situations as the basic questionnaire, with teachers having the ability to select the extra items that they wish to use. A similar scheme operates for the Massey University Student Evaluation of Teaching (SET) programme trialed in 1995. In this way teachers can focus on particular areas of interest or concern.

The Consultation Conference

There was strong support from the teachers for the form of the consultation and the outcomes from it. Two thirds of the teachers thought that the consultation helped them to benefit from the data provided by the student evaluation. What did appear to occur was that the teachers were able to use the feedback and consultation to put into perspective several aspects of their teaching, and this helped them focus and clarify issues of concern. This was an example of reflection in action.

Almost one in five of the teachers do not believe that it is necessary to complement student evaluation data with some form of collegial consultation. Of those who do, they felt that a colleague helps to bring a balance to the process, and opens up a variety of approaches when teaching is viewed in other professional ways. While the consultation was seen as being not so useful for supporting teachers, this is not at all surprising. The nature of the work done in the research did not make it possible for the teachers to be supported in their efforts to improve. The teachers were left to their own devices to action the points discussed in the consultation, and no support was offered, yet 72% of the teachers found that they gained some support from the conference. It is difficult to read too much into this and it may be that teachers do not have an expectation of support when engaged in forms of this kind of research. In a school appraisal system, support from colleagues is a vital component that should not be overlooked.

On other measures the consultation was judged to be a success. Two thirds of the teachers found the collaborative, non-judgmental nature of the consultation very helpful, although one teacher commented that perhaps it was too devoid of positive reinforcement thus causing them to feel uncomfortable. One teacher did comment that although they were able to deal with the less than flattering responses without reassurances from the researcher, they felt that other teachers might find this difficult. On the other hand, others expressed appreciation of this style as it allowed them to

decide for themselves, was very impartial and focussed, was far more constructive and was perhaps the only valid way.

It would seem that teachers often operate in a vacuum, and use informal means of gaining feedback on their performance. This may be simply that the absence of any criticism means that they must be doing a good job. Another explanation is that the traditional visits of the Principal, Head of Department or even the Inspectors of old provided them with no hard data about the quality of their teaching but rather platitudes such as "I like your teaching - the students seem to like being in your class too" or criticisms such as "You will need to watch their homework more" which leave the teacher with no idea what they can do to improve. Student evaluations provide some numerical data and comments designed to pinpoint strengths and weaknesses and enable the teacher and a colleague to devise strategies to turn the weaknesses into strengths.

Reaching out for some tangible form of evaluation like this may also reflect the essentially solitary nature of a teacher's work - when it comes to analysing classroom performance most teachers have only their own experiences to call on, and even fewer resources and support mechanisms. In the absence of any other source of "objective" data, it seems that teachers are ready to accept any useful information about their teaching. Therefore, student feedback was a key element in helping them to develop an action plan for improvement. The information the teachers received was perceived as being useful because it was specific about some aspect of teaching that they were able to exercise some control over. They could do something about it.

Teacher Appraisal and Evaluation

While the primary purpose of this research is teacher development, there is no escaping the issues of teacher accountability. Since the advent of Tomorrow's Schools (Lange, 1988), there have been significant changes which have imposed formal accountability on educational institutions and teachers (Department of Education, 1990a, 1990b). One of the most recent of these changes is the accreditation requirements for schools wishing to teach units registered on the National Qualifications Framework, while the Draft National Guidelines for Performance Management in Schools (1995) have been published for comment this year. Therefore, teachers were asked to consider the process under investigation within the context of appraisal systems concerned with performance and accountability as well as development.

Over two-thirds of the schools had appraisal schemes, and almost all of them were focussed on formative appraisal - that is, focussed on teacher development, not personnel decision-making. Therefore, the teachers had been exposed to systems of appraisal. Only two of them felt that student evaluations had no place in an appraisal scheme One of those was a teacher who received one of the least favourable student evaluations and expressed a preference for peer appraisal because peers at least have a knowledge of the teaching process. As a result of their participation, very few changed their minds about the inclusion of student evaluations in teacher development programmes.

The teachers in this research had a clear and sceptical view of student evaluations and their use for performance management when conducted in an environment to support performance pay or personnel decision-making. This merely reflects teacher attitudes based on observations overseas (Barber, 1984; Glass, 1990; Good and Mulryan, 1990; Knapp, 1984) that found that any competency system linked to performance pay in teaching produces insecurity and defensive behaviour, and can lead to "suspicion, acrimony, inflexibility, cheating and finally control" (MacDonald, 1976, 238). This view also reinforces the assertion that "any data that can be used summatively will be" (Barber, 1990, 217).

However, the teachers in this study could see potential benefits in using student evaluations for the data collection phase of a performance management appraisal. In this way, the teacher and the appraiser could enter into an appraisal cycle without the "classroom observation" material being called into dispute - the appraiser has no vested interest in the data used for the appraisal. Furthermore, the time taken to collect the data was acknowledged to be more economic than the use of a teacher-observer. The collection time of about half an hour was less than would be necessary if an appraiser were to attend the class and collect objective information that is more than "one-off" data.

There appears to be no literature on the cost effectiveness of student ratings as a means of teacher evaluation. Analyses of costs and benefits are well established as a means of evaluation, particularly in the modern business world, but they are notably absent from the realm of teacher evaluation which uses students as the source of information. For teacher evaluation, the use of other teachers to conduct evaluations is a significant cost to the system. The teachers who conduct the evaluation are frequently senior teachers and the cost of their time is great compared with the

alternative of using a class to provide suitable data for a teacher development programme. If the data provided by students is seen as being valid and reliable, then this alternative is more cost effective than using senior teachers to provide this information. This whole area of the cost effectiveness of student evaluations would benefit from further study.

Two recent occurrences will impact on teacher attitudes to teacher evaluation. The first is the new requirement that all schools implement a performance management programme involving a formal appraisal linked to roles, responsibilities and accountabilities (Ministry of Education, 1995). The second comes from an unexpected source - the recent settlement of the teachers' collective employment contract. Teachers will no longer automatically progress through the basic salary scale. Instead, a process of attestation will be required for a teacher to progress to the next step. Each year, the school principal has to attest to the competence of the teacher before that teacher can secure the next step on the salary scale. In part, it will be a default mechanism - a teacher will progress unless the principal makes a decision to withhold that step from the teacher. This can only be done if a performance management appraisal has taken place in the past twelve months and clearly shows that the teacher is under-performing. Therefore, under this system there is a link between the appraisal of a teacher and their pay and tenure, although the link is not one that establishes payment for enhanced performance but progression through the pay scale upon meeting a minimum level of competency.

Should student evaluations be used when the stakes are high? The answer is a yes and a no. No, student evaluations should not be used as the sole means of obtaining data for the evaluation of a teacher. To do so is to place too much faith on student evaluations. This research was premised on student evaluations being used for teacher development and improvement. The students were aware of this objective and could see something in the exercise for them - the likelihood of better teaching and learning. If students knew that their evaluations were to be used for pay and tenure purposes, then they may respond differently. Yes, student evaluations can be used as *one* of the tools of teacher evaluation. Alongside other tools of evaluation (for example, peer evaluation, self evaluation, portfolios, diaries, interviews), they provide corroborative (and sometimes indicative) information.

Teacher and Student Reactions

Only six of the students (out of 306) felt that the whole process was not worthwhile. This provides overwhelming support from the students for the opportunity to evaluate their teachers in this manner. Similarly, only ten students claimed that it did not feel good to be asked for their opinion via student evaluations about the teaching they receive. By and large, the students therefore feel that their opinions are being listened to, and appreciate the opportunity to air these in a confidential manner.

Students feel good about being at school, being in class and with their friends. They are more certain and positive about the affective climate in the classroom than they are about matters in the cognitive domain and their learning as a result of what the teacher does in the classroom. The students feel that they try hard at school and are valued by their teachers. Teachers, according to the students, are in control of their classrooms and are able to do this without being unnecessarily strict.

What influenced the teachers to change? The student feedback provided information that was sufficiently at variance with the teacher's assessment of how they were teaching the class to make the teacher feel uncomfortable. This is in accord with the theory of cognitive dissonance espoused by Festinger (1957) where a person tries to achieve equilibrium between external information and their own perception of a situation in one of three ways - the person can adjust their behaviour so that it matches the external perceptions; the person can try to make the external perceptions change to match their behaviour; or the person can dismiss the external perceptions as being inconsequential and therefore there is no mismatch between their behaviour and perception others have of that behaviour. None of the teachers chose the third option of totally dismissing the students' perceptions although there were those who questioned them with respect to a few particular points. The student feedback prompted teachers to act in one of the first two ways. There were those who found that the feedback focussed on areas that needed attention, with the effect that they endeavoured to change their own behaviour on those areas and thereby bring the behaviour into balance with the perception.

Some others chose to take the second way of reducing the dissonance between their own perception and that of the students. This involved talking to the class and explaining their actions so that the class had a better understanding of what the teacher was trying to achieve. This was a more common response and indicates that the teachers tried to reduce the dissonance by changing the student perceptions, and

not by changing their own behaviour with regard to that aspect of their teaching. This would indicate a belief that what they were doing was right and that their actions had been misinterpreted by the students. Therefore, the teacher tried to correct the impression that the students had of their behaviour, not the behaviour itself.

Implications for Future Research

Some changes in the design of the study would be in order for future research purposes. In addition to the design problems outlined above, the development of an "ideas" booklet described by Marsh and Roche (1993, 1994) would seem to be a positive way to improve the consultation service to teachers. In discussing their strengths as revealed by the student evaluations, teachers would contribute to the pool of knowledge that teachers in need of help in that area could use. In this way, they would be depositors as well as drawers from the bank of knowledge. The "ideas" booklet would mean that during the consultation the teacher would have available the best practices of many other teachers who are described by their students as being particularly good in that attribute.

Much work still has to be done in the secondary sector of education, and additional research in the primary sector would be revealing. One of the myths that secondary teachers intuitively hold is that secondary students are not mature enough to make useful and valid evaluations of their performance. It is likely that primary teachers would also hold this view, and research in this area would be fruitful.

Summary

The questions to be answered relate to the place and effectiveness of student evaluations of teacher performance in a teacher development programme. Are student evaluations valid and reliable? Are they useful and credible devices for gathering data about teacher performance? Are the time demands of gathering data in this way too great? What can we learn about secondary teachers and secondary classrooms as a result of asking the students what they think about their teachers? Do teachers improve their performance as classroom teachers as a result of feedback from student evaluations augmented by consultation? Are teachers prepared to accept student evaluations as part of a formative appraisal scheme within their schools? And finally, should student evaluations be used for formative or summative purposes, or both?

The quantitative and qualitative data were used to triangulate a situation in which students and teachers had positive views on the place and usefulness of student evaluations in teacher evaluation and development. Both sets of opinions agreed with the work done in this study. In addition, teachers were shown to have made some modest improvement in their classroom teaching as result of using student evaluation feedback. This data, augmented with a consultation conference designed to devise strategies for improving certain aspects of the student-teacher relationship and the work done in the classroom, was a catalyst for the teacher to adopt a reflective attitude towards their classroom teaching. The modest gains made, however, need to be tempered with some caution as a result of the methodological difficulties outlined.

Bibliography

Abbott R D, Wulff D H, Nyquist J D, Ropp V A and Hess C W (1990) Satisfaction With Processes of Collecting Student Opinions About Instruction: The Student Perspective. Journal of Educational Psychology, 82, 2, 201-206

Abrami P C, Cohen P A and d'Apollonia S (1988) Implementation Problems in Meta-Analysis. Review of Educational Research 58, 2, 151-179

Abrami P C and d'Apollonia S (1991) Multidimensional Students' Evaluations of Teaching Effectiveness - Generalizability of "N=1" Research: Comment on Marsh (1991). Journal of Educational Psychology, 83, 3, 411-415

Abrami P C, d'Apollonia S and Cohen P A (1990) Validity of Student Ratings of Instruction: What We Know and What We Do Not. Journal of Educational Psychology 82, 2, 219-231

Acheson K A and Gall M D (1992) Techniques in the Clinical Supervision of Teachers: Preservice and Inservice Applications 3rd ed New York: Longman

Aitken G (1994) Discussing the Observation (Mimeo notes) Auckland: Auckland College of Education

Aleamoni L M (1974) The Usefulness of Student Evaluations in Improving College Teaching (Research Report No 356) Urbana, Illinois: University Of Illinois. (ERIC Document Reproduction Service No ED 088 317)

Aleamoni L M and Stevens J J (1983) The Effectiveness of Consultation in Support of Student Evaluation Feedback: A Ten Year Follow-up Paper presented at the Annual Meeting of the Rocky Mountain Psychological Association, Albuquerque NM, Apr 28-May 1

Amidon E J and Flanders N A (1971) The Role of the Teacher in the Classroom: A Manual for Understanding and Improving Teacher Classroom Behaviour (rev ed.) St Paul MN: Association for Productive Teaching

Amos L W, Purkey W W and Smith C H (1987) Using the Invitational Teaching Survey (ITS) and the Student Attitudinal Outcome Measures (SAOM) to Improve Teaching. Paper presented at the Annual Meeting of the American Educational Research Association (Washington DC, April 1987) ERIC Document Reproduction Service No. ED 281 891

Anderson G (1990) Fundamentals of Educational Research London: Falmer Press

Anderson L W and Burns R B (1989) Research in Classrooms: The Study of Teachers, Teaching and Instruction Oxford: Pergamon

Arreola R A (1984) Evaluation of Faculty Performance: Key Issues. In Seldin P Changing Practices in Faculty Evaluation Jossey Bass Higher Education Series

Ashcroft E (1984) Effective Instruction. In Nicholson K, Ashcroft E, Shouksmith G and Renwick M What Makes Schools Effective for Pupil Learning [Lectures supported by Newspapers in Education, The Dominion and The Evening Post]

Aspinwall K, Simkins T, Wilkinson J F and McAuley J (1992) Managing Evaluation in Education: A Developmental Approach London: Routledge

Aubrecht J D (1984) Better Faculty Evaluation Systems. In Seldin P Changing Practices in Faculty Evaluation Jossey Bass Higher Education Series

Aubrecht J D, Hanna G S & Hoyt D P (1986) A comparison of high school student ratings of teaching effectiveness with teacher self-ratings: Factor analytic and multitrait-multimethod analyses. Educational and Psychological Measurement 46, 1, 223-231

Ballard R and Duncan P (1989) Role of the Principal and Trustees in Tomorrow's Schools (Implementation Unit, Department of Education) Paper presented at the Secondary Principals' Association of New Zealand (SPANZ) Annual Conference, Palmerston North, 21 April

Barber L (1984) Teacher Evaluation and Merit Pay (Working Paper No TF-83-5) Denver CO, Education Commission of the States [Background Papers for the Task Force on Education for Economic Growth]

Barber L W (1990) Self-Assessment. In Millman J and Darling-Hammond L (eds) The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Barnes W (1995) [Interview, Auckland University, 20 October.]

Batten M (1989) Teacher and Pupil Perspectives on the Positive Aspects of Classroom Experience. Scottish Educational Review, 21, 1, 48-57

Batten M (1994) Students' Perceptions of Effective Teaching. Score 2, 6, 4-5

Batten M, Marland P and Khamis M (1993) Knowing How to Teach Well: Teachers Reflect on their Classroom Practice ACER Research Monograph No 44. Hawthorn, Victoria

Beecher D E (1949) The Evaluation of Teaching: Backgrounds and Concepts Syracuse: Syracuse University Press

Bennett H, Lister M and McManus M (1992) Teacher Appraisal: Survival and Beyond Harlow, Essex: Longman

Berk R A (1979) The Construction of Rating Instruments for Faculty Evaluation. Journal of Higher Education 50, 5, 650-669

Berliner D C (1980) Using Research on Teaching for the Improvement of Classroom Practice. Theory Into Practice 19, 4, 302-308

Berliner D C (1987) Knowledge is Power: A Talk to Teachers About a Revolution in the Teaching Profession. In Berliner D C and Rosenshine B V (eds) Talks to Teachers: A Festschrift for N L Gage New York: Random House, Chapter 1, pp3-33.

Berliner D C and Rosenshine B V (eds) (1987) Talks to Teachers: A Festschrift for N L Gage New York: Random House

Blackburn R T and Clark M J (1975) An Assessment of Faculty Performance: Some Correlates Between Administrator, Colleague, Student and Self-Ratings. Sociology of Education 48, 242-256

Boud D and McDonald R (1981) Educational Development Through Consultancy Research in Higher Education Monographs. Surrey: Society for Research into Higher Education

Braskamp L A (1984) Guidelines for Using Rating Scales and Checklists Phi Delta Kappa: Centre on Evaluation, Development and Research

Braskamp L A, Brandenburg D C and Ory J C (1984) Evaluating Teaching Effectiveness: A Practical Guide Beverly Hills, CA: Sage Publications

Braskamp L A, Caulley D and Costin F (1979) Student Ratings and Instructor Self-ratings and Their Relationship to Student Achievement. American Educational Research Journal 16, 3, 295-306

Braunstein D N, Klein G A & Pachla M (1973) Feedback expectancy shifts in student ratings of college faculty. Journal of Applied Psychology 58, 254-258

Bridges E M (1986) The incompetent teacher: The challenge and the response Philadelphia: Taylor & Francis

Bridges E M and Groves B (1984) Managing the Incompetent Teacher Eugene, Oregon: ERIC Clearinghouse on Educational Management

Brinko K T (1990) Instructional consultation with feedback in higher education. Journal of Higher Education 61, 1, 65-83

Brinko K T (1991) The Interactions of Teaching Improvement. In Theall M and Franklin J (eds) New Directions for Teaching and Learning: No 48 Effective Practices for Improving Teaching San Francisco, CA: Jossey-Bass

Brophy J (1987) On Motivating Students. In Berliner D C and Rosenshine B V (eds) Talks to Teachers: A Festschrift for N L Gage New York: Random House. Chapter 9, pp201-245.

Bush, R N (1954) The Teacher-Pupil Relationship New York: Prentice-Hall

Callahan R E (1962) Education and the Cult of Efficiency Chicago: The University of Chicago Press

Campbell D T and Stanley J C (1963) Experimental and Quasi-experimental Designs for Research on Teaching. In Gage N L Handbook of Research on Teaching Chicago: Rand-McNally.

Carroll M (1994) [Interview, 12 May, Massey University, Palmerston North]

Carter K R (1974) The effect of student feedback in modifying teaching performance. Dissertation Abstracts International 35,5110A

Cashin W E (1988) Student Ratings of Teaching: A Summary of the Research. Center for Faculty Evaluation and Development, IDEA Paper No 20: Kansas State University (ERIC Document Reproduction Service No ED 302 567)

Centra J A (1973) The effectiveness of student feedback in modifying college instruction. Journal of Educational Psychology 65, 395-401

Centra J (1975) Colleagues as Raters of Classroom Instruction. Journal of Higher Education 46, 1, 327-337

Centra J (1977) Student ratings of instruction and their relationship to student learning. American Educational Research Journal 14, 17-24

Charters W W & Gage N L (eds) (1963) Readings in social psychology of education Boston: Allyn and Bacon

Clark D L (1987) High School Seniors React to Their Teachers and Their Schools. Phi Delta Kappan March, 503-509

Coffman S J (1991) Improving Your Teaching Through Small-Group Diagnosis. College Teaching 39, 2, 80-82

Cohen J (1977) Statistical Power Analysis for the Behavioral Sciences (Rev ed) New York: Academic Press

Cohen P A (1981) Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies. Review of Educational Research 51, 3, 281-309

Cohen P A (1982) Validity of Student Ratings in Psychology Courses: A Research Synthesis. Teaching of Psychology 9, 2, 78-82

Cohen P A (1983) Comment on A Selective Review of the Validity of Student Ratings of Teaching. Journal of Higher Education 54, 4, 448-457

Cohen P A & Herr G (1979) A procedure for diagnostic instructional feedback: The Formative Assessment of College Teaching (FACT) model. Educational Technology 19, 18-23

Cohen P A & Herr G (1982) Using an Interactive Feedback Procedure to Improve College Teaching. Teaching of Psychology 9, 3, 138-140

Costin F, Greenough W T and Menges R J (1971) Student Ratings of College Teaching: Reliability, Validity, and Usefulness. Review of Educational Research 41, 5, 511-535

Cranton P and Smith R A (1990) Reconsidering the Unit of Analysis: A Model of Student Ratings of Instruction. Journal of Educational Psychology, 82, 2, 207-212

Cross K P (1987) The Adventures of Education in Wonderland: Implementing Education Reform. Phi Delta Kappan, March, 496-502

Cross K P (1988) Feedback in the Classroom: Making Assessment Matter. Paper collected as part of the American Association for Higher Education Assessment Forum. ERIC Document Reproduction Service No. ED 299 922

Cruickshank D R and Haefele D L (1990) Research-Based Indicators: Is the Glass Half-Full or Half-Empty? In Schwab R L (ed.) Research-Based Teacher Evaluation. Boston MA: Kluwer Academic Publishers, pp33-39

Cuban L (1983) Effective Schools: A Friendly but Cautionary Note. Phi Delta Kappan, June, 695-696

Darling-Hammond L, Wise A E and Pease S R (1983) Teacher Evaluation in the Organizational Context: A Review of the Literature. Review of Educational Research 53, 3, 285-328

Delamont S (1992) Fieldwork in Educational Settings: Methods, Pitfalls & Perspectives London: Falmer Press

Department of Education (1990a) A Guide to Governance and Management A Report of the Principals' Implementation Task Force. Wellington: Governement Printer

Department of Education (1990b) A Guide to Personnel Management A Report of the Principals' Implementation Task Force. Wellington: Governement Printer

Donald J G and Sullivan A M (eds) (1985) Using Research to Improve Teaching. New Directions for Teaching and Learning, No 23, September

Dowell D A and Neal J A (1982) A Selective Review of the Validity of Student Ratings of Teaching. Journal of Higher Education 53, 1, 51-62

Doyle K O (1983) Evaluating Teaching Lexington MA: Lexington Books

Doyle K O and Crichton L I (1978) Student, Peer and Self Evaluations of College Instructors. Journal of Educational Psychology 70, 815-826

Doyle K O & Whitely S E (1974) Student Ratings as Criteria for Effective Teaching. American Educational Research Journal 11, 3, 259-274

Drews D R, Burroughs W J and Nokovich D (1987) Teacher Self-Ratings as a Validity Criterion for Student Evaluations. Teaching of Psychology 14, 1, 23-25

Duke D L and Stiggins R J (1990) Beyond Minimum Competence: Evaluation for Professional Development. In Millman J and Darling-Hammond L (eds) The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Eble K E (1984) New Directions in Faculty Evaluation. In Seldin P Changing Practices in Faculty Evaluation San Francisco: Jossey Bass

Edwards W L (1992a) Appraising Teachers: Is There Light Without Heat? set 92, 1 Wellington: New Zealand Council for Educational Research

Edwards W L (1992b) The Appraisal Process: Cycling Towards Professional Development. [Paper presented to the Principals' Training - Curriculum Leadership Project, Dunedin College of Education, New Zealand, 2 October 1992]

Erickson G R & Erickson B L (1979) Improving college teaching: An evaluation of a teaching consultation procedure. Journal of Higher Education 50, 670-683

Festinger L (1957) A Theory of Cognitive Dissonance. Evanston, Illinois: Row, Paterson and Co

Finlay I (1992) Observation and Feedback: Using Appraisal to Enhance Professionalism Paper presented at the Annual Conference of the British Educational Management and Administration Society, Bristol

Fox R, Peck R F, Blattstein A, & Blattstein D (1983) Student evaluation of teacher as measure of teacher behaviour and teacher impact on students. Journal of Educational Research 77, 1, 16-21

Fraenkel J R and Wallen N E (1993) How to Design and Evaluate Research in Education (2nd ed) New York: McGraw-Hill Inc.

Fraser B J (1986) Classroom Environment London: Croom Helm

Fraser G S (1994) Ready Set Go. Extramural News, September, Palmerston North: Massey University

Freiberg H J and Waxman (1988) Alternative Feedback Approaches for Improving Student Teacher's Classroom Instruction. Journal of Teacher Education 39, 4, 8-14

Frey P W (1973) Student Ratings of Teaching: Validity of Several Rating Factors. Science 182, 83-85

Frey P W, Leonard D W & Beatty W M (1975) Student Ratings of Instruction. American Educational Research Journal 12, 4, 435-447

Friedlander J (1978) Student perceptions on the effectiveness of midterm feedback to modify college instruction. Journal of Educational Research 71, 140-143

Froman R D & Owen S V (1980) Influence of Different Types of Student Ratings Feedback upon Later Instructional Behavior. (Paper presented at the Annual Meeting of the American Educational Research Association, Boston) (ERIC Document Reproduction Service No ED 187 724)

Fry E B (1972) Reading Instruction for Classroom and Clinic New York: McGraw Hill

Fullan M G and Hargreaves A (1991) Working Together for Your School: Strategies for Developing Interactive Professionalism in Your School Hawthorn, Victoria: Australian Council for Educational Administration

Furtwengler C B (1987) Use of Student Achievement Information in the Performance Evaluation of Teachers in the Tennessee Career Ladder Program. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Mobile AL, November.

Gage N L (1963) Handbook of Research on Teaching Chicago: Rand-McNally.

Gage N L, Runkel P J and Chatterjee B B (1963) Changing Teacher Behaviour Through Feedback from Pupils: An Application of Equilibrium Theory. In Charters W W and Gage N L (eds) Reading in the Social Psychology of Education. Boston: Allyn and Bacon

Geva-May I (1993) Do Student Questionnaire Responses Reflect Actual Active Behavior? A Study in Instruction Evaluation. Studies in Educational Evaluation 19, 383-396

Gibbs G, Habeshaw S and Habeshaw T (1989) 53 Interesting Ways to Appraise Your Teaching Bristol: Technical and Educational Services Ltd

Gies G L (1984) The Context of Evaluation. In Seldin P Changing Practices in Faculty Evaluation Jossey Bass Higher Education Series

Gitlin A and Smyth J (1989) Teacher Evaluation: Educative Alternatives New York: Falmer Press

Glass G V (1978) Integrating Findings: The Meta-analysis of Research. In Shulman L S (ed) Review of Research in Education Vol 5 Itasca, IL:Peacock

Glass G V (1990) Using Student Test Scores to Evaluate Teachers. In Millman J and Darling-Hammond L (eds) The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Glasser W (1990) The Quality School. Phi Delta Kappan, February, 425-435

Gold Y and Roth R A (1993) Teachers Managing Stress and Preventing Burnout London: The Falmer Press

Goldsmid C A, Gruber J E and Wilson E K (1977) Perceived Attributes of Superior Teachers (PAST): An Inquiry into the Giving of Teacher Awards. American Educational Research Journal 14, 4, 423-440

Good T L (1984) Observation and the Evaluation of Teaching Phi Delta Kappan: Centre on Evaluation, Development and Research

Good T L (1987) Teacher Expectations. In Berliner D C and Rosenshine B V (eds) Talks to Teachers: A Festschrift for N L Gage New York: Random House, Chapter 8, pp159-200.

Good T L and Brophy J E (1991) Looking in Classrooms (5th ed.) New York: Harper Collins

Good T L and Mulryan C (1990) Teacher Ratings: A Call for Teacher Control and Self-Evaluation. In Millman J and Darling-Hammond L (eds) The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Grace G (1985) Judging Teachers: The Social and Political Contexts of Teacher Evaluation. British Journal of Sociology of Education 6, 1, 3-16

Greenwood G E, Bridges C M, Ware W B and McLean J E (1973)   Student Evaluation of College Teaching Behaviors Instrument: A Factor Analysis. Journal of Higher Education 44, 596-604

Grush J E and Costin F (1975) The Student as Consumer of the Teaching Process. American Educational Research Journal 12, 1, 55-66

Haertel E H (1991) Teacher Assessment. Review of Research in Education, 17, 3-29

Hair J F, Anderson R E, Tatham R L and Black W C (1995) Multivariate Data Analysis: With Readings (4th ed) Englewood Cliffs, New Jersey: Prentice-Hall

Hall C and Fitzgerald C (1993) Student Summative Evaluation of Teaching: (Draft) Code of Practice Wellington: Joint publication between Association of University Staff of New Zealand (Inc) and University Teaching Development Centre, Victoria University of Wellington

Haller E J and Strike K A (eds) (1986) An Introduction to Educational Administration: Legal and Ethical Perspectives. New York: Longman

Hart P, Taylor M and Robottom I (1994) Dilemmas of Participatory Enquiry: A Case Study of Method-in-Action. Assessment and Evaluation in Higher Education 19, 3, 201-214

Hepworth D and Oviatt B E (1985) Using Student Course Evaluations: Findings, Issues, and Recommendations. Journal of Social Work Education 3, 105-112

Hickcox E S, Lawton S B, Leithwood K A and Musella D F (1988) Making a Difference through Performance Appraisal [Research in Education Series, No 16] Toronto: Ontario Institute for the Study of Education

Hildebrand M (1972) How to recommend promotion of a mediocre teacher without actually lying. Journal of Higher Education 43, 44-62

Hobar N and Sullivan D K (1984) Systematic Observation of Instruction: Genesis, Research, Practice and Potential. Journal of Classroom Instruction 19, 2, 26-34

Hosford P L (ed.) (1984) Using What We Know About Teaching Association for Supervision and Curriculum Development Year-book: Alexandria VA.

Hoy W K and Tarter C J (1993) A Normative Theory of Participative Decision Making in Schools. Journal of Educational Administration, 31, 3, 4-19

Isaacson R L, McKeachie W J, Milholland J E, Lin Y G, Hofeller M, Baerwaldt J W and Zinn K L (1964) Dimensions of Student Evaluations of Teaching. Journal of Educational Psychology 55, 344-351

Iwanicki E F (1990) Teacher Evaluation for School Improvement. In Millman J and Darling-Hammond L (eds) The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Jones J (1989) Students' Ratings of Teacher Personality and Teaching Competence. Higher Education 18, 551-558

Kauchak D, Peterson K and Driscoll A (1984) An Interview Study of Teachers' Attitudes Toward Teacher Evaluation Practices Paper presented at the Annual Meeting of the American Education Research Association, New Orleans (ERIC Document Reproduction Service No ED 246 106)

Ker P (1994) Teacher Appraisal and Teacher Development: The Experience of a Polytechnic Teaching Department Unpublished master's thesis, Massey University, Palmerston North

Ker P (1995) [Personal interview, Auckland, 7 July 1995]

Knapp M S (1984) Teacher Evaluation Practices Within Schools: An Overview of the Literature and Issues from the Practitioner's Point of View. The CEDR Quarterly

Koffman M L (1974) A study of the effects of instructional analysis and feedback on the classroom behavior and student achievement of university teaching assistants. Dissertation Abstracts International 35, 302A

Kohlan R G (1973) A Comparison of Faculty Faculty Evaluations Early and Late in the Course. Journal of Higher Education 44, 587-595

Kulik J A and Kulik C-L C (1974) Student Ratings of Instruction. Teaching of Psychology 1, 2, 51-57

Kulik J A & McKeachie W J (1975) The Evaluation of Teachers in Higher Education. In Kerlinger XX (ed) Review of Research in Education 3, 210-240. Itasca, IL: Peacock

Lange D (1988) Tomorrow's Schools: The Reform of Education Administration in New Zealand Wellington: Government Printer

Larsgaard J O (1971) An experiment using a student rating scale as a form of feedback for the modification of teaching behavior at the college level. Dissertation Abstracts International 32, 2485A

Leventhal L, Abrami P C, & Perry R P (1976) Do Teacher Rating Forms Reveal as Much About Students as About Teachers? Journal of Educational Psychology 68, 441-445

Levy J, Wubbels T and Brekelmans M (1992) Student and Teacher Characteristics and Perceptions of Teacher Communication Style. Journal of Classroom Interaction 23, 1, 23-29

Levinson-Rose J and Menges R J (1981) Improving college teaching: A critical review of research. Review of Educational Research 51, 403-434

L'Hommedieu R, Menges R J and Brinko K T (1990) Methodological Explanations for the Modest Effects of Feedback From Student Ratings. Journal of Educational Psychology 82, 2, 232-241

Lincoln University (n d) Student Rating: Myths Vs Research Facts Lincoln University: Education Centre

Lokan J and McKenzie P (eds) (1989) Teacher Appraisal: Issues and Approaches [Australian Education Review No. 28] Victoria, Australia: Australian Council for Educational Research

MacDonald B (1976) Evaluation and the control of education. In Tawney D A (ed.) Curriculum Evaluation Today: Trends and Implications MacMillan, London

Marsh H W (1977) The Validity of Students' Evaluations: Classroom Evaluations of Instructors Independently Nominated As Best and Worst Teachers by Graduating Seniors. American Educational Research Journal 14, 4, 441-447

Marsh H W (1984) Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases, Utility. Journal of Educational Psychology 76, 5, 707-754

Marsh H W (1987) Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research. International Journal of Educational Research 11, 253-388

Marsh H W (1991a) Multidimensional Students' Evaluations of Teaching Effectiveness: A Test of Alternative Higher-Order Structures. Journal of Educational Psychology, 83, 2, 285-296

Marsh H W (1991b) A Multidimensional Perspective on Students' Evaluations of Teaching Effectiveness: Reply to Abrami and d'Apollonia (1991). Journal of Educational Psychology, 83, 3, 416-421

Marsh H W (1994) Student evaluation of teaching. In Husen T and Postlethwaite T N (eds) International Encyclopaedia of Education: Research and Studies (2nd ed) Oxford: Pergamon Vol 10, pp 6227-6235.

Marsh H W and Hocevar D (1991a) The Multidimensionality of Students' Evaluations of Teaching Effectiveness: The Generality of Factor Structures Across Academic Discipline, Instructor Level, and Course Level. Teaching and Teacher Education 7, 1, 9-18

Marsh H W and Hocevar D (1991b) Students' Evaluations of Teaching Effectiveness: The Stability of Mean Ratings of the Same Teachers Over a 13-Year Period. Teaching and Teacher Education 7, 4, 303-314

Marsh H W & Overall J U (1980) Validity of student evaluations of teaching effectiveness: Cognitive and affective criteria. Journal of Educational Psychology 72, 468-475

Marsh H W and Roche L A (1993) The Use of Students' Evaluations and an Individually Structured Intervention To Enhance University Teaching Effectiveness American Educational Research Journal 30, 217-251. [ERIC Document Reproduction Service No. ED 353 283]

Marsh H W and Roche L A (1994) The Use of Students' Evaluations of University Teaching to Improve Teaching Effectiveness. (Final Project Report for Department of Employment, Education and Training) Canberra: Australian Government Publishing Service

Massey University (1993) New System to SET Direction for Evaluating Teaching Performance. Massey Alumnus Christmas issue, 1-2

McCall W A and Krause G R (1959) Measurement of Teacher Merit. Journal of Educational Research 53, 2, 73-75

McGreal T L (1990) Rating Scales in Teacher Evaluation. In Schwab R L (ed.) Research-Based Teacher Evaluation. Boston MA: Kluwer Academic Publishers, pp41-58.

McKeachie W J, Lin Y-G and Mann W (1971) Student Ratings of Teacher Effectiveness: Validity Studies. American Educational Research Journal 8, 3, 435-445

McKeachie W J, Lin Y-G, Daugherty M, Moffett M M, Neigler C, Nork J, Walz M, & Baldwin R (1980) Using student ratings and consultation to improve instruction. British Journal of Educational Psychology 50, 168-174

McLaughlin M W (1990) Embracing Contraries: Implementing and Sustaining Teacher Evaluation. In Millman J and Darling Hammond L (eds) The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Medley-Mark V and Weston C B (1988) A Comparison of Student Feedback Obtained From Three Methods of Formative Evaluation of Instructional Materials. Instructional Science 17, 3-27

Mehrens W A (1990) Combining Evaluation Data from Multiple Sources. In Millman J and Darling-Hammond L (eds) The New Handbook of Teacher Evaluation:

Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Menges R J (1977) The Intentional Teacher: Controller, Manager, Helper Monterey CA: Brooks/Cole Publishing

Menges R J (1984) Evaluation in the Service of Faculty. In Seldin P Changing Practices in Faculty Evaluation Jossey Bass Higher Education Series

Menges R J (1988) Why Research May Underestimate Effects of Feedback from Student Ratings Paper presented at the Annual Meeting of the Association for the Study of Higher Education, St Louis MO, Nov 3-8

Menges R J and Brinko K T (1986) Effects of Student Evaluation Feedback: A Meta-Analysis of Higher Education Research Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco CA, April 16-20.

Meredith G M (1976) Structure of the Teaching Analysis by Students (TABS). Journal of Psychology 92, 235-242

Miles P L (1989) A Communication Based Strategy to Improve Teaching: The Continuous Feedback Technique Paper presented to the Annual Meeting of the Speech Communication Association, San Francisco CA

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

Millman J and Darling-Hammond L (eds) (1990) The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Ministry of Education (1995) Draft National Guidelines for Performance Management in Schools Wellington: Learning Media

Moses I (1989) Student Input into Teacher Appraisal. In Lokan J and McKenzie P (eds) Teacher Appraisal: Issues and Approaches [Australian Education Review No. 28] Victoria, Australia: Australian Council for Educational Research

Mousley J A, Sullivan P and Gervasoni A (1994) The Analysis of Teaching: Constraints on Lesson Description and Critique. Proceedings of the 17th Annual Conference of the Mathematics Education Research Group of Australasia (MERGA), July, Southern Cross University: Lismore, NSW

Munn P, Johnstone M and Holligan C (1990) Pupils' Perceptions of "Effective Disciplinarians". British Educational Research Journal 16, 2, 191-198

Nicholson K, Ashcroft E, Shouksmith G and Renwick M (1984) What Makes Schools Effective for Pupil Learning [Lectures supported by Newspapers in Education, The Dominion and The Evening Post]

Ory J C, Braskamp L A & Pieper D M (1980) Congruence of Student Evaluative Information Collected by Three Methods. Journal of Educational Psychology 72, 2, 181-185

Overall J U and Marsh H W (1979) Midterm Feedback from Students: Its Relationship to Instructional Improvement and Students' Cognitive and Affective Outcomes. Journal of Educational Psychology 71, 6, 856-865

Overall J U and Marsh H W (1980) Student's Evaluation of Instruction : A Longitudinal Study of Their Stability. Journal of Educational Psychology 72, 3, 321-325

Page C F (1974) Student Evaluation of Teaching: The American Experience. Research in Higher Education Monographs

Pajak E and Glickman C D (1989) Informational and Controlling Language in Simulated Supervisory Conferences. American Educational Research Journal 26, 1, 93-106

Pambookian H S (1972) The effect of feedback from students to college instructors on their teaching behavior. Dissertation Abstracts International 33, 4950A

Pambookian H S (1974) The initial level of student evaluation of instruction as a source of influence on instructor change after feedback. Journal of Educational Psychology 66, 52-56

Parsons R (1995) The Loneliness of the Dedicated Teacher. Education Today 7,3, 8

Perry R P (1985) Instructor Expressiveness: Implications for Improved Teaching. In Donald J G and Sullivan A M (eds) Using Research to Improve Teaching New Directions for Teaching and Learning, No 23, September

Petersen K D (1990) Assistance and Assessment for Beginning Teachers. In Millman J and Darling-Hammond L (eds) The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Petersen D, Kromrey J and Smith D C (1990) Research Based Teacher Evaluation: A response to Scriven. In Schwab R L (ed.) Research-Based Teacher Evaluation. Boston MA: Kluwer Academic Publishers, pp7-17

Phi Delta Kappa Evaluation of Teaching: The Formative Process Centre on Evaluation Development and Research, Hot Topic Series, 1983-84

Polahar K L (1981) The effect of student feedback to college business instructors on the evaluation of instruction. Dissertation Abstracts International 43, 2208A

Pollitt C (1983) Models of Staff Appraisal: Some Political Implications. Studies in Higher Education, 13, 1, 7-15

Porter A C and Brophy J (1988) Synthesis of Research on Good Teaching: Insights from the Work of the Institute for Research on Teaching. Educational Leadership, May, 74-85

Poster C and Poster D (1992) Teacher Appraisal: A Guide to Training. London: Routledge

Ramsden P (1990) Evaluating Teaching: Supporting Learning Paper presented at the Annual Conference of the Higher Education Research & Development Society of Australasia, Griffiths University, July

Ramsey R S (1980) The effect of evaluative feedback from students on faculty performance at the University of Mississippi. Dissertation Abstracts International 41, 4622A

Rayder N F (1968) College Student Ratings of Instructors. The Journal of Experimental Education 37, 2, 76-81

Remmers H H and Gage N L (1955) Educational Measurement and Evaluation (rev ed.) New York: Harper and Brothers

Rodin M and Rodin B (1972) Student Evaluations of Teachers. Science 177, 1164-1166

Rosenshine B (1971) Teaching Behaviours and Student Achievement Slough: NFER

Rotem A & Glasman N S (1979) On the effectiveness of students' evaluative feedback to university instructors. Review of Educational Research 49, 497-511

Rutter M, Maughan B, Mortimore P, Ouston J with Smith A (1979) Fifteen Thousand Hours: Secondary Schools and Their Effects on Children London: Open Books

Savage T V and McCord M K (1986) The Use of Student Evaluation in the Assessment of Teacher Competence Paper presented at the Annual Meting of the American Educational Research Association, San Francisco CA, April

Schuh A J and Crivelli M A (1973) Animadversion Error in Student Evaluations of Faculty Teaching Effectiveness. Journal of Educational Psychology 58, 2, 259-260

Schwab R L (ed.) (1990) Research-Based Teacher Evaluation. Boston MA: Kluwer Academic Publishers

Scott N H (1976) A pilot study testing the effects of feedback on teaching behavior as measured by student ratings at New River Community College, Dublin, Virginia. Dissertation Abstracts International 37, 1950A

Scriven M (1988) The Validity of Student Ratings. Instructional Evaluation 9, 2, 5-19

Scriven M (1990) Can Research Based Teacher Evaluation be Saved? In Schwab R L (ed.) Research-Based Teacher Evaluation. Boston MA: Kluwer Academic Publishers, pp19-32

Seldin P (1984) Changing Practices in Faculty Evaluation San Francisco: Jossey Bass

Sergiovanni T J (ed.) (1982) Supervision of Teaching Association for Supervision and Curriculum Development Year-book: Alexandria VA.

Sergiovanni T J (1982) Toward a Theory of Supervisory Practice: Integrating Scientific, Clinical and Artistic Views. In Sergiovanni T J (ed.) Supervision of Teaching Association for Supervision and Curriculum Development Year-book: Alexandria VA., 67-78

Siagan E (1983) The impact of student evaluative feedback on the perception of the teaching faculty at Andrews University. Dissertation Abstracts International 44, 1254A

Smith D L (1977) The relationship of feedback to professors of the results of student ratings of their teaching effectiveness at mid-semester to their end-of-semester ratings. Dissertation Abstracts International 38, 187A

Smith W G (1975) The Influence of Student Feedback on Changing Teaching Behaviors of College Faculty Members. Dissertation Abstracts International 36, 4424A

Smyth W J (1988) A Critical Pedagogy of Teacher Evaluation Geelong: Deakin University

Snedeker J H and Remmers H H (nd) The Purdue Instructor Performance Indicator - Form A Purdue Research Foundation, Purdue University

Soar R S, Medley D M, and Coker H (1983) Teacher Evaluation: A Critique of Currently Used Methods. Phi Delta Kappan, December, 239-246

SPSS Inc (1994) SPSS for Windows, Version 6.11 Chicago

Stanley S J and Popham W J (eds) (1988) Teacher Evaluation: Six Prescriptions for Success Alexandria, VA: Association for Supervision and Curriculum

Stevens J J and Aleamoni L M (1985) The Use of Evaluative Feedback for Instructional Improvement: A Longitudinal Perspective. Instructional Science 13, 285-304

Stewart D and Prebble T (1985) Making It Happen: A School Development Process Palmerston North: The Dunmore Press

Stewart D and Prebble T (1993) The Reflective Principal: School Development Within a Learning Community Palmerston North: ERDC Press

Stiggins R J and Duke D L (1988) The Case for Commitment to Teacher Growth: Research on Teacher Evaluation New York: State University of New York Press

Strike K A (1990) The Ethics of Educational Evaluation. In Millman J and Darling-Hammond L (eds) The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Stufflebeam D L, Foley W J, Gephart W J, Guba E G, Hammond R L, Merriman H O & Provus M M (1971) Educational Evaluation and Decision Making Phi Delta Kappa National Study: Committee on Evaluation. Itasca, Illinois: F E Peacock

Sullivan A M (1985) The Role of Two Types of Research on the Evaluation and Improvement of University Teaching. In Donald J G & Sullivan A M (eds) Using Research to Improve Teaching. New Directions for Teaching and Learning, No 23. San Francisco: Jossey-Bass

Sullivan A M & Skanes G R (1974) Validity of student evaluation of teaching and the characteristics of successful instructors. Journal of Educational Psychology 66, 584-590

Thomas H B Jr (1969) Changing Teacher Behavior: An Experiment in Feedback from Students to Teachers. Dissertation Abstracts International 30, 1903A

Tiberius R G, Sackin H D and Cappe L (1987) A Comparison of Two Teaching Methods for Evaluating Teaching. Studies in Higher Education 12, 3, 287-297

Tollefson N and Wigington H (1986) Teacher-generated and Student-generated Variability in Teacher Effectiveness Ratings. Instructional Science 15, 109-120

Trank D M (1978) The Effect of Student Feedback on the Improvement of Instruction (Report CS 502 659) Iowa City, Iowa: University of Iowa. (ERIC Document Reproduction Service No ED 177 642)

Toney J W (1973) The effects of feedback to teachers from student evaluations of the instructional process. Dissertation Abstracts International 34, 7652A

Tuckman B W (1976) The Tuckman Teacher Feedback Form (TTFF). Journal of Educational Measurement 13, 3, 233-7

Tuckman B W & Yates D (1980) Evaluating the student feedback strategy of changing teacher style. Journal of Educational Research 74, 2, 74-78

Tuckman B W and Oliver W F (1968) Effectiveness of Feedback to Teachers as a Function of Source. Journal of Educational Psychology XX, 4, 297-301

Vogt K E & Lasher H (1973) Does Student Evaluation Stimulate Improved Teaching? Ohio: Bowling Green State University, College of Business Administration. ERIC Document Reproduction Service No ED 078 748

Watkins D, Marsh H W and Young D (1987) Evaluating Tertiary Teaching: A New Zealand Perspective. Teaching and Teacher Education 3, 1, 41-53

Webb G (1994) Making the Most of Appraisal: Career and Professional Development Planning for Lecturers London: Kogan Page

Webb W B and Nolan C Y (1955) Student, Supervisor, and Self-ratings of Instructional Proficiency. Journal of Educational Psychology 46, 42-46

Weinbach R W (1988) Manipulations of Student Evaluations: No Laughing Matter, Journal of Social Work Education 1, 27-34

Whitely S E and Doyle K O (1976) Implicit Theories in Student Ratings. American Educational Research Journal 13, 4, 241-253

Wilson R C (1986) Improving faculty teaching: Effective use of student evaluations and consultants. Journal of Higher Education 57, 196-211

Winder J (1996) Peer Coaching. A paper presented to the Auckland Secondary Deputy and Assistant Principals Association, August 1996

Wise A E and Gendler T (1990) Governance Issues in the Evaluation of Elementary and Secondary Schoolteachers. In Millman J and Darling-Hammond L (eds) The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers Newbury Park, CA: Sage Publications

Wragg E C (ed.) (1984) Classroom Teaching Skills Beckenham: Croom Helm

Wragg E C and Wood E K (1984) Pupil Appraisals of Teaching. In Wragg E C (ed.) Classroom Teaching Skills Beckenham: Croom Helm

Zumwalt K K (ed.) (1986) Improving Teaching Association for Supervision and Curriculum Development Year-book: Alexandria VA.

A	B	C	D	E	F	G	H	I	J	K	K-I
		1	2	3	4	5	6	Mean	Std Dev	Teacher	Difference
1	Pride	2	13	8	0	0	0	2.26	0.60558	2.00	-0.26
2	Discuss.n	7	5	6	5	0	0	2.39	1.13211	2.00	-0.39
3	Handling *	2	4	6	7	1	0	2.95	1.07121	4.00	1.05
4	Clarity	2	9	5	5	2	0	2.83	1.12876	1.00	-1.83
5	Prompt	3	7	6	2	5	0	2.96	1.33444	3.00	0.04
6	Happy	7	10	4	2	0	0	2.04	0.90785	2.00	-0.04
7	Control *	0	1	1	12	8	0	1.77	0.73434	5.00	3.23
8	Challenge	2	6	9	5	1	0	2.87	0.99146	2.00	-0.87
9	Approachable	1	0	8	4	9	1	3.91	1.08331	3.00	-0.91
10	Related	0	3	4	11	4	1	3.73	0.91363	4.00	0.27
11	Enthuses	0	6	10	6	1	0	3.09	0.82951	2.00	-1.09
12	Humour	1	9	5	4	4	0	3.04	1.19703	2.00	-1.04
13	Impt Pts	5	11	5	1	0	0	2.09	0.79253	1.00	-1.09
14	Strict *	2	6	8	4	2	0	3.09	1.08331	2.00	-1.09
15	Disrupt. Std	2	7	6	6	2	0	2.96	1.12204	2.00	-0.96
16	Asmt Fair	4	13	2	3	0	0	2.18	0.88607	1.00	-1.18
17	Mat Prep	6	5	7	5	0	0	2.48	1.0982	2.00	-0.48
18	Know Wld	3	7	6	1	4	1	2.81	1.29538	3.00	0.19
19	Chall/Intrst	2	3	11	5	2	0	3.09	1.0178	2.00	-1.09
20	Improve	1	6	8	5	3	0	3.13	1.07559	2.00	-1.13
21	Asmt Crit	4	7	6	3	1	0	2.52	1.09627	2.00	-0.52
22	Hmwk Ovld *	3	2	9	9	0	0	2.96	0.99905	3.00	0.04
23	Try Hard	6	7	9	1	0	0	2.22	0.88251	2.00	-0.22
24	Boys Qns *	0	3	9	4	5	1	2.48	1.00565	5.00	2.52
25	Ignore Tch *	0	2	5	13	3	0	2.26	0.79221	3.00	0.74
26	Feel Gd	2	6	7	5	3	0	3.04	1.16014	2.00	-1.04
27	Org/Expct	4	6	5	7	1	0	2.78	1.17793	1.00	-1.78
28	Too Much	3	5	10	3	2	0	2.83	1.08956	4.00	1.17
29	Incr Think	4	10	5	2	1	0	2.36	1.02449	2.00	-0.36
30	Esy Undrst	1	8	2	7	5	0	3.30	1.26611	2.00	-1.30
31	Gen Conc	0	6	3	9	4	0	3.50	1.07661	2.00	-1.50
32	No Hm Hlp *	2	5	6	3	7	0	2.65	1.33868	3.00	0.35
33	Stdy Skls	0	5	7	8	3	0	3.39	0.96635	4.00	0.61
34	Team	5	7	6	3	1	0	2.45	1.11711	1.00	-1.45
35	Adpts Ablty	1	7	6	7	2	0	3.09	1.05966	1.00	-2.09
36	Hmk Cpes	5	8	4	5	1	0	2.52	1.17472	1.00	-1.52
37	St Disply	0	5	8	5	4	0	3.36	1.02449	3.00	-0.36
38	Inclusive	9	8	4	1	1	0	2.00	1.06322	1.00	-1.00
39	Get Job	10	4	5	2	2	0	2.22	1.31733	1.00	-1.22
40	S Ask Qns	2	6	9	1	5	0	3.04	1.23282	2.00	-1.04
41	Motivates	0	5	6	5	7	0	3.61	1.13211	2.00	-1.61
42	Difficulty *	0	3	5	7	8	0	2.13	1.03438	4.00	1.87
43	Cmnts Hlp	3	6	5	7	2	0	2.96	1.19703	2.00	-0.96
44	On Time	0	1	5	8	9	0	4.09	0.88037	4.00	-0.09
45	Speed *	1	9	4	8	1	0	3.04	1.04167	3.00	-0.04
46	Pronounce	9	8	2	3	1	0	2.09	1.17633	2.00	-0.09
47	Cont Nxt Yr	4	5	10	2	2	0	2.70	1.12036	2.00	-0.70
48	Examples	0	11	8	4	0	0	2.70	0.74803	2.00	-0.70
49	Fin Early *	1	2	4	9	6	1	2.23	1.08426	3.00	0.77
50	Talk A Lot *	1	5	5	7	5	0	2.57	1.17311	1.00	-1.57
51	Bst Tchr	0	3	12	3	5	0	3.43	0.97026	2.00	-1.43
52	Bst Class	2	7	7	3	3	1	2.91	1.1642	2.00	-0.91

	A	B	C	D	E	F	G	H	I	J	K	"K-I
Dimension			1	2	3	4	5	6	Average	Std Dev	Teacher	Difference
Attitude	1	Pride	2	13	8	0	0	0	2.26	0.61	2.00	-0.26
	23	Try Hard	6	7	9	1	0	0	2.22	0.88	2.00	-0.22
	26	Feel Gd	2	6	7	5	3	0	3.04	1.16	2.00	-1.04
	47	Cont Nxt Yr	4	5	10	2	2	0	2.70	1.12	2.00	-0.70
Utility	10	Related	0	3	4	11	4	1	3.73	0.91	4.00	0.27
	39	Get Job	10	4	5	2	2	0	2.22	1.32	1.00	-1.22
Workload	28	Too Much	3	5	10	3	2	0	2.83	1.09	4.00	1.17
	35	Adpts Ablty	1	7	6	7	2	0	3.09	1.06	1.00	-2.09
	42	Difficulty *	0	3	5	7	8	0	2.13	1.03	4.00	1.87
	45	Speed *	1	9	4	8	1	0	3.04	1.04	3.00	-0.04
	49	Fin Early *	1	2	4	9	6	1	2.23	1.08	3.00	0.77
Org/Clarity	4	Clarity	2	9	5	5	2	0	2.83	1.13	1.00	-1.83
	13	Impt Pts	5	11	5	1	0	0	2.09	0.79	1.00	-1.09
	17	Mat Prep	6	5	7	5	0	0	2.48	1.10	2.00	-0.48
	27	Org/Expct	4	6	5	7	1	0	2.78	1.18	1.00	-1.78
	30	Esy Undrst	1	8	2	7	5	0	3.30	1.27	2.00	-1.30
	44	On Time	0	1	5	8	9	0	4.09	0.88	4.00	-0.09
Other Studs	2	Discuss.n	7	5	6	5	0	0	2.39	1.13	2.00	-0.39
	6	Happy	7	10	4	2	0	0	2.04	0.91	2.00	-0.04
	34	Team	5	7	6	3	1	0	2.45	1.12	1.00	-1.45
	40	S Ask Qns	2	6	9	1	5	0	3.04	1.23	2.00	-1.04
Rapport	3	Handling *	2	4	6	7	1	0	2.95	1.07	4.00	1.05
	9	Approachable	1	0	8	4	9	1	3.91	1.08	3.00	-0.91
	24	Boys Qns *	0	3	9	4	5	1	2.48	1.01	5.00	2.52
	31	Gen Conc	0	6	3	9	4	0	3.50	1.08	2.00	-1.50
	37	St Disply	0	5	8	5	4	0	3.36	1.02	3.00	-0.36
	38	Inclusive	9	8	4	1	1	0	2.00	1.06	1.00	-1.00
	46	Pronounce	9	8	2	3	1	0	2.09	1.18	2.00	-0.09
Homework	22	Hmwk Ovld *	3	2	9	9	0	0	2.96	1.00	3.00	0.04
	32	No Hm Hlp *	2	5	6	3	7	0	2.65	1.34	3.00	0.35
	36	Hmk Cpes	5	8	4	5	1	0	2.52	1.17	1.00	-1.52
Enthusiasm	11	Enthuses	0	6	10	6	1	0	3.09	0.83	2.00	-1.09
	12	Humour	1	9	5	4	4	0	3.04	1.20	2.00	-1.04
	41	Motivates	0	5	6	5	7	0	3.61	1.13	2.00	-1.61
Learning/Val	8	Challenge	2	6	9	5	1	0	2.87	0.99	2.00	-0.87
	18	Know Wld	3	7	6	1	4	1	2.81	1.30	3.00	0.19
	19	Chall/Intrst	2	3	11	5	2	0	3.09	1.02	2.00	-1.09
	29	Incr Think	4	10	5	2	1	0	2.36	1.02	2.00	-0.36
	33	Stdy Skls	0	5	7	8	3	0	3.39	0.97	4.00	0.61
	48	Examples	0	11	8	4	0	0	2.70	0.75	2.00	-0.70
	50	Talk A Lot *	1	5	5	7	5	0	2.57	1.17	1.00	-1.57
Control	7	Control *	0	1	1	12	8	0	1.77	0.73	5.00	3.23
	14	Strict *	2	6	8	4	2	0	3.09	1.08	2.00	-1.09
	15	Disrupt. Std	2	7	6	6	2	0	2.96	1.12	2.00	-0.96
	25	Ignore Tch *	0	2	5	13	3	0	2.26	0.79	3.00	0.74
Assessm.nt	5	Prompt	3	7	6	2	5	0	2.96	1.33	3.00	0.04
	16	Asmt Fair	4	13	2	3	0	0	2.18	0.89	1.00	-1.18
	20	Improve	1	6	8	5	3	0	3.13	1.08	2.00	-1.13
	21	Asmt Crit	4	7	6	3	1	0	2.52	1.10	2.00	-0.52
	43	Cmnts Hlp	3	6	5	7	2	0	2.96	1.20	2.00	-0.96
Overall Eval	51	Bst Tchr	0	3	12	3	5	0	3.43	0.97	2.00	-1.43
	52	Bst Class	2	7	7	3	3	1	2.91	1.16	2.00	-0.91

Pride

Happy

Discuss.n

Control *

Handling *

Challenge

Clarity

Approachable

Prompt

Related

<Q1
Q6>

<Q2
Q7>

<Q3
Q8>

<Q4
Q9>

<Q5
Q10>

Enthuses

Humour

Impt Pts

Strict *

Disrupt. Std

<Q11
Q16>

<Q12
Q17>

<Q13
Q18>

<Q14
Q19>

<Q15
Q20>

Asmt Fair

Mat Prep

Know Wld

Chall/Intrst

Improve

Asmt Crit

Feel Gd

Hmwk Ovld *

Org/Expct

Try Hard

Too Much

Boys Qns *

Incr Think

Ignore Tch *

Esy Undrst

<Q21
Q26>

<Q22
Q27>

<Q23
Q28>

<Q24
Q29>

<Q25
Q30>

Gen Conc

No Hm Hlp *

Stdy Skls

Team

Adpts Ablty

<Q31
Q36>

<Q32
Q37>

<Q33
Q38>

<Q34
Q39>

<Q35
Q40>

Hmk Cpes

St Disply

Inclusive

Get Job

S Ask Qns

Motivates

Pronounce

Difficulty *

Cont Nxt Yr

Cmnts Hlp

Examples

On Time

Fin Early *

Speed *

Talk A Lot *

<Q41
Q46>

<Q42
Q47>

<Q43
Q48>

<Q44
Q49>

<Q45
Q50>

Bst Tchr

<Q51 Sex >

Bst Class

<Q52 Ethnic >

Ability >

READING THE DATA

The following notes are intended to assist you to read the data sheets that are attached.

There are a number of different data sheets for you, all of which I hope will increase your understanding of the student responses. These are:

1 Class data in order of statement (labelled Page 15 at the bottom)
2 Class data according to the dimensions shown to you at the initial interview (Page 16)
3 Graphs which show the frequency of student responses to each statement in the order of the questionnaire (Pages 19-24)
4 Graphs which show the demographics of the respondents (on Page 24).

All of the above have a code like SCHD08.XLS (a code is unique for each teacher according to the three characters before the decimal point)at the top of the page, which means that they represent the responses of the whole class.

In addition, there may be what appears to be repeats of this information, but which in reality represent the subgroups within your class, where such subgroups have at least five members. These repeats are indicated by the following codes at the top of the pages:

Male	SCHD08M.XLS
Female	SCHD08F.XLS
European	SCHD08E.XLS
Maori	SCHD08MA.XLS
Pacific Island	SCHD08P.XLS
Asian	SCHD08A.XLS
Other	SCHD08O.XLS
Top	SCHD08T.XLS
Middle	SCHD08MI.XLS
Bottom	SCHD08B.XLS

Not all sets of data will contain these reports, so do not worry about that. By examining the graphs on Page 24 you can tell whether there are enough students in these subsets to justify the printing of these reports. Also, if a substantial majority of a class belonged to only one category, then there may be no point in printing the reports for that group - for example, if 20

out of 24 students were European there will be no separate European report.

Page 15

Top Line is a set of headers showing the categories scored on the questionnaire plus headings for some calculated information:

1 2 3 4 5 6   Average Teacher Difference

Second Line:

1   Pride   5 5 5 0 2 0   2.35   2   0.35

Question 1
A short hand summary for the statement
The number of students who scored you in each category
The average of the student responses to categories 1 to 5 only.
Your response.
The difference between the last two columns.

The other lines follow this format, except the ones that are marked with an asterisk after the summary word. eg.

3   Handling* 1 0 3 7 6 0   2.00   4   0.00

In these statements you would expect students to disagree with the statement, so to keep the results consistent I have **reverse scored** these. In other words, a student response of 1 counts as 5, 2 counts as 4, and so on for the purpose of finding the average. Your response is coded as you responded but also reversed for calculation purposes.

In general, a low average is desirable, but this is not always the case, and there may be good reasons for this. You are in the best position to judge that.

The graphs should be self explanatory, and simply reflect the data shown in the tables, in the order of the questionnaire.

USING

STUDENT EVALUATIONS

FOR

TEACHER DEVELOPMENT

RESEARCH BY EARL IRVING

CONTENTS

* *ATTITUDE TO SUBJECT*

* *UTILITY OF SUBJECT*

* *WORKLOAD/DIFFICULTY*

* *ORGANISATION/CLARITY*

* *OTHER STUDENTS/GROUP INTERACTION*

* *RAPPORT*

* *HOMEWORK*

* *ENTHUSIASM*

* *LEARNING/VALUE*

* *CONTROL*

* *EXAMS/GRADING*

* *OVERALL EVALUATIONS*

* *OPEN ENDED QUESTIONS* (free student response)

* *BACKGROUND STUDENT/SUBJECT/CLASS CHARACTERISTICS*

ATTITUDE TO SUBJECT

2 I am usually proud of the work I do in this subject.

3 The thought of coming to this class makes me feel good.

5 In this class I usually try to do as well as I can.

6 I want to keep doing this subject after this year because I am enjoying it.

Deleted:

4 I don't enjoy many of the activities we do in this subject.

7 This is a hard/easy(?) subject.

1 I usually like this subject

UTILITY OF SUBJECT

3 If I do well in this subject, it will help me to get a job. ,

4   The teacher shows us how this subject is related to the other subjects taught at school.

Deleted:

1 I find this subject less useful than other subjects.

2 I don't expect to make much use of what I learn in this subject.

WORKLOAD/DIFFICULTY

1 This subject teacher does not allow me to work at my own speed.

3 In this subject, students who finish their work early are usually left with nothing to do.

5 I find it hard to keep up in this subject.

6 This teacher can adapt (change) teaching and language to the level of students' ability.

8 The teacher attempts to cover too much material in class.

Deleted:

2 I can usually handle the work I am given in this subject.

4 In this class I don't try to do work that I find difficult.

7 This course has a heavy workload.

ORGANISATION/CLARITY

1   This subject teacher always explains clearly what she/he would like us to do.

3   My teacher is able to make this subject easier for me to understand.

5   The aims, objectives and structure of this subject class were clearly outlined at the start of the course.

6   The class materials we use in class are well prepared and help me to learn this subject.

7   My teacher arrives at class on time.

9   My teacher emphasizes the important points to learn.

Deleted:

2   Activities in this subject are carefully planned.

4   In lessons in this subject I usually don't understand the work we are given.

8   The students in this class have a choice in the topics studied.

10   The teacher is flexible and can change the lesson to meet new situations.

11   The teacher makes good use of examples or personal experiences to help get the point across.

12   Laboratory (workshop/practical??) lessons are well prepared and help me to understand this subject. (Is this point covered by general lesson preparation above?)

OTHER STUDENTS/GROUP INTERACTION

2 I am happy to work with most of the students in this class.

4 Students in this class are encouraged to take part in class discussions.

5 In this class we are encouraged to ask questions, and the teacher answers them well.

7 Class members are encouraged to work as a team.

Deleted:

1 I like most of the students in this class.

3 Many of the students in this class just want to waste time.

6 The teacher uses student ideas in teaching this class.

RAPPORT

2   This subject teacher is not interested in whether or not the students can handle the work.

5   Boys are more likely to be asked to answer questions during lessons in this subject.

6   I feel that I can talk to this subject teacher about things that are on my mind.

9   The teacher uses language and examples that include all students (i.e. non-racist, non-sexist etc)

10   This teacher seems to be genuinely concerned for my progress.

12   Student work is displayed in our classroom, and changes during the year.

14   The teacher tries to pronounce our names properly.

Deleted:

1   This subject teacher is very understanding.

3   This subject teacher rarely helps students who are having difficulties with school work.

4   This class teacher seems to like most of the students in this class.

7   The teacher of this class embarrasses students who don't know the right answer.

8   My teacher in this subject is not interested in our opinions.

11   This teacher seems to have favourites in the class.

13   The teacher gets angry and shouts when we are naughty.

HOMEWORK

1 I can usually cope with the homework we are given in this subject.

3 If I needed help to my homework in this subject, I would not be able to get it.

5 We get too much homework in this subject.

Deleted:

2 Doing homework in this subject helps me to understand this subject.

4 My parents do not encourage me to do my homework.

ENTHUSIASM

3 The teacher is enthusiastic when teaching this class.

4 My teacher makes the lessons more interesting with a good sense of humour.

6 The teacher has motivated me to do my best work.

Deleted:

1 Lessons in this subject are usually boring.

2 This subject teacher usually arranges interesting things for us to do.

5 My teacher likes teaching this class.

LEARNING/VALUE

4 I find this class challenging and stimulating.

5   The teacher in this class talks too much, and we don't get enough time to do our work.

6 In this class we are taught study skills special to this subject.

7 The teacher often asks us challenging questions that really make me think.

8 The teacher makes good use of examples and illustrations.

10 This class has increased my skills in thinking.

11 In this class I have learned how this subject relates to other subjects.

Deleted:

1 I feel that I learn a lot in this subject.

2 Most things we learn in this subject are interesting.

3 What we do in this subject will help me understand more of the world around us.

9 I have gained new viewpoints and appreciations in this class.

CONTROL

1 The teacher is not able to control the students in this class.

2 The students in this class don't pay attention to the teacher.

3 Some of the students in this subject class prevent me from working.

7 My teacher is strict.

Deleted:

4 Rules in this class are kept to a minimum.

5 It is easy to side-track the teacher.

6 In this class we can get away without doing much work.

EXAMS/GRADING

1   Criteria for assessment are clearly stated for this subject.

3   The teacher in this subject always marks and returns assignments and tests promptly.

4   The teachers comments on my tests and assignments are constructive and helpful.

5   Tests and assignments in this subject are fair and cover the work we have been taught.

9   My teacher helps me to improve my performance in this subject.

Deleted:

2   I know what I have to do to get good grades in this subject.

6   It is easy to get good grades in this subject.

7   There are too many tests and assignments in this subject.

8   The teacher expects us to do well in this subject.

10   In this class, too much emphasis is placed on getting good results in the external exams.

OVERALL EVALUATIONS

1 Overall, this teacher is one of the best teachers I have this year.

2 Overall, this subject is one of the best classes I have this year.

Deleted:

3 This is an excellent class.

OPEN ENDED QUESTIONS (free student response)
[NB: Responses in this section will be typed up verbatim before the teacher receives them.]

2 What do you like about this class?

3 What do you think can be improved in this class?

Deleted: ,

1   What questions do you think should be in the above sections? Has anything important been missed out?

4 What suggestions would you make to improve this class?

BACKGROUND STUDENT/SUBJECT/CLASS CHARACTERISTICS

School

Class

Sex

Ethnic (Pakeha/Maori/Pacific Islands/Asian/Other)

Age

If this class was divided into three groups on the basis of marks/grades, would you expect to be in the top, middle or bottom group?

STUDENT EVALUATION FORM

School _____   Class _____   Date _____

Sex: Ethnic: Ability:
M F   E M P A O   T M B

Please complete all the questions by circling the number which applies, using a pen. Put a X through any mistake, and circle the number you want to be counted.

If the question does not apply to you, please use the NOT APPLICABLE column.

		1. Strongly Agree	2. Agree	3. Neither Agree/Disagree	4. Disagree	5. Strongly Disagree	6. Not Applicable
01	I am usually proud of the work I do in this class.	1	2	3	4	5	6
02	Students in this class are encouraged to take part in class discussions.	1	2	3	4	5	6
03	This subject teacher is not interested in whether the students can handle the work.	1	2	3	4	5	6
04	This subject teacher always explains clearly what she/he would like us to do.	1	2	3	4	5	6
05	The teacher in this subject always marks and returns assignments and tests promptly.	1	2	3	4	5	6	5
06	I am happy to work with most of the students in this class.	1	2	3	4	5	6
07	The teacher is not able to control the students in this class.	1	2	3	4	5	6
08	The teacher often asks us challenging questions that really make me think.	1	2	3	4	5	6
09	I feel that I can talk to this subject teacher about things that are on my mind.	1	2	3	4	5	6
10	The teacher shows us how this subject is related to the other subjects taught at school.	1	2	3	4	5	6	10

1	2	3	4	5	6

11 The teacher is enthusiastic when teaching this class.

| 1 | 2 | 3 | 4 | 5 | 6 |

12 My teacher makes the lessons more interesting with a good sense of humour.

| 1 | 2 | 3 | 4 | 5 | 6 |

13 My teacher emphasizes the important points to learn.

| 1 | 2 | 3 | 4 | 5 | 6 |

14 My teacher is strict.

| 1 | 2 | 3 | 4 | 5 | 6 |

15 Some of the students in this subject class prevent me from working.

| 1 | 2 | 3 | 4 | 5 | 6 | 15

16 Tests and assignments in this subject are fair and cover the work we have been taught.

| 1 | 2 | 3 | 4 | 5 | 6 |

17 The materials we use in class are well prepared and help me to learn this subject.

| 1 | 2 | 3 | 4 | 5 | 6 |

18 What we do in this subject will help me understand more of the world around us.

| 1 | 2 | 3 | 4 | 5 | 6 |

19 I find this class challenging and interesting.

| 1 | 2 | 3 | 4 | 5 | 6 |

20 My teacher helps me to improve my performance in this subject.

| 1 | 2 | 3 | 4 | 5 | 6 | 20

21 Criteria for assessment are clearly stated for this subject.

| 1 | 2 | 3 | 4 | 5 | 6 |

22 We get too much homework in this subject.

| 1 | 2 | 3 | 4 | 5 | 6 |

23 In this class I usually try to do as well as I can.

| 1 | 2 | 3 | 4 | 5 | 6 |

24 Boys are more likely to be asked to answer questions during lessons in this subject.

| 1 | 2 | 3 | 4 | 5 | 6 |

25 The students in this class don't pay attention to the teacher.

| 1 | 2 | 3 | 4 | 5 | 6 | 25

26 The thought of coming to this class makes me feel good.

| 1 | 2 | 3 | 4 | 5 | 6 |

27 At the start of the course, the teacher told us about the course, its organisation and what is expected of us.

| 1 | 2 | 3 | 4 | 5 | 6 |

					6. Not Applicable
				5. Strongly Disagree	
			4. Disagree		
		3. Neither Agree/Disagree			
	2. Agree				
1. Strongly Agree					

28 The teacher attempts to cover too much material in class.

1 2 3 4 5 6

29 This class has increased my skills in thinking.

1 2 3 4 5 6

30 My teacher is able to make this subject easier for me to understand.

1 2 3 4 5 6 30

31 This teacher seems to be genuinely concerned for my progress.

1 2 3 4 5 6

32 If I needed help with my homework in this subject, I would not be able to get it.

1 2 3 4 5 6

33 In this class we are taught study skills special to this subject.

1 2 3 4 5 6

34 Class members are encouraged to work as a team.

1 2 3 4 5 6

35 This teacher can change their teaching and language to suit the level of students' ability.

1 2 3 4 5 6 35

36. I can usually cope with the homework we are given in this subject.

1 2 3 4 5 6

37 Student work is displayed in our classroom, and ↑ is changed during the year.

1 2 3 4 5 6

38 The teacher uses language and examples that include all students (i.e. non-racist, non-sexist, etc.).

1 2 3 4 5 6

39 If I do well in this subject, it will help me to get a job.

1 2 3 4 5 6

40 In this class we are encouraged to ask questions, and the teacher answers them well.

1 2 3 4 5 6 40

41 The teacher has motivated me to do my best work.

1 2 3 4 5 6

42 I find it hard to keep up in this subject.

1 2 3 4 5 6

43 The teacher's comments on my tests and assignments
    are constructive and helpful.   1 | 2 | 3 | 4 | 5 | 6

44 My teacher arrives at class on time.   1 | 2 | 3 | 4 | 5 | 6

45 This subject teacher does not allow me to work
    at my own speed.   1 | 2 | 3 | 4 | 5 | 6   45

46 The teacher tries to pronounce our names properly.   1 | 2 | 3 | 4 | 5 | 6

47 I want to keep doing this subject after this year
    because I am enjoying it.   1 | 2 | 3 | 4 | 5 | 6

48 The teacher makes good use of examples
    and illustrations.   1 | 2 | 3 | 4 | 5 | 6

49 In this subject, students who finish their work early
    are usually left with nothing to do.   1 | 2 | 3 | 4 | 5 | 6

50 The teacher in this class talks too much, and we
    don't get enough time to do our work.   1 | 2 | 3 | 4 | 5 | 6   50

51 Overall, this teacher is one of the best teachers
    I have this year.   1 | 2 | 3 | 4 | 5 | 6

52 Overall, this subject is one of the best classes I
    have this year.   1 | 2 | 3 | 4 | 5 | 6

FREE RESPONSE

[NB: Your responses to these two questions will be typed up word-for-word before your teacher
receives them].

53 What do you like about this class?

54 What do you think can be improved in this class?

USING STUDENT EVALUATIONS FOR TEACHER DEVELOPMENT

FOLLOW UP QUESTIONNAIRE - STUDENTS

1 Using student evaluations like this one is worthwhile.		1	2	3	4	5
2 It felt good to be asked my opinion about my teacher and their teaching.		1	2	3	4	5
3 The teacher thanked the class for the information the questionnaire provided.		1	2	3	4	5
4 My teacher discussed the results of the first evaluation with the class.		1	2	3	4	5
5 I think that our teacher **tried to improve** the way that the class was taught as a result of the first evaluation.		1	2	3	4	5
6 I think that our teacher **did improve** the way that the class was taught as a result of the first evaluation.		1	2	3	4	5
7 I think that some students were "picked on" by the teacher as a result of the first evaluation.		1	2	3	4	5
8 I would like to complete an evaluation like this one on all my teachers once a year.		1	2	3	4	5

9 Comment on any aspect of this project, the use of student evaluations like these and how you felt about being involved in evaluating your teacher that you feel will help.

```



```

CONSULTATION PROCEDURES

Share the data
This is the way the students in your class see you and your teaching.

Discuss the teacher perception first
What are your general impressions of the results?

What aspects of the results are you pleased with? List item numbers.
    Can you think of anything about your teaching that leads the students to
    describe you in this way? Are these items important for student learning?

Which items do you wish to comment on, just to clarify the situation. What needs
explaining? List item numbers.

What aspects of the results are you not pleased with? List item numbers.
    Can you think of anything about your teaching that leads the students to .
    describe you in this way? Are these items important for student learning?

Identify the areas for change
.What two or three items would you most like to change? Why?

Are there any institutional barriers that need to be removed for this change to take
place?

Write the action plan
What practical changes can you make that will improve this behaviour?

Write them down. The changes should include any resources that are needed, courses
that might help, any key personnel that should/could be consulted, etc. Refer to Marsh
and Roche (1994) for additional ideas that may help - present these as suggestions.
A timetable for the implementation of the changes should be discussed.

Finish the interview
Encourage the teacher to put ideas into action, and seek support if required. Thank
teacher for participation, and outline next steps in process.

Follow up
Notes made during the conference to be typed up. Copy of the action plan for the
consultant and another copy posted to the teacher, together with the timetable for the
next class meeting set.
Refer: Acheson and Gall, 1992; Wilson, 1986.

The Use of Student Evaluations for Teacher Development

A Research Project conducted by Earl Irving

Teacher Questionnaire

Please indicate your response by placing a tick in the appropriate space.

This research project involved a student questionnaire, written feedback and a discussion with the researcher to focus on action for teacher development.

1 The Student Questionnaire

a To what extent do you think that the student evaluation form covered the teaching job that you do in the classroom?

Comprehensively ☐   To some extent ☐   Not at all ☐

Any Comments? _____

_____

b Do you think that the student evaluation form was a valid way of evaluating you as a teacher?

Extremely valid ☐   Of some validity ☐   Invalid ☐

Any Comments? _____

_____

c What features of the student evaluation questionnaire did you like?

_____

_____

d What features of the student evaluation questionnaire did you dislike?

_____

_____

e What questions did you think were the most useful? (Use the numbers
on the questionnaire form.)

_____

Any Comments?   _____
_____

f What questions did you think were the least useful? (Use the numbers
on the questionnaire form.)

_____

Any Comments?   _____
_____

g "Students are not competent to evaluate certain aspects of teaching
(such as the teacher's subject knowledge), and these matters should not be
in a student evaluation questionnaire."
To what extent do you agree with this statement?

Fully agree ☐   Neither agree/disagree ☐   Totally disagree ☐
Any Comments?   _____
_____

2 **Written Feedback**

a How useful was the written feedback you received in terms of planning
for teacher development?

Very useful ☐   Of some use ☐   No use ☐
Any Comments?   _____
_____

b How much credibility do you think student evaluation data has?

Very credible ☐   Some credibility ☐   Not credible ☐
Any Comments?   _____
_____

From which class would you give more credibility to feedback?

Senior class ☐   Junior class ☐
Any class ☐   No class level ☐
Any Comments? _____

_____

c Which data did you find the more useful?

     Numerical tables ☐   Verbatim written comments ☐
Any Comments? _____

_____

d To what extent do you think that the feedback reflected your teaching of the class concerned?

Accurately ☐   With some accuracy ☐   Inaccurately ☐
Any Comments? _____

_____

e Did you discuss the results of the evaluation with your class?

Yes ☐   No ☐

Describe their response?

_____

_____

Do you think that this discussion helped?

Yes ☐   No ☐
Any Comments? _____

_____

f Did the information you received reinforce any particular approaches you have to teaching?

Considerably ☐   To some extent ☐   Not at all ☐
Any Comments? _____

_____

3 The Consultation

a How useful was the consultation in helping you to benefit from the student evaluation data?

Very useful ☐   Of some use ☐   Not useful ☐
Any Comments? _____

_____

b How useful was the consultation in terms of receiving support?

Very useful ☐   Of some use ☐ ,   Not useful ☐
Any Comments? _____

_____

c How useful was the consultation in terms of planning your action for teacher development?

Very useful ☐   Of some use ☐   Not useful ☐
Any Comments? _____

_____

d The researcher was deliberately non-judgmental in the consultation, and sought to help you interpret the data and plan for improvement. To what extent do you think this approach is helpful.

Very helpful ☐   Of some help ☐   Not helpful ☐
Any Comments? _____

_____

e In almost all cases you received the student evaluation data prior to the consultation so that you had time to consider it. How helpful was this?

Very helpful ☐   Of some help ☐   Not helpful ☐
Any Comments?   _____
                     _____

f At the conclusion of the consultation you received a written report covering the main points covered in the discussion. To what extent do you think that this was helpful?

Very helpful ☐   Of some help ☐   Not helpful ☐
Any Comments?   _____
                     _____

g How important is it to complement student feedback with some form of collegial consultation?

Very important ☐   Of some importance ☐   Not important ☐
Any Comments?   _____
                     _____

4 Teacher Development

a The purpose of the process was to help you improve some aspect(s) of your teaching using information from students about your teaching. To what extent do you think this purpose is worthwhile?

Very worthwhile ☐   Some worthiness ☐   Not worthwhile ☐
Any Comments?   _____
                     _____

b To what extent do you think that the student feedback has been a factor leading to improvements in your teaching?

Very useful ☐   Some use ☐   Not useful ☐
Any Comments?   _____
                     _____

c What factors influenced you to change your teaching practice and why?

_____

_____

d To what extent did you **try to improve** the identified aspects of your teaching?

Very hard ☐   To some extent ☐   Not at all ☐
Any Comments? _____

      _____

e To what extent do you think that you **actually did improve** those aspects of your teaching?

Considerably ☐   To some extent ☐   Not at all ☐
Any Comments? _____

      _____

f To what extent do you think that the learning of your students has been enhanced by the student feedback process?

Greatly ☐   To some extent ☐   Not at all ☐
Any Comments? _____

      _____

5 Your reactions

a Describe your reaction to having participated in this research project?

_____

_____

b Describe your reaction when you received the results of the student evaluation.

_____

_____

c To what extent did participation in this project enhance or constrain your teaching?

Greatly enhanced ☐   Enhanced a little ☐   Constrained ☐
Any Comments?   _____
                        _____

d What impact has this appraisal had on your job satisfaction and morale?

Great impact ☐   Some impact ☐   No impact ☐
Any Comments?   _____
                        _____

6 Evaluation of this project

a What are the benefits and disadvantages you think you received as a result of this project?
Benefits:_____
_____
Disadvantages:_____
_____

Any Comments?   _____
                        _____

b What suggestions would you make to improve the process.

_____
_____

c To what extent do you think that this type of teacher evaluation has a part to play in the development of better teachers in schools?

Considerable ☐   Some ☐   None ☐
Any Comments?   _____
                        _____

d To what extent do you think that the process would have been more effective if it had been supported by a training component?

Considerably ☐   Somewhat ☐   Not at all ☐
Any Comments? _____

_____

7 Teacher Appraisal

This research project sits within the context of teacher appraisal.

a What do you think is the purpose of teacher appraisal?

    Teacher Development (Formative evaluation)   ☐

    Personnel Decisions (Summative evaluation)
        - hiring/firing, promotion, competency etc   ☐

    Both of these.   ☐

Any Comments? _____

_____

b Does your school have an teacher appraisal scheme?

      Yes ☐   No ☐

    Is participation compulsory?

      Yes ☐   No ☐

Any Comments? _____

_____

What is the purpose of that scheme?

Teacher Development (Formative evaluation) ☐

Personnel Decisions (Summative evaluation)
- hiring/firing, promotion, competency etc ☐

Both of these. ☐

Any Comments? _____

c As a result of your experience in this project, to what extent should student evaluations be incorporated in an appraisal scheme?

To a large extent ☐   To some extent ☐   Not at all ☐
Any Comments? _____
_____

Have you changed your mind about this since entering into this project?

Yes ☐   No ☐
Any Comments? _____
_____

d Considering other teacher development strategies you have been involved in, what features of this project do you like/prefer or dislike compared with the others?

Like/Prefer:_____
_____

Dislike:_____
_____

Any Comments? _____
_____

e Teacher appraisal is a time consuming business. To what extent do you think that the time demands of this form of teacher appraisal (ie student evaluations) are reasonable given the potential outcomes?

Very reasonable ☐   Reasonable ☐   Unreasonable ☐
Any Comments?   _____
                     _____

f As part of the process you were offered the right to have a support person with you at the consultation. Do you think that this is a desirable feature of an appraisal system?

Very desirable ☐   Desirable ☐   Undesirable ☐
Any Comments?   _____
                     _____

g Imagine that your school has decided to introduce an appraisal scheme which uses student evaluations. List three features that you would want to include in, and three features you would want to exclude from the new scheme.

Include: 1 _____
             2 _____
             3 _____

Exclude: 1 _____
             2 _____
             3 _____

Any Comments?   _____
                     _____

h To what extent should the results of student evaluations be available to course supervisors/HODs/senior management team?

Freely available ☐   Available at your discretion ☐   Unavailable ☐
Any Comments?   _____
                     _____

i For the purpose of teacher development, how often do you think that a teacher should collect and act on student evaluations?

Every term ☐   Every year ☐   Every two years ☐   Never ☐
Any Comments? _____
_____

j From whom should this data be obtained?
All classes ☐   A sample of classes ☐   One class ☐

Any Comments? _____
_____

8 General comments
Feel free to make any comments that you wish about this project or about teacher appraisal in general.

_____
_____
_____
_____
_____
_____
_____

In conclusion

> **"Student feedback is the most important source of feedback for improving teaching."**

To what extent do you agree with this statement?

Totally agree	☐
Agree	☐
Neither agree/disagree	☐
Disagree	☐
Totally disagree	☐

Any comments: _____

_____

Thank you for your participation in this project and the time you have put into these responses.

Appendix 4.1

<u>A Priori Factors Used in Student Questionnaire</u>

The questionnaire items were arranged into the following *a priori* factors:

Learning/Value
>   08 The teacher often asks us challenging questions that really make me think.
>   18 What we do in this subject will help me understand more of the world around us.
>   19 I find this class challenging and interesting.
>   29 This class has increased my skills in thinking.
>   33 In this class we are taught study skills special to this subject.
>   48 The teacher makes good use of examples and illustrations.
>   50 The teacher in this class talks too much, and we don't get enough time to do our work.

Organisation/Clarity
>   04 This subject teacher always explains clearly what she/he would like us to do
>   13 My teacher emphasises the important points to learn.
>   17 The materials we use in class are well prepared and help me to learn this subject.
>   27 At the start of the course, the teacher told us about the course, its organisation and what is expected of us.
>   30 My teacher is able to make this subject easier for me to understand.
>   44 My teacher arrives at class on time.

Teacher (instructor) Enthusiasm
>   11 The teacher is enthusiastic when teaching this class.
>   12 My teacher makes the lessons more interesting with a good sense of humour
>   41 The teacher has motivated me to do my best work.

Other Students (group interaction)
>   02 Students in this class are encouraged to take part in class discussions.
>   06 I am happy to work with most of the students in this class.
>   34 Class members are encouraged to work as a team.
>   40 In this class we are encouraged to ask questions, and the teacher answers them well.

Individual Rapport
>   03 This teacher is not interested in whether the students can handle the work.
>   09 I feel that I can talk to this subject teacher about things that are on my mind.
>   24 Boys are more likely to be asked to answer questions during lessons in this subject.
>   31 This teacher seems to be genuinely concerned for my progress.
>   37 Student work is displayed in our classroom and is changed during the year.
>   38 The teacher uses language and examples that include all students.
>   46 The teacher tries to pronounce our names properly.

Assessment (examinations/grading)

05 The teacher in this subject always marks and returns assignments and tests promptly.

16 Tests and assignments in this subject are fair and cover the work we have been taught.

20 My teacher helps me to improve my performance in this subject.

21 Criteria for assessment are clearly stated for this subject.

43 The teacher's comments on my tests and assignments are constructive and helpful.

Workload/Difficulty

28 The teacher attempts to cover too much material in class.

35 This teacher is able to change their teaching and language to suit the level of students' ability.

42 I find it hard to keep up in this subject.

45 This subject teacher does not allow me to work at my own speed.

49 In this subject, students who finish their work early are usually left with nothing to do.

Classroom Control

07 The teacher is not able to control the students in this class.

14 My teacher is strict

15 Some of the students in this subject class prevent me from working.

25 The students in this class don't pay attention to the teacher.

Homework

22 We get too much homework in this subject.

32 If I needed help with my homework in this subject, I would not be able to get it.

36 I can usually cope with the homework we are given in this subject.

Overall Evaluations

51 Overall, this teacher is one of the best teachers I have this year.

52 Overall, this subject is one of the best classes I have this year.

Attitude

01 I am usually proud of the work I do in this class.

23 In this class I usually try to do as well as I can.

26 The thought of coming to this class makes me feel good.

47 I want to keep doing this subject after this year because I am enjoying it.

Utility

10 The teacher shows us how this subject is related to the other subjects taught at school.

39 If I do well in this subject, it will help me to get a job.

Appendix 4.2

Rotated Factor Matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
09 Approachable	0.63449	0.29429	-0.09259	0.10009	-0.03617	-0.02232	0.01077
30 Esy Undrst	0.60706	0.44289	-0.08880	0.09871	-0.23619	-0.01700	0.05407
31 Gen Conc	0.60369	0.40149	-0.09653	0.18781	0.03088	-0.09816	0.08315
35 Adpts Ablty	0.54688	0.11702	-0.07792	0.14676	-0.16061	0.23622	-0.00578
03 Handling *	-0.53811	-0.15650	0.33123	-0.03668	0.11864	-0.18561	0.12885
12 Humour	0.51323	0.15989	0.01624	0.35018	-0.10041	0.22966	0.05835
40 S Ask Qns	0.49007	0.48781	-0.08630	0.08028	-0.08480	0.22784	0.04534
11 Enthuses	0.48645	0.21267	0.01716	0.32047	0.03733	0.29300	0.13989
48 Examples	0.46804	0.25124	0.02227	0.16813	-0.16230	0.23739	0.25089
04 Clarity	0.44405	0.25984	-0.12060	0.24166	-0.21751	0.03022	-0.02615
38 Inclusive	0.43143	-0.02233	-0.27744	0.02007	-0.13252	0.08657	0.11428
13 Impt Pts	0.39971	0.36177	-0.20643	0.12929	-0.16157	0.20822	0.17999
41 Motivates	0.43633	0.65653	-0.14424	0.02398	-0.06487	0.05687	-0.00931
29 Incr Think	0.14058	0.64821	-0.18063	0.15696	-0.14978	0.15890	0.06620
20 Improve	0.43069	0.62767	-0.16091	0.10494	-0.14443	-0.02478	0.06673
19 Chall/Intrst	0.21862	0.59607	-0.02629	0.32896	-0.14876	0.10285	0.17743
08 Challenge	0.13164	0.55657	-0.12451	0.09996	0.25808	0.24604	0.15960
33 Stdy Skls	0.22847	0.49662	0.02734	-0.02781	-0.12852	0.29058	-0.06857
18 Know wld	0.02858	0.45709	0.17565	0.42456	-0.25291	0.21569	0.14103
25 Ignore Tch*	-0.16671	-0.07152	0.72263	-0.11992	0.04252	-0.08672	-0.09668
49 Fin Early *	-0.01243	-0.16387	0.69425	0.00965	0.00426	-0.11512	0.16974
07 Control *	-0.19851	-0.06400	0.65026	-0.17515	0.04540	-0.02782	-0.15906
50 Talk A Lot *	-0.06063	-0.18520	0.49325	-0.16925	0.38978	0.27231	-0.19635
05 Prompt	0.18035	0.07881	-0.16039	0.70525	-0.03596	0.15910	0.00993
44 On Time	0.19878	0.31502	-0.11356	0.56476	-0.05744	0.24990	-0.01411
17 Mat Prep	0.39214	0.31947	-0.15320	0.43147	-0.20131	-0.09714	0.09661
46 Pronounce	0.36338	-0.01036	-0.20980	0.40825	0.04588	-0.09148	-0.01627
42 Difficulty *	0.02350	-0.29724	0.02495	-0.17007	0.69613	0.01496	-0.02624
28 Too Much	-0.21234	0.03838	0.08540	0.17285	0.69402	-0.21161	-0.00882
45 Speed *	-0.26237	-0.08322	0.09328	-0.14172	0.59178	-0.04565	-0.01213
34 Team	0.09219	0.20520	-0.10627	0.10617	-0.11881	0.71303	-0.02813
02 Discuss.n	0.14758	0.18317	-0.07014	0.13858	-0.02132	0.68622	0.13949
06 Happy	-0.05849	0.15532	0.00025	0.06896	0.03299	0.00815	0.74887
15 Disrupt. Std	-0.09177	0.08861	0.34868	0.13449	0.25442	0.03169	-0.51483
27 Org/Expct	0.43771	0.20725	0.02476	-0.03462	0.09678	0.24942	0.45414
21 Asmt Crit	0.33822	0.05410	-0.03710	0.22992	-0.03066	0.33749	0.40782

<u>Post Facto Factors Determined Using Factor Analysis</u>

The full wording of each of the items making up each factor (in the order the items appear in the questionnaire) is:

Factor One: General Teaching Skills
    03 This teacher is not interested in whether the students can handle the work.
    04 This subject teacher always explains clearly what she/he would like us to do
    09 I feel that I can talk to this subject teacher about things that are on my mind.
    11 The teacher is enthusiastic when teaching this class.
    12 My teacher makes the lessons more interesting with a good sense of humour
    13 My teacher emphasises the important points to learn.
    30 My teacher is able to make this subject easier for me to understand.
    31 This teacher seems to be genuinely concerned for my progress.
    35 This teacher is able to change their teaching and language to suit the level of
        students' ability.
    38 The teacher uses language and examples that include all students.
    40 In this class we are encouraged to ask questions, and the teacher answers
        them well.
    48 The teacher makes good use of examples and illustrations.

Factor Two: Value of Learning
    08 The teacher often asks us challenging questions that really make me think.
    18 What we do in this subject will help me understand more of the world
        around us.
    19 I find this class challenging and interesting.
    20 My teacher helps me to improve my performance in this subject.
    29 This class has increased my skills in thinking.
    33 In this class we are taught study skills special to this subject.
    41 The teacher has motivated me to do my best work.

Factor Three: Classroom Management and Control
    07 The teacher is not able to control the students in this class.
    25 The students in this class don't pay attention to the teacher.
    49 In this subject, students who finish their work early are usually left with
        nothing to do.
    50 The teacher in this class talks too much, and we don't get enough time to do
      our work.

Factor Four: Teacher Organisation
    05 The teacher in this subject always marks and returns assignments and tests
        promptly.
    17 The materials we use in class are well prepared and help me to learn this
        subject.
    44 My teacher arrives at class on time.
    46 The teacher tries to pronounce our names properly.

Factor Five: **Workload/Difficulty**

28 The teacher attempts to cover too much material in class.

42 I find it hard to keep up in this subject.

45 This subject teacher does not allow me to work at my own speed.

Factor Six: **Student Cooperation**

02 Students in this class are encouraged to take part in class discussions.

34 Class members are encouraged to work as a team.

Factor Seven: **Student Contentment**

06 I am happy to work with most of the students in this class.

15 Some of the students in this subject class prevent me from working.

21 Criteria for assessment are clearly stated for this subject.

27 At the start of the course, the teacher told us about the course, its organisation and what is expected of us.

		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	2	12	6	1	0	0	2.29	0.70	2.26	0.61	0.02	0.04	2.00	0.26
2	Discuss.n	3	10	6	2	0	0	2.33	0.84	2.39	1.13	-0.06	-0.05	2.00	0.39
3	Handling *	2	1	6	8	4	0	2.48	1.14	2.95	1.07	-0.47	-0.44	4.00	-1.05
4	Clarity	3	12	2	4	0	0	2.33	0.94	2.83	1.13	-0.49	-0.44	1.00	1.83
5	Prompt	1	4	5	9	2	0	3.33	1.04	2.96	1.33	0.38	0.28	3.00	-0.04
6	Happy	3	14	2	1	1	0	2.19	0.91	2.04	0.91	0.15	0.16	2.00	0.04
7	Control *	0	0	1	11	9	0	1.62	0.58	1.77	0.73	-0.15	-0.21	5.00	-3.23
8	Challenge	1	12	6	2	0	0	2.43	0.73	2.87	0.99	-0.44	-0.44	2.00	0.87
9	Approachable	0	1	8	10	2	0	3.62	0.72	3.91	1.08	-0.29	-0.27	3.00	0.91
10	Related	0	7	9	2	3	0	3.05	1.00	3.73	0.91	-0.68	-0.74	4.00	-0.27
11	Enthuses	2	9	7	3	0	0	2.52	0.85	3.09	0.83	-0.56	-0.68	2.00	1.09
12	Humour	2	9	5	4	1	0	2.67	1.04	3.04	1.20	-0.38	-0.31	2.00	1.04
13	Impt Pts	2	13	5	1	0	0	2.24	0.68	2.09	0.79	0.15	0.19	1.00	1.09
14	Strict *	3	3	11	1	3	0	3.10	1.15	3.09	1.08	0.00	0.00	2.00	1.09
15	Disrupt. Std	1	5	5	9	1	0	3.19	1.01	2.96	1.12	0.23	0.21	2.00	0.96
16	Asmt Fair	2	11	7	1	0	0	2.33	0.71	2.18	0.89	0.15	0.17	1.00	1.18
17	Mat Prep	0	8	9	3	0	1	2.75	0.70	2.48	1.10	0.27	0.25	2.00	0.48
18	Know Wld	3	12	5	1	0	0	2.19	0.73	2.81	1.30	-0.62	-0.48	3.00	-0.19
19	Chall/Intrst	0	4	13	4	0	0	3.00	0.62	3.09	1.02	-0.09	-0.09	2.00	1.09
20	Improve	1	10	6	4	0	0	2.62	0.84	3.13	1.08	-0.51	-0.48	2.00	1.13
21	Asmt Crit	6	7	6	0	0	1	2.00	0.79	2.52	1.10	-0.52	-0.48	2.00	0.52
22	Hmwk Ovld *	2	3	7	8	1	0	2.86	1.04	2.96	1.00	-0.10	-0.10	3.00	-0.04
23	Try Hard	2	12	5	2	0	0	2.33	0.78	2.22	0.88	0.12	0.13	2.00	0.22
24	Boys Qns *	0	1	7	4	5	3	2.24	0.94	2.48	1.01	-0.24	-0.24	5.00	-2.52
25	Ignore Tch *	0	1	5	12	3	0	2.19	0.73	2.26	0.79	-0.07	-0.09	3.00	-0.74
26	Feel Gd	0	5	8	6	2	0	3.24	0.92	3.04	1.16	0.19	0.17	2.00	1.04
27	Org/Expct	2	8	4	6	1	0	2.81	1.10	2.78	1.18	0.03	0.02	1.00	1.78
28	Too Much	0	2	9	9	1	0	3.43	0.73	2.83	1.09	0.60	0.55	4.00	-1.17
29	Incr Think	1	10	10	0	0	0	2.43	0.58	2.36	1.02	0.06	0.06	2.00	0.36
30	Esy Undrst	0	9	4	8	0	0	2.95	0.90	3.30	1.27	-0.35	-0.28	2.00	1.30
31	Gen Conc	1	8	5	5	2	0	2.95	1.09	3.50	1.08	-0.55	-0.51	2.00	1.50
32	No Hm Hlp *	1	4	6	6	4	0	2.62	1.13	2.65	1.34	-0.03	-0.02	3.00	-0.35
33	Stdy Skls	2	9	7	3	0	0	2.52	0.85	3.39	0.97	-0.87	-0.90	4.00	-0.61
34	Team	2	10	6	3	0	0	2.48	0.85	2.45	1.12	0.02	0.02	1.00	1.45
35	Adpts Ablty	1	7	7	4	1	0	2.85	0.96	3.09	1.06	-0.24	-0.22	1.00	2.09
36	Hmk Cpes	1	11	4	4	1	0	2.67	0.99	2.52	1.17	0.14	0.12	1.00	1.52
37	St Disply	6	14	1	0	0	0	1.76	0.53	3.36	1.02	-1.60	-1.56	3.00	0.36
38	Inclusive	6	12	2	1	0	0	1.90	0.75	2.00	1.06	-0.10	-0.09	1.00	1.00
39	Get Job	6	6	9	0	0	0	2.14	0.83	2.22	1.32	-0.07	-0.06	1.00	1.22
40	S Ask Qns	2	6	8	4	1	0	2.81	1.01	3.04	1.23	-0.23	-0.19	2.00	1.04
41	Motivates	1	6	8	5	1	0	2.95	0.95	3.61	1.13	-0.66	-0.58	2.00	1.61
42	Difficulty *	0	2	4	13	2	0	2.29	0.76	2.13	1.03	0.16	0.15	4.00	-1.87
43	Cmnts Hlp	2	8	3	8	0	0	2.81	1.05	2.96	1.20	-0.15	-0.12	2.00	0.96
44	On Time	0	0	8	6	6	1	3.90	0.83	4.09	0.88	-0.19	-0.21	4.00	0.09
45	Speed *	1	8	3	7	2	0	2.95	1.13	3.04	1.04	-0.09	-0.09	3.00	0.04
46	Pronounce	7	10	1	1	1	0	1.95	1.02	2.09	1.18	-0.14	-0.12	2.00	0.09
47	Cont Nxt Yr	2	6	7	5	1	0	2.86	1.04	2.70	1.12	0.16	0.14	2.00	0.70
48	Examples	3	10	5	2	1	0	2.43	1.00	2.70	0.75	-0.27	-0.36	2.00	0.70
49	Fin Early *	0	4	2	8	6	1	2.20	1.08	2.23	1.08	-0.03	-0.03	3.00	-0.77
50	Talk A Lot *	1	4	8	6	2	0	2.81	1.01	2.57	1.17	0.24	0.21	1.00	1.57
51	Bst Tchr	2	7	8	3	1	0	2.71	0.98	3.43	0.97	-0.72	-0.74	2.00	1.43
52	Bst Class	2	7	3	7	2	0	3.00	1.20	2.91	1.16	0.09	0.08	2.00	0.91
53	Worthwhile	7	9	4	0	1	0	2.00	0.98						

54	Felt Good	9	9	1	1	1	0	1.86	1.04
55	Thanked	1	6	8	5	1	0	2.95	0.95
56	Discussed	1	2	4	7	7	0	3.81	1.14
57	Tried Impr	1	7	8	3	2	0	2.90	1.02
58	Did Impr	2	5	7	4	3	0	3.05	1.17
59	Picked On *	0	1	2	13	5	0	1.95	0.72
60	Do For All	11	4	3	1	2	0	2.00	1.31

Sex -->

Male	11
Female	10
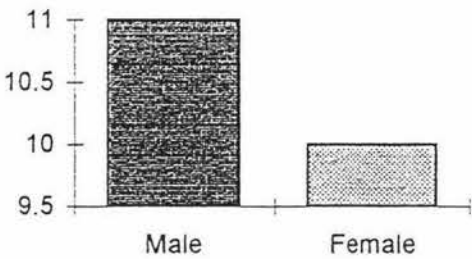
Ethnic -->

European	15
Maori	2
Pacific Is	0
Asian	3
Other	1
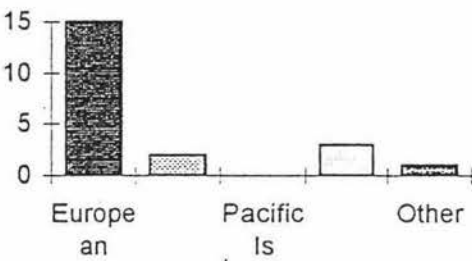
Ability -->

Top	3
Middle	18
Bottom	0
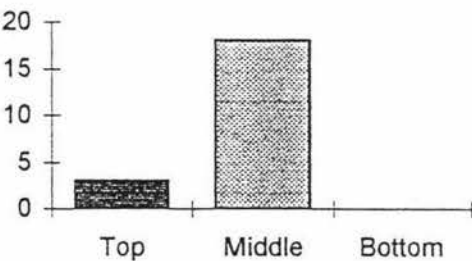
		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	2	12	10	0	0	0	2.33	0.62	2.41	0.65	-0.08	-0.12		2.41
2	Discuss.n	3	13	6	2	0	0	2.29	0.79	2.86	0.87	-0.57	-0.66	1.00	1.86
3	Handling *	0	3	4	9	8	0	2.08	1.00	2.23	0.79	-0.14	-0.18	5.00	-2.77
4	Clarity	6	11	7	0	0	0	2.04	0.73	2.86	1.10	-0.82	-0.75	2.00	0.86
5	Prompt	1	0	4	16	3	0	3.83	0.80	3.86	1.14	-0.03	-0.03	3.00	0.86
6	Happy	5	16	2	1	0	0	1.96	0.68	2.00	0.90	-0.04	-0.05	1.00	1.00
7	Control *	3	0	0	7	14	0	1.79	1.29	1.91	1.08	-0.12	-0.11	5.00	-3.09
8	Challenge	5	10	5	3	1	0	2.38	1.07	2.59	0.98	-0.22	-0.22	1.00	1.59
9	Approachable	2	9	9	3	1	0	2.67	0.94	3.82	1.15	-1.15	-1.00	2.00	1.82
10	Related	3	5	7	7	2	0	3.00	1.15	3.64	1.11	-0.64	-0.57	2.00	1.64
11	Enthuses	3	7	11	2	1	0	2.63	0.95	3.23	0.79	-0.60	-0.76	1.00	2.23
12	Humour	2	11	7	3	1	0	2.58	0.95	3.14	1.10	-0.55	-0.50	1.00	2.14
13	Impt Pts	7	9	7	0	0	0	2.00	0.78	2.45	0.78	-0.45	-0.58	1.00	1.45
14	Strict *	4	7	12	0	1	0	3.54	0.91	3.64	0.71	-0.09	-0.13	1.00	2.64
15	Disrupt. Std	1	2	8	8	5	0	3.58	1.04	3.32	1.36	0.27	0.19	2.00	1.32
16	Asmt Fair	7	11	4	2	0	0	2.04	0.89	2.14	0.69	-0.09	-0.14	1.00	1.14
17	Mat Prep	4	10	7	2	0	1	2.30	0.86	2.59	0.94	-0.29	-0.31	1.00	1.59
18	Know Wld	4	6	7	5	2	0	2.79	1.19	3.10	1.31	-0.30	-0.23	1.00	2.10
19	Chall/Intrst	0	11	7	4	2	0	2.88	0.97	3.05	0.93	-0.17	-0.18	1.00	2.05
20	Improve	2	13	6	3	0	0	2.42	0.81	2.91	1.04	-0.49	-0.47	2.00	0.91
21	Asmt Crit	3	9	10	1	0	0	2.39	0.77	3.09	0.85	-0.70	-0.83	1.00	2.09
22	Hmwk Ovld *	0	4	9	8	3	0	2.58	0.91	3.18	1.07	-0.60	-0.56	4.00	-0.82
23	Try Hard	7	11	5	1	0	0	2.00	0.82	2.36	1.26	-0.36	-0.29	1.00	1.36
24	Boys Qns *	0	2	9	3	9	1	2.17	1.05	2.32	1.06	-0.14	-0.14	5.00	-2.68
25	Ignore Tch *	0	2	5	12	5	0	2.17	0.85	2.36	1.02	-0.20	-0.19	3.00	-0.64
26	Feel Gd	1	4	12	4	3	0	3.17	0.99	3.64	0.77	-0.47	-0.61	3.00	0.64
27	Org/Expct	9	9	4	0	0	1	1.77	0.73	2.27	1.01	-0.50	-0.50	1.00	1.27
28	Too Much	0	8	6	9	0	0	3.04	0.86	3.23	1.17	-0.18	-0.16	4.00	-0.77
29	Incr Think	1	13	9	0	0	0	2.35	0.56	3.00	0.80	-0.65	-0.82	3.00	0.00
30	Esy Undrst	2	11	9	1	0	0	2.39	0.71	3.32	1.18	-0.93	-0.78	2.00	1.32
31	Gen Conc	3	11	6	2	1	0	2.43	0.97	2.82	0.98	-0.38	-0.39	1.00	1.82
32	No Hm Hlp *	0	2	1	10	10	0	1.78	0.88	2.32	1.10	-0.54	-0.49	5.00	-2.68
33	Stdy Skls	1	4	11	5	1	0	3.05	0.88	3.77	1.08	-0.73	-0.67	1.00	2.77
34	Team	2	11	8	2	0	0	2.43	0.77	3.55	0.89	-1.11	-1.25	1.00	2.55
35	Adpts Ablty	2	10	9	2	0	0	2.48	0.77	3.14	1.10	-0.66	-0.60	1.00	2.14
36	Hmk Cpes	3	15	3	2	0	0	2.17	0.76	2.32	0.82	-0.14	-0.18	2.00	0.32
37	St Disply	6	8	4	3	2	0	2.43	1.25	2.91	1.20	-0.47	-0.39	1.00	1.91
38	Inclusive	10	9	2	1	1	0	1.87	1.03	2.18	0.89	-0.31	-0.35	1.00	1.18
39	Get Job	13	6	3	1	0	0	1.65	0.87	1.91	1.08	-0.26	-0.24	1.00	0.91
40	S Ask Qns	5	11	7	0	0	0	2.09	0.72	3.00	1.17	-0.91	-0.78	1.00	2.00
41	Motivates	1	12	9	1	0	0	2.43	0.65	3.14	1.01	-0.70	-0.69	1.00	2.14
42	Difficulty *	2	3	4	8	6	0	2.43	1.25	3.05	1.26	-0.61	-0.48	3.00	0.05
43	Cmnts Hlp	4	5	8	3	3	0	2.83	1.24	3.18	0.94	-0.36	-0.38	2.00	1.18
44	On Time	1	2	4	9	7	0	3.83	1.09	3.77	1.08	0.05	0.05	4.00	-0.23
45	Speed *	2	4	9	6	1	0	3.00	1.00	2.64	0.83	0.36	0.44	3.00	-0.36
46	Pronounce	12	10	1	0	0	0	1.52	0.58	1.77	0.73	-0.25	-0.34	1.00	0.77
47	Cont Nxt Yr	3	6	8	1	5	0	2.96	1.30	3.09	1.28	-0.13	-0.11	2.00	1.09
48	Examples	3	11	7	2	0	0	2.35	0.81	2.86	1.06	-0.52	-0.49	1.00	1.86
49	Fin Early *	0	2	1	9	11	0	1.74	0.90	1.77	1.04	-0.03	-0.03	5.00	-3.23
50	Talk A Lot *	0	4	5	11	3	0	2.43	0.92	2.45	0.99	-0.02	-0.02	4.00	-1.55
51	Bst Tchr	3	7	9	2	2	0	2.70	1.08	3.45	0.94	-0.76	-0.81	3.00	0.45
52	Bst Class	1	8	7	4	3	0	3.00	1.10	3.68	1.02	-0.68	-0.67	3.00	0.68
53	Worthwhile	10	12	2	0	0	0	1.67	0.62						

54	Felt Good	11	10	3	0	0	0	1.67	0.69
55	Thanked	7	11	6	0	0	0	1.96	0.73
56	Discussed	17	6	1	0	0	0	1.33	0.55
57	Tried Impr	4	13	6	1	0	0	2.17	0.75
58	Did Impr	3	7	12	2	0	0	2.54	0.82
59	Picked On *	0	0	8	6	10	0	1.92	0.86
60	Do For All	18	5	1	0	0	0	1.29	0.54

Sex -->

| Male | 15 |
| Female | 9 |

Eth nic -->

European	15
Maori	2
Pacific Is	3
Asian	2
Other	2

Abi lity -->

Top	4
Middle	14
Bottom	5

		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	5	8	5	1	0	0	2.11	0.85	2.26	1.02	-0.16	-0.16	3.00	-0.74
2	Discuss.n	2	8	6	2	1	0	2.58	0.99	2.63	1.09	-0.05	-0.05	2.00	0.63
3	Handling *	4	1	4	7	3	0	2.79	1.36	2.68	1.17	0.11	0.09	4.00	-1.32
4	Clarity	2	11	4	1	0	0	2.22	0.71	2.95	1.28	-0.73	-0.57	2.00	0.95
5	Prompt	1	5	7	5	0	0	2.89	0.87	3.42	1.14	-0.53	-0.47	5.00	-1.58
6	Happy	7	6	3	0	1	1	1.94	1.06	2.05	1.23	-0.11	-0.09	2.00	0.05
7	Control *	0	2	6	7	4	0	2.32	0.92	2.11	1.02	0.21	0.21	4.00	-1.89
8	Challenge	1	6	7	3	1	0	2.83	0.96	2.79	0.83	0.04	0.05	3.00	-0.21
9	Approachable	1	3	6	7	2	0	3.32	1.03	3.42	0.99	-0.11	-0.11	4.00	-0.58
10	Related	2	4	9	2	1	1	2.78	0.97	3.26	1.16	-0.49	-0.42	2.00	1.26
11	Enthuses	4	10	4	1	0	0	2.11	0.79	2.32	1.17	-0.21	-0.18	2.00	0.32
12	Humour	6	5	5	2	1	0	2.32	1.17	2.47	1.39	-0.16	-0.11	2.00	0.47
13	Impt Pts	2	10	6	0	1	0	2.37	0.87	2.53	0.99	-0.16	-0.16	3.00	-0.47
14	Strict *	2	5	7	4	0	1	3.28	0.93	3.11	0.91	0.17	0.19	4.00	-0.89
15	Disrupt. Std	7	4	2	4	2	0	2.47	1.43	2.53	1.27	-0.05	-0.04	2.00	0.53
16	Asmt Fair	1	8	6	1	1	1	2.59	0.91	2.42	1.09	0.17	0.15	2.00	0.42
17	Mat Prep	3	8	3	4	0	0	2.44	1.01	2.84	1.09	-0.40	-0.37	4.00	-1.16
18	Know Wld	2	13	2	2	0	0	2.21	0.77	2.39	1.16	-0.18	-0.15	2.00	0.39
19	Chall/Intrst	2	6	8	3	0	0	2.63	0.87	2.74	0.85	-0.11	-0.12	3.00	-0.26
20	Improve	2	9	5	1	1	0	2.44	0.96	2.74	1.07	-0.29	-0.27	4.00	-1.26
21	Asmt Crit	4	6	5	2	0	1	2.29	0.96	2.42	0.82	-0.13	-0.16	5.00	-2.58
22	Hmwk Ovld *	3	3	4	6	3	0	2.84	1.31	2.79	1.24	0.05	0.04	5.00	-2.21
23	Try Hard	9	5	3	1	0	0	1.78	0.92	1.89	0.97	-0.12	-0.12	3.00	-1.11
24	Boys Qns *	1	3	7	5	1	1	2.88	1.03	2.32	1.08	0.57	0.53	3.00	-0.68
25	Ignore Tch *	2	7	4	5	1	0	3.21	1.10	2.63	1.04	0.58	0.56	2.00	0.63
26	Feel Gd	3	6	7	3	0	0	2.53	0.94	3.11	0.97	-0.58	-0.60	3.00	0.11
27	Org/Expct	7	5	2	2	1	1	2.12	1.23	2.33	1.20	-0.22	-0.18	2.00	0.33
28	Too Much	3	5	6	3	1	0	2.67	1.08	2.89	1.07	-0.23	-0.21	4.00	-1.11
29	Incr Think	5	3	8	2	0	0	2.39	0.98	2.47	1.09	-0.08	-0.08	3.00	-0.53
30	Esy Undrst	5	5	5	1	2	0	2.44	1.23	2.63	1.13	-0.19	-0.17	3.00	-0.37
31	Gen Conc	4	7	6	1	0	0	2.22	0.85	2.61	0.89	-0.39	-0.44	2.00	0.61
32	No Hm Hlp *	2	4	5	2	4	1	2.88	1.29	2.84	1.09	0.04	0.04	4.00	-1.16
33	Stdy Skls	1	6	7	4	0	0	2.78	0.83	2.33	0.88	0.44	0.50	2.00	0.33
34	Team	5	9	3	0	1	0	2.06	0.97	2.84	1.35	-0.79	-0.58	2.00	0.84
35	Adpts Ablty	1	7	6	3	1	0	2.78	0.96	2.95	1.23	-0.17	-0.14	4.00	-1.05
36	Hmk Cpes	4	8	5	0	0	0	2.06	0.75	2.11	0.99	-0.05	-0.05	2.00	0.11
37	St Disply	6	6	3	1	1	0	2.12	1.10	1.68	0.73	0.43	0.59	4.00	-2.32
38	Inclusive	9	5	3	0	0	0	1.65	0.76	1.74	0.85	-0.09	-0.11	3.00	-1.26
39	Get Job	3	5	5	4	1	0	2.72	1.12	3.11	1.25	-0.38	-0.31	4.00	-0.89
40	S Ask Qns	6	6	3	2	1	0	2.22	1.15	2.74	1.02	-0.51	-0.51	3.00	-0.26
41	Motivates	4	8	3	1	2	0	2.39	1.18	2.53	0.94	-0.14	-0.15	3.00	-0.47
42	Difficulty *	1	2	4	8	3	0	2.44	1.04	2.11	0.99	0.33	0.34	2.00	0.11
43	Cmnts Hlp	5	5	4	1	1	1	2.25	1.13	3.00	1.21	-0.75	-0.62	4.00	-1.00
44	On Time	4	2	8	4	0	0	2.67	1.03	2.84	1.14	-0.18	-0.15	4.00	-1.16
45	Speed *	1	5	4	5	3	0	2.78	1.18	2.84	1.39	-0.06	-0.05	2.00	0.84
46	Pronounce	7	4	1	5	0	1	2.24	1.24	1.88	1.18	0.35	0.30	2.00	-0.12
47	Cont Nxt Yr	2	2	7	6	1	0	3.11	1.10	3.16	1.31	-0.05	-0.04	3.00	0.16
48	Examples	1	5	7	4	0	0	2.82	0.83	2.37	0.98	0.46	0.46	3.00	-0.63
49	Fin Early *	1	4	4	4	3	1	2.75	1.20	2.42	1.46	0.33	0.22	2.00	0.42
50	Talk A Lot *	3	1	4	9	1	0	2.78	1.15	3.16	1.27	-0.38	-0.30	2.00	1.16
51	Bst Tchr	3	6	4	3	2	0	2.72	1.22	3.11	1.12	-0.38	-0.34	4.00	-0.89
52	Bst Class	3	4	3	4	4	0	3.11	1.37	3.21	1.06	-0.10	-0.09	4.00	-0.79
53	Worthwhile	8	7	4	0	0	0	1.79	0.77						

54	Felt Good	12	5	2	0	0	0	1.47	0.68
55	Thanked	5	5	4	3	2	0	2.58	1.31
56	Discussed	3	7	6	2	1	0	2.53	1.04
57	Tried Impr	4	9	5	0	1	0	2.21	0.95
58	Did Impr	4	6	6	2	1	0	2.47	1.09
59	Picked On *	1	3	6	5	4	0	2.58	1.14
60	Do For All	15	4	0	0	0	0	1.21	0.41

Sex -->

| Male | 11 |
| Female | 8 |

Eth nic -->

European	2
Maori	4
Pacific Is	9
Asian	3
Other	1

Abi lity -->
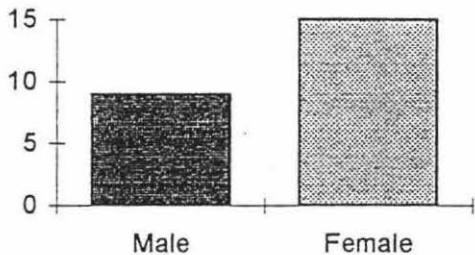
Top	2
Middle	10
Bottom	5
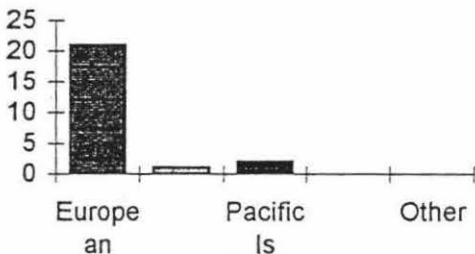
		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	3	12	6	1	2	0	2.46	1.04	2.27	0.75	0.19	0.25	2.00	0.27
2	Discuss.n	2	9	10	2	1	0	2.63	0.90	2.50	0.84	0.13	0.15	2.00	0.50
3	Handling *	1	2	3	14	4	0	2.25	0.97	3.00	1.09	-0.75	-0.69	5.00	-2.00
4	Clarity	4	9	6	3	2	0	2.58	1.15	2.77	1.17	-0.19	-0.16	1.00	1.77
5	Prompt	0	4	9	9	2	0	3.38	0.86	4.05	1.02	-0.67	-0.66	2.00	2.05
6	Happy	11	11	2	0	0	0	1.63	0.63	1.59	0.65	0.03	0.05	2.00	-0.41
7	Control *	1	0	4	8	11	0	1.83	0.99	2.14	1.10	-0.30	-0.28	5.00	-2.86
8	Challenge	1	3	11	6	3	0	3.29	0.98	2.86	1.01	0.43	0.42	2.00	0.86
9	Approachable	3	5	5	9	2	0	3.08	1.19	4.05	1.07	-0.96	-0.90	2.00	2.05
10	Related	0	5	8	8	3	0	3.38	0.95	3.91	1.00	-0.53	-0.54	3.00	0.91
11	Enthuses	2	15	5	1	1	0	2.33	0.85	3.23	1.04	-0.89	-0.86	2.00	1.23
12	Humour	2	4	9	5	4	0	3.21	1.15	3.73	1.09	-0.52	-0.47	2.00	1.73
13	Impt Pts	7	10	6	0	1	0	2.08	0.95	2.23	1.00	-0.14	-0.14	1.00	1.23
14	Strict *	3	4	8	8	1	0	3.00	1.08	2.82	0.89	0.18	0.21	2.00	0.82
15	Disrupt. Std	1	6	4	7	6	0	3.46	1.22	3.27	1.05	0.19	0.18	3.00	0.27
16	Asmt Fair	9	10	3	1	1	0	1.96	1.02	2.18	0.83	-0.22	-0.27	1.00	1.18
17	Mat Prep	5	9	7	1	2	0	2.42	1.11	2.52	0.91	-0.11	-0.12	1.00	1.52
18	Know Wld	1	7	11	4	1	0	2.88	0.88	3.27	0.86	-0.40	-0.46	1.00	2.27
19	Chall/Intrst	2	7	9	3	3	0	2.92	1.11	3.05	0.82	-0.13	-0.16	1.00	2.05
20	Improve	3	10	5	4	2	0	2.67	1.14	3.09	1.08	-0.42	-0.39	1.00	2.09
21	Asmt Crit	7	13	1	3	0	0	2.00	0.91	2.10	0.87	-0.10	-0.11	1.00	1.10
22	Hmwk Ovld *	0	0	7	10	6	1	2.04	0.75	2.41	1.03	-0.37	-0.36	3.00	-0.59
23	Try Hard	6	11	5	1	1	0	2.17	0.99	2.14	1.01	0.03	0.03	1.00	1.14
24	Boys Qns *	0	1	3	8	12	0	1.71	0.84	2.23	0.95	-0.52	-0.55	3.00	-0.77
25	Ignore Tch *	2	1	4	11	6	0	2.25	1.13	2.73	1.17	-0.48	-0.41	5.00	-2.27
26	Feel Gd	0	6	11	2	5	0	3.25	1.05	3.55	1.03	-0.30	-0.29	2.00	1.55
27	Org/Expct	8	8	5	2	1	0	2.17	1.11	2.23	1.13	-0.06	-0.05	1.00	1.23
28	Too Much	1	2	8	9	4	0	3.54	1.00	3.23	0.90	0.31	0.35	2.00	1.23
29	Incr Think	0	12	6	6	0	0	2.75	0.83	3.23	1.04	-0.48	-0.46	1.00	2.23
30	Esy Undrst	1	10	4	7	2	0	2.96	1.10	3.32	1.33	-0.36	-0.27	2.00	1.32
31	Gen Conc	4	13	3	1	2	0	2.30	1.08	3.05	1.15	-0.74	-0.65	1.00	2.05
32	No Hm Hlp *	1	2	6	4	10	1	2.13	1.19	2.45	1.23	-0.32	-0.26	5.00	-2.55
33	Stdy Skls	0	4	10	8	2	0	3.33	0.85	3.41	0.98	-0.08	-0.08	1.00	2.41
34	Team	0	5	10	8	1	0	3.21	0.82	3.14	0.97	0.07	0.07	2.00	1.14
35	Adpts Ablty	2	8	8	4	2	0	2.83	1.07	3.41	0.89	-0.58	-0.65	2.00	1.41
36	Hmk Cpes	8	9	4	1	1	1	2.04	1.04	1.95	1.07	0.09	0.08	2.00	-0.05
37	St Disply	12	9	3	0	0	0	1.63	0.70	1.41	0.49	0.22	0.44	1.00	0.41
38	Inclusive	11	9	2	1	1	0	1.83	1.03	1.86	0.69	-0.03	-0.04	1.00	0.86
39	Get Job	14	9	0	0	1	0	1.54	0.87	1.77	1.04	-0.23	-0.22	1.00	0.77
40	S Ask Qns	3	10	9	0	2	0	2.50	1.00	2.91	1.12	-0.41	-0.36	2.00	0.91
41	Motivates	3	7	7	5	2	0	2.83	1.14	3.27	1.21	-0.44	-0.36	2.00	1.27
42	Difficulty *	2	5	6	5	6	0	2.67	1.28	2.36	1.19	0.30	0.25	2.00	0.36
43	Cmnts Hlp	4	7	6	4	3	0	2.79	1.26	3.00	1.13	-0.21	-0.18	1.00	2.00
44	On Time	1	5	5	8	5	0	3.46	1.15	4.18	1.03	-0.72	-0.70	4.00	0.18
45	Speed *	2	4	7	8	3	0	2.75	1.13	2.77	1.13	-0.02	-0.02	4.00	-1.23
46	Pronounce	12	8	3	1	0	0	1.71	0.84	1.68	0.76	0.03	0.03	2.00	-0.32
47	Cont Nxt Yr	9	3	4	2	5	1	2.61	1.58	3.00	1.38	-0.39	-0.28	1.00	2.00
48	Examples	8	5	9	1	1	0	2.25	1.09	2.45	0.66	-0.20	-0.31	1.00	1.45
49	Fin Early *	3	10	5	3	3	0	3.29	1.21	2.86	1.22	0.43	0.35	4.00	-1.14
50	Talk A Lot *	2	3	6	10	3	0	2.63	1.11	2.45	1.08	0.17	0.16	4.00	-1.55
51	Bst Tchr	2	6	9	4	3	0	3.00	1.12	3.77	1.17	-0.77	-0.66	2.00	1.77
52	Bst Class	4	5	5	5	5	0	3.08	1.38	3.64	1.33	-0.55	-0.41	2.00	1.64
53	Worthwhile	13	9	2	0	0	0	1.54	0.64						

54	Felt Good	10	13	1	0	0	0	1.63	0.56
55	Thanked	10	11	2	0	1	0	1.79	0.91
56	Discussed	18	4	2	0	0	0	1.33	0.62
57	Tried Impr	8	13	1	1	1	0	1.92	0.95
58	Did Impr	5	10	6	1	2	0	2.38	1.11
59	Picked On *	1	0	2	6	15	0	1.58	0.95
60	Do For All	12	4	5	2	1	0	2.00	1.19
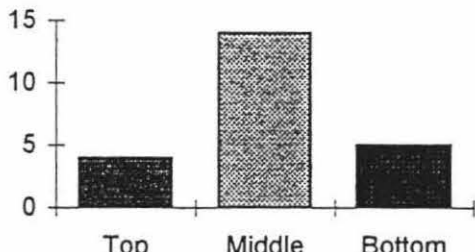
Sex -->

Male	9
Female	15

Eth nic -->

European	21
Maori	1
Pacific Is	2
Asian	0
Other	0

Abi lity -->

Top	4
Middle	14
Bottom	5

		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	4	13	4	0	0	0	2.00	0.62	2.24	0.55	-0.24	-0.43	2.00	0.24
2	Discuss.n	6	12	3	0	0	0	1.86	0.64	1.76	0.73	0.09	0.13	1.00	0.76
3	Handling *	0	0	1	8	11	1	1.50	0.59	2.38	1.36	-0.88	-0.64	5.00	-2.63
4	Clarity	10	9	2	0	0	0	1.62	0.65	1.76	0.81	-0.15	-0.18	1.00	0.76
5	Prompt	0	5	2	1	1	12	2.78	1.03	3.00	0.82	-0.22	-0.27	2.00	1.00
6	Happy	2	15	2	0	1	0	2.15	0.79	1.94	1.00	0.21	0.21	2.00	-0.06
7	Control *	0	0	1	12	8	0	1.67	0.56	2.12	1.28	-0.45	-0.35	5.00	-2.88
8	Challenge	0	4	9	4	1	3	3.11	0.81	2.88	0.86	0.24	0.28	2.00	0.88
9	Approachable	2	8	8	2	1	0	2.62	0.95	2.94	1.06	-0.32	-0.31	6.00	-3.06
10	Related	0	7	12	1	0	1	2.70	0.56	3.25	0.90	-0.55	-0.61	4.00	-0.75
11	Enthuses	9	9	1	0	2	0	1.90	1.15	1.53	0.70	0.38	0.54	2.00	-0.47
12	Humour	8	12	1	0	0	0	1.67	0.56	1.71	0.57	-0.04	-0.07	2.00	-0.29
13	Impt Pts	6	8	6	1	0	0	2.10	0.87	2.06	0.87	0.04	0.04	1.00	1.06
14	Strict *	0	0	9	4	8	0	2.05	0.90	2.82	0.62	-0.78	-1.26	3.00	-0.18
15	Disrupt. Std	0	4	9	7	1	0	3.24	0.81	3.38	1.05	-0.14	-0.13	2.00	1.38
16	Asmt Fair	1	0	3	0	0	17	2.50	0.87	2.00	0.82	0.50	0.61	1.00	1.00
17	Mat Prep	2	10	5	1	0	2	2.28	0.73	2.08	0.47	0.20	0.42	1.00	1.08
18	Know Wld	2	4	11	1	1	2	2.74	0.91	2.81	0.95	-0.08	-0.08	2.00	0.81
19	Chall/Intrst	7	12	2	0	0	0	1.76	0.61	1.71	0.75	0.06	0.07	2.00	-0.29
20	Improve	11	7	2	1	0	0	1.67	0.84	1.50	0.61	0.17	0.27	2.00	-0.50
21	Asmt Crit	6	4	6	1	1	3	2.28	1.15	1.40	0.71	0.88	1.23	1.00	0.40
22	Hmwk Ovld *	0	0	0	1	1	19	1.50	0.50	2.00	1.00	-0.50	-0.50	6.00	-4.00
23	Try Hard	9	10	2	0	0	0	1.67	0.64	1.88	0.96	-0.22	-0.22	2.00	-0.12
24	Boys Qns *	0	0	2	5	14	0	1.43	0.66	1.65	0.76	-0.22	-0.29	4.00	-2.35
25	Ignore Tch *	0	1	10	4	6	0	2.29	0.93	2.25	0.97	0.04	0.04	4.00	-1.75
26	Feel Gd	7	10	2	2	0	0	1.95	0.90	2.12	0.83	-0.17	-0.20	2.00	0.12
27	Org/Expct	5	14	2	0	0	0	1.86	0.56	1.69	0.68	0.17	0.25	1.00	0.69
28	Too Much	0	1	4	10	5	0	3.95	0.80	3.53	1.19	0.42	0.35	5.00	-1.47
29	Incr Think	3	11	5	1	0	0	2.20	0.75	2.24	0.94	-0.04	-0.04	2.00	0.24
30	Esy Undrst	2	14	3	0	0	1	2.05	0.51	2.35	0.68	-0.30	-0.44	2.00	0.35
31	Gen Conc	7	9	1	2	0	1	1.89	0.91	2.50	0.87	-0.61	-0.70	1.00	1.50
32	No Hm Hlp *	0	1	1	2	1	14	2.40	1.02	3.00	1.67	-0.60	-0.36	5.00	-2.00
33	Stdy Skls	4	2	4	0	0	9	2.00	0.89	2.44	1.17	-0.44	-0.38	1.00	1.44
34	Team	12	7	0	0	0	0	1.37	0.48	1.18	0.38	0.19	0.50	1.00	0.18
35	Adpts Ablty	2	8	8	1	0	0	2.42	0.75	2.50	0.87	-0.08	-0.09	1.00	1.50
36	Hmk Cpes	2	0	0	0	0	17	1.00	0.00	2.00	0.00	-1.00	#####	6.00	-4.00
37	St Disply	1	1	5	3	1	8	3.18	1.03	3.83	1.34	-0.65	-0.48	6.00	-2.17
38	Inclusive	4	10	4	0	1	0	2.16	0.93	1.47	0.61	0.69	1.13	2.00	-0.53
39	Get Job	1	1	13	4	0	0	3.05	0.69	2.81	0.95	0.24	0.25	1.00	1.81
40	S Ask Qns	8	6	4	1	0	0	1.89	0.91	1.88	0.58	0.01	0.02	1.00	0.88
41	Motivates	6	10	2	1	0	0	1.89	0.79	2.00	0.84	-0.11	-0.13	2.00	0.00
42	Difficulty *	0	0	3	5	10	0	1.61	0.76	1.71	1.07	-0.09	-0.09	3.00	-1.29
43	Cmnts Hlp	1	4	2	0	0	12	2.14	0.64	2.80	1.17	-0.66	-0.56	2.00	0.80
44	On Time	6	10	2	1	0	0	1.89	0.79	1.88	0.83	0.01	0.01	1.00	0.88
45	Speed *	1	0	1	7	9	0	1.72	0.99	2.00	0.73	-0.28	-0.38	5.00	-3.00
46	Pronounce	9	7	2	1	0	0	1.74	0.85	1.76	1.00	-0.03	-0.03	1.00	0.76
47	Cont Nxt Yr	7	0	10	1	0	1	2.28	1.04	1.88	1.05	0.40	0.38	2.00	-0.13
48	Examples	5	7	5	1	0	1	2.11	0.87	2.00	0.91	0.11	0.12	2.00	0.00
49	Fin Early *	0	0	4	9	5	1	1.94	0.70	2.53	1.09	-0.58	-0.54	5.00	-2.47
50	Talk A Lot *	0	2	6	7	4	0	2.32	0.92	2.29	0.89	0.02	0.02	5.00	-2.71
51	Bst Tchr	13	4	2	0	0	0	1.42	0.67	1.94	0.90	-0.52	-0.57	6.00	-4.06
52	Bst Class	11	5	3	0	0	0	1.58	0.75	1.81	0.73	-0.23	-0.32	2.00	-0.19
53	Worthwhile	3	13	5	0	0	0	2.10	0.61						

54	Felt Good	7	11	1	1	0	0	1.80	0.75					
55	Thanked	5	15	0	0	1	0	1.90	0.81					
56	Discussed	14	5	2	0	0	0	1.43	0.66					
57	Tried Impr	11	8	1	0	0	0	1.50	0.59					
58	Did Impr	7	10	3	0	1	0	1.95	0.95					
59	Picked On *	0	1	4	5	11	0	1.76	0.92					
60	Do For All	8	7	4	1	1	0	2.05	1.09					

Sex -->

| Male | 6 |
| Female | 15 |

Eth nic -->

European	20
Maori	1
Pacific Is	0
Asian	0
Other	0

Abi lity -->

Top	3
Middle	17
Bottom	0

		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	4	7	2	1	0	0	2.00	0.85	2.20	0.75	-0.20	-0.27	2.00	0.20
2	Discuss.n	1	4	5	3	1	0	2.93	1.03	3.50	0.82	-0.57	-0.69	2.00	1.50
3	Handling *	3	0	2	5	4	0	2.50	1.45	2.47	0.81	0.03	0.04	4.00	-1.53
4	Clarity	1	6	5	0	1	0	2.54	0.93	2.80	0.98	-0.26	-0.27	2.00	0.80
5	Prompt	4	5	4	1	0	0	2.14	0.91	2.79	1.21	-0.64	-0.53	2.00	0.79
6	Happy	5	5	4	0	0	0	1.93	0.80	1.80	0.75	0.13	0.17	1.00	0.80
7	Control *	2	0	6	4	2	0	2.71	1.16	2.93	1.06	-0.22	-0.21	5.00	-2.07
8	Challenge	0	2	8	4	0	0	3.14	0.64	3.62	0.92	-0.47	-0.51	6.00	-2.38
9	Approachable	1	5	3	3	2	0	3.00	1.20	3.00	1.10	0.00	0.00	2.00	1.00
10	Related	0	7	4	2	1	0	2.79	0.94	2.73	1.18	0.05	0.04	1.00	1.73
11	Enthuses	3	3	5	1	2	0	2.71	1.28	2.60	0.80	0.11	0.14	1.00	1.60
12	Humour	3	5	3	2	1	0	2.50	1.18	2.60	0.88	-0.10	-0.11	2.00	0.60
13	Impt Pts	1	6	5	2	0	0	2.57	0.82	2.53	0.81	0.04	0.05	2.00	0.53
14	Strict *	2	1	7	4	0	0	3.07	0.96	2.29	0.88	0.79	0.89	1.00	1.29
15	Disrupt. Std	1	1	4	6	2	0	3.50	1.05	3.80	1.28	-0.30	-0.24	5.00	-1.20
16	Asmt Fair	6	5	3	0	0	0	1.79	0.77	2.20	0.83	-0.41	-0.50	1.00	1.20
17	Mat Prep	8	5	1	0	0	0	1.50	0.63	1.80	0.83	-0.30	-0.36	1.00	0.80
18	Know Wld	6	4	4	0	0	0	1.86	0.83	2.47	1.31	-0.61	-0.47	1.00	1.47
19	Chall/Intrst	4	3	6	0	0	0	2.15	0.86	2.33	0.94	-0.18	-0.19	1.00	1.33
20	Improve	2	8	3	0	1	0	2.29	0.96	2.67	1.14	-0.38	-0.34	1.00	1.67
21	Asmt Crit	1	4	5	3	0	0	2.77	0.89	2.80	0.91	-0.03	-0.03	1.00	1.80
22	Hmwk Ovld *	0	0	2	6	5	0	1.77	0.70	1.64	0.72	0.13	0.18	5.00	-3.36
23	Try Hard	5	5	2	1	1	0	2.14	1.19	1.60	0.71	0.54	0.76	1.00	0.60
24	Boys Qns *	0	0	0	0	0	13	#####	#####	#####	#####	#####	#####	6.00	#####
25	Ignore Tch *	2	2	4	5	1	0	2.93	1.16	3.00	1.10	-0.07	-0.07	2.00	1.00
26	Feel Gd	2	5	7	0	0	0	2.36	0.72	2.67	1.07	-0.31	-0.29	2.00	0.67
27	Org/Expct	4	6	4	0	0	0	2.00	0.76	2.29	1.16	-0.29	-0.25	1.00	1.29
28	Too Much	2	1	2	8	1	0	3.36	1.17	3.33	0.94	0.02	0.03		3.33
29	Incr Think	2	7	5	0	0	0	2.21	0.67	2.87	1.26	-0.65	-0.52		2.87
30	Esy Undrst	0	8	4	2	0	0	2.57	0.73	3.00	1.32	-0.43	-0.33		3.00
31	Gen Conc	2	6	6	0	0	0	2.29	0.70	2.73	1.00	-0.45	-0.45		2.73
32	No Hm Hlp *	2	1	2	3	6	0	2.29	1.44	2.80	1.42	-0.51	-0.36		2.80
33	Stdy Skls	2	1	5	2	3	0	3.23	1.31	2.93	1.00	0.30	0.30		2.93
34	Team	0	1	5	3	4	1	3.77	0.97	3.47	1.36	0.30	0.22		3.47
35	Adpts Ablty	1	5	6	1	1	0	2.71	0.96	2.73	1.24	-0.02	-0.02		2.73
36	Hmk Cpes	5	4	4	0	0	1	1.92	0.83	2.62	1.08	-0.69	-0.64		2.62
37	St Disply	0	0	5	0	8	1	4.23	0.97	3.75	1.23	0.48	0.39		3.75
38	Inclusive	6	3	2	1	1	1	2.08	1.27	2.07	1.16	0.01	0.00		2.07
39	Get Job	7	4	3	0	0	0	1.71	0.80	1.47	0.72	0.25	0.34		1.47
40	S Ask Qns	1	6	5	1	1	0	2.64	0.97	2.80	1.38	-0.16	-0.11		2.80
41	Motivates	2	4	4	3	1	0	2.79	1.15	2.57	0.73	0.21	0.29		2.57
42	Difficulty *	1	0	2	5	6	0	1.93	1.10	2.27	1.53	-0.34	-0.22		2.27
43	Cmnts Hlp	3	3	5	3	0	0	2.57	1.05	2.57	0.98	0.00	0.00		2.57
44	On Time	0	3	3	4	4	0	3.64	1.11	3.71	0.88	-0.07	-0.08		3.71
45	Speed *	0	0	2	3	8	0	1.54	0.75	2.07	1.12	-0.53	-0.47		2.07
46	Pronounce	4	6	3	1	0	0	2.07	0.88	2.07	1.10	0.00	0.00		2.07
47	Cont Nxt Yr	2	2	7	3	0	0	2.79	0.94	2.50	0.98	0.29	0.29		2.50
48	Examples	1	5	5	3	0	0	2.71	0.88	3.13	0.88	-0.42	-0.47		3.13
49	Fin Early *	0	1	2	2	9	0	1.64	0.97	2.20	1.22	-0.56	-0.46		2.20
50	Talk A Lot *	0	1	2	4	7	0	1.79	0.94	1.60	0.71	0.19	0.26		1.60
51	Bst Tchr	0	5	5	2	2	0	3.07	1.03	3.00	1.03	0.07	0.07		3.00
52	Bst Class	3	5	3	3	0	0	2.43	1.05	2.71	0.80	-0.29	-0.36		2.71
53	Worthwhile	7	4	2	0	1	0	1.86	1.12						

54	Felt Good	6	6	0	1	1	0	1.93	1.16
55	Thanked	1	3	5	3	2	0	3.14	1.12
56	Discussed	0	0	5	3	6	0	4.07	0.88
57	Tried Impr	3	4	4	2	1	0	2.57	1.18
58	Did Impr	2	5	5	1	1	0	2.57	1.05
59	Picked On *	2	1	0	4	7	0	2.07	1.44
60	Do For All	3	4	6	0	1	0	2.43	1.05

Sex -->

Male	0
Female	14
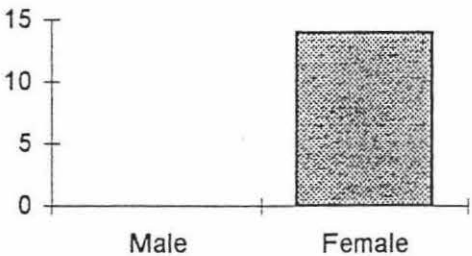
Ethnic -->

European	5
Maori	1
Pacific Is	4
Asian	2
Other	2
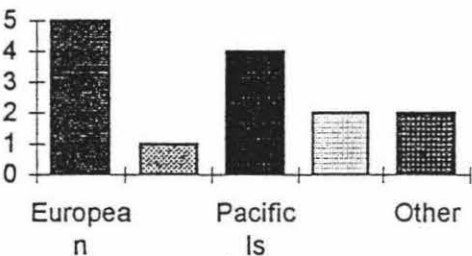
Ability -->

Top	4
Middle	7
Bottom	3
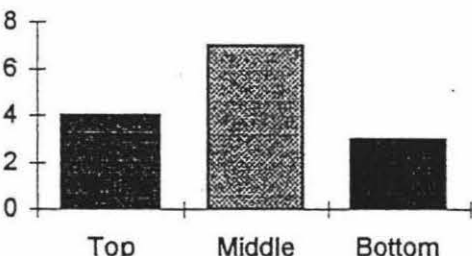
		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	2	8	5	0	1	0	2.38	0.93	2.00	0.82	0.38	0.46	2.00	0.00
2	Discuss.n	6	6	3	1	0	0	1.94	0.90	1.87	0.88	0.07	0.08	1.00	0.87
3	Handling *	0	4	4	6	2	0	2.63	0.99	2.13	0.88	0.49	0.56	6.00	-3.87
4	Clarity	1	6	7	2	0	0	2.63	0.78	1.87	0.62	0.76	1.23	5.00	-3.13
5	Prompt	9	6	1	0	0	0	1.50	0.61	2.13	1.02	-0.63	-0.62	5.00	-2.87
6	Happy	5	7	2	1	1	0	2.13	1.11	2.13	1.09	-0.01	-0.01	1.00	1.13
7	Control *	0	4	7	5	0	0	2.94	0.75	2.27	0.68	0.67	0.99	4.00	-1.73
8	Challenge	1	11	2	2	0	0	2.31	0.77	2.60	0.88	-0.29	-0.33	2.00	0.60
9	Approachable	0	5	4	5	2	0	3.25	1.03	2.43	1.12	0.82	0.74	1.00	1.43
10	Related	0	6	6	3	1	0	2.94	0.90	3.33	0.94	-0.40	-0.42	3.00	0.33
11	Enthuses	3	7	5	1	0	0	2.25	0.83	2.47	0.72	-0.22	-0.30	2.00	0.47
12	Humour	4	7	2	3	0	0	2.25	1.03	2.60	0.95	-0.35	-0.37	1.00	1.60
13	Impt Pts	5	7	3	1	0	0	2.00	0.87	2.33	0.70	-0.33	-0.48	1.00	1.33
14	Strict *	1	2	8	3	2	0	2.81	1.01	3.07	0.93	-0.25	-0.27	1.00	2.07
15	Disrupt. Std	0	4	4	6	2	0	3.38	0.99	3.00	1.03	0.38	0.36	1.00	2.00
16	Asmt Fair	2	8	5	1	0	0	2.31	0.77	2.07	1.00	0.25	0.25	2.00	0.07
17	Mat Prep	3	7	2	4	0	0	2.44	1.06	2.21	0.67	0.22	0.33	2.00	0.21
18	Know Wld	2	7	5	2	0	0	2.44	0.86	2.67	0.70	-0.23	-0.33	1.00	1.67
19	Chall/Intrst	6	4	3	3	0	0	2.19	1.13	2.07	0.85	0.12	0.14	2.00	0.07
20	Improve	3	5	3	5	0	0	2.63	1.11	2.13	1.15	0.49	0.43	1.00	1.13
21	Asmt Crit	3	11	1	1	0	0	2.00	0.71	2.64	0.89	-0.64	-0.72	2.00	0.64
22	Hmwk Ovld *	0	1	10	3	2	0	2.63	0.78	2.53	1.09	0.09	0.08	6.00	-3.47
23	Try Hard	5	7	3	1	0	0	2.00	0.87	1.73	0.57	0.27	0.46	6.00	-4.27
24	Boys Qns *	0	0	0	0	0	16	#####	#####	#####	#####	#####	#####	6.00	#####
25	Ignore Tch *	0	5	6	5	0	0	3.00	0.79	2.40	0.95	0.60	0.63	1.00	1.40
26	Feel Gd	5	2	6	3	0	0	2.44	1.12	2.27	0.77	0.17	0.22	2.00	0.27
27	Org/Expct	2	7	3	4	0	0	2.56	1.00	2.93	1.12	-0.37	-0.33	6.00	-3.07
28	Too Much	0	3	4	7	2	0	3.50	0.94	3.87	0.50	-0.37	-0.73	6.00	-2.13
29	Incr Think	2	9	4	0	1	0	2.31	0.92	2.60	1.08	-0.29	-0.27	2.00	0.60
30	Esy Undrst	3	3	5	5	0	0	2.75	1.09	2.53	1.02	0.22	0.21	2.00	0.53
31	Gen Conc	1	6	6	3	0	0	2.69	0.85	2.60	0.80	0.09	0.11	1.00	1.60
32	No Hm Hlp *	1	2	6	6	1	0	2.75	0.97	1.73	0.93	1.02	1.09	6.00	-4.27
33	Stdy Skls	1	6	5	3	1	0	2.81	1.01	2.67	1.07	0.15	0.14	2.00	0.67
34	Team	1	13	2	0	0	0	2.06	0.43	2.40	0.88	-0.34	-0.38	1.00	1.40
35	Adpts Ablty	2	10	3	1	0	0	2.19	0.73	2.14	0.91	0.04	0.05	1.00	1.14
36	Hmk Cpes	6	7	3	0	0	0	1.81	0.73	1.60	0.49	0.21	0.43	1.00	0.60
37	St Disply	0	3	6	5	2	0	3.38	0.93	2.53	0.88	0.84	0.95	4.00	-1.47
38	Inclusive	9	4	2	1	0	0	1.69	0.92	1.93	0.77	-0.25	-0.32	1.00	0.93
39	Get Job	5	7	3	1	0	0	2.00	0.87	2.20	1.05	-0.20	-0.19	1.00	1.20
40	S Ask Qns	5	6	4	1	0	0	2.06	0.90	2.13	0.72	-0.07	-0.10	1.00	1.13
41	Motivates	3	4	4	4	1	0	2.75	1.20	2.47	1.09	0.28	0.26	2.00	0.47
42	Difficulty *	0	3	5	2	6	0	2.31	1.16	1.60	0.61	0.71	1.17	1.00	0.60
43	Cmnts Hlp	7	6	1	1	1	0	1.94	1.14	2.13	0.88	-0.20	-0.22	1.00	1.13
44	On Time	4	7	4	1	0	0	2.13	0.86	1.93	0.77	0.19	0.25	2.00	-0.07
45	Speed *	0	3	5	7	1	0	2.63	0.86	2.07	0.93	0.56	0.60	6.00	-3.93
46	Pronounce	10	5	0	1	0	0	1.50	0.79	1.67	0.60	-0.17	-0.28	1.00	0.67
47	Cont Nxt Yr	5	5	4	1	1	0	2.25	1.15	1.86	0.74	0.39	0.53	6.00	-4.14
48	Examples	1	12	3	0	0	0	2.13	0.48	2.13	0.81	-0.01	-0.01	2.00	0.13
49	Fin Early *	0	2	5	6	3	0	2.38	0.93	2.14	1.06	0.23	0.22	2.00	0.14
50	Talk A Lot *	0	5	2	5	4	0	2.50	1.17	2.07	0.77	0.43	0.56	2.00	0.07
51	Bst Tchr	3	4	5	2	2	0	2.75	1.25	1.93	0.77	0.82	1.06		1.93
52	Bst Class	3	5	3	3	2	0	2.75	1.30	2.00	0.89	0.75	0.84		2.00
53	Worthwhile	5	5	5	0	1	0	2.19	1.07						

54	Felt Good	6	4	4	2	0	0	2.13	1.05						
55	Thanked	2	9	3	1	1	0	2.38	0.99						
56	Discussed	2	6	4	3	1	0	2.69	1.10						
57	Tried Impr	5	8	2	0	1	0	2.00	1.00						
58	Did Impr	4	2	6	3	1	0	2.69	1.21						
59	Picked On *	3	1	2	4	6	0	2.44	1.50						
60	Do For All	7	2	4	0	3	0	2.38	1.49						

Sex -->

Male	0
Female	16
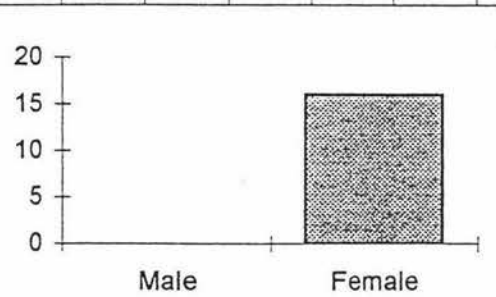
Eth nic -->

European	9
Maori	2
Pacific Is	5
Asian	0
Other	0
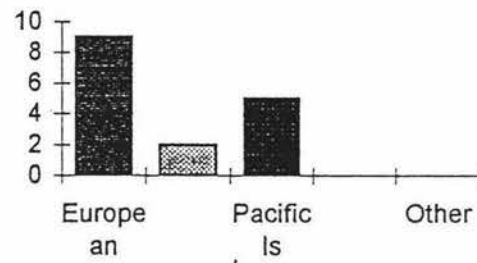
Abi lity -->

Top	2
Middle	14
Bottom	0
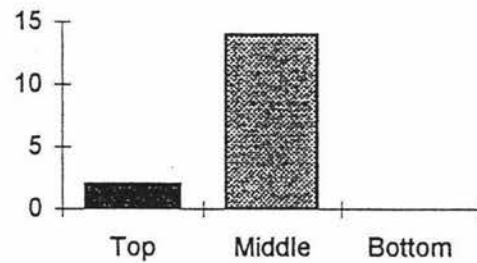
		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	5	8	5	2	0	0	2.20	0.93	2.61	1.30	-0.41	-0.32	2.00	0.61
2	Discuss.n	1	6	6	5	2	0	3.05	1.07	2.89	0.99	0.16	0.16	1.00	1.89
3	Handling *	1	4	5	9	1	0	2.75	0.99	2.89	1.20	-0.14	-0.12	1.00	1.89
4	Clarity	1	10	6	3	0	0	2.55	0.80	2.61	1.06	-0.06	-0.06	6.00	-3.39
5	Prompt	0	2	8	6	1	1	3.35	0.76	3.47	1.09	-0.12	-0.11	6.00	-2.53
6	Happy	6	9	3	0	2	0	2.15	1.15	2.28	1.45	-0.13	-0.09	1.00	1.28
7	Control *	1	7	6	2	4	0	2.95	1.20	3.29	1.18	-0.34	-0.29	6.00	-2.71
8	Challenge	0	8	6	5	0	0	2.84	0.81	3.00	1.15	-0.16	-0.14	6.00	-3.00
9	Approachable	1	5	7	4	3	0	3.15	1.11	3.29	1.36	-0.14	-0.11	6.00	-2.71
10	Related	2	10	7	1	0	0	2.35	0.73	2.71	1.32	-0.36	-0.27	6.00	-3.29
11	Enthuses	0	11	6	2	1	0	2.65	0.85	2.82	0.92	-0.17	-0.19	1.00	1.82
12	Humour	4	9	5	2	0	0	2.25	0.89	2.69	1.10	-0.44	-0.40	2.00	0.69
13	Impt Pts	3	8	8	1	0	0	2.35	0.79	2.61	0.95	-0.26	-0.27	1.00	1.61
14	Strict *	1	0	6	7	6	0	2.15	1.01	2.67	1.20	-0.52	-0.43	2.00	0.67
15	Disrupt. Std	4	7	7	1	1	0	2.40	1.02	3.12	1.28	-0.72	-0.56	1.00	2.12
16	Asmt Fair	3	8	5	4	0	0	2.50	0.97	3.00	1.28	-0.50	-0.39	2.00	1.00
17	Mat Prep	5	7	6	2	0	0	2.25	0.94	2.50	1.21	-0.25	-0.21		2.50
18	Know Wld	5	6	7	2	0	0	2.30	0.95	2.83	1.42	-0.53	-0.37	6.00	-3.17
19	Chall/Intrst	4	5	6	3	2	0	2.70	1.23	2.53	1.33	0.17	0.13	1.00	1.53
20	Improve	2	8	5	5	0	0	2.65	0.96	2.83	1.42	-0.18	-0.13	6.00	-3.17
21	Asmt Crit	1	5	12	1	1	0	2.80	0.81	2.89	0.87	-0.09	-0.10	1.00	1.89
22	Hmwk Ovld *	0	1	3	7	7	2	1.89	0.87	3.00	1.24	-1.11	-0.90	6.00	-3.00
23	Try Hard	12	4	3	1	0	0	1.65	0.91	1.94	0.91	-0.29	-0.32	6.00	-4.06
24	Boys Qns *	1	4	6	3	5	0	2.63	1.22	3.28	1.10	-0.65	-0.59	2.00	1.28
25	Ignore Tch *	4	5	8	2	1	0	3.45	1.07	3.53	1.19	-0.08	-0.07	6.00	-2.47
26	Feel Gd	2	7	6	3	2	0	2.80	1.12	3.22	1.23	-0.42	-0.34	2.00	1.22
27	Org/Expct	3	4	10	3	0	0	2.65	0.91	2.61	1.06	0.04	0.04	1.00	1.61
28	Too Much	2	6	8	2	1	1	2.68	0.98	3.11	0.94	-0.43	-0.46	3.00	0.11
29	Incr Think	2	7	6	3	2	0	2.80	1.12	3.00	1.03	-0.20	-0.19	1.00	2.00
30	Esy Undrst	2	9	5	4	0	0	2.55	0.92	3.06	1.27	-0.51	-0.40	2.00	1.06
31	Gen Conc	3	6	10	0	1	0	2.50	0.92	3.00	1.33	-0.50	-0.38	1.00	2.00
32	No Hm Hlp *	1	2	8	9	0	0	2.75	0.83	2.94	0.91	-0.19	-0.21	5.00	-2.06
33	Stdy Skls	1	7	5	6	1	0	2.95	1.02	2.89	0.74	0.06	0.08	3.00	-0.11
34	Team	3	6	5	4	2	0	2.80	1.21	2.78	1.44	0.02	0.02	2.00	0.78
35	Adpts Ablty	5	5	5	5	0	0	2.50	1.12	2.50	0.83	0.00	0.00	4.00	-1.50
36	Hmk Cpes	4	12	3	1	0	0	2.05	0.74	1.94	0.91	0.11	0.12	6.00	-4.06
37	St Disply	1	3	8	2	6	0	3.45	1.20	3.50	1.12	-0.05	-0.04	5.00	-1.50
38	Inclusive	1	6	6	3	4	0	3.15	1.19	3.00	0.94	0.15	0.16	1.00	2.00
39	Get Job	9	6	3	2	0	0	1.90	0.99	1.78	1.13	0.12	0.11	6.00	-4.22
40	S Ask Qns	3	10	4	2	1	0	2.40	1.02	2.41	1.24	-0.01	-0.01	2.00	0.41
41	Motivates	2	12	3	2	1	0	2.40	0.97	2.56	1.26	-0.16	-0.12	6.00	-3.44
42	Difficulty *	1	6	6	3	4	0	2.85	1.19	2.61	1.16	0.24	0.21	6.00	-3.39
43	Cmnts Hlp	0	11	5	2	1	1	2.63	0.87	2.94	1.00	-0.31	-0.31	6.00	-3.06
44	On Time	3	5	2	7	3	0	3.10	1.34	2.47	0.92	0.63	0.69	4.00	-1.53
45	Speed *	0	4	7	5	3	1	2.63	0.98	3.18	1.04	-0.54	-0.52	6.00	-2.82
46	Pronounce	4	10	3	2	1	0	2.30	1.05	2.61	1.38	-0.31	-0.23	2.00	0.61
47	Cont Nxt Yr	1	4	7	5	2	1	3.16	1.04	2.94	1.47	0.21	0.15	6.00	-3.06
48	Examples	0	9	8	3	0	0	2.70	0.71	2.65	0.84	0.05	0.06	1.00	1.65
49	Fin Early *	3	3	7	5	2	0	3.00	1.18	2.78	1.36	0.22	0.16	4.00	-1.22
50	Talk A Lot *	3	11	0	4	2	0	3.45	1.24	3.33	1.49	0.12	0.08	5.00	-1.67
51	Bst Tchr	3	4	6	2	4	0	3.00	1.34	2.94	1.21	0.06	0.05	6.00	-3.06
52	Bst Class	7	2	6	1	4	0	2.65	1.49	3.35	1.41	-0.70	-0.50	5.00	-1.65
53	Worthwhile	6	11	3	0	0	0	1.85	0.65						

54	Felt Good	12	5	3	0	0	0	1.55	0.74					
55	Thanked	5	8	7	0	0	0	2.10	0.77					
56	Discussed	8	4	6	1	1	0	2.15	1.15					
57	Tried Impr	5	11	2	1	1	0	2.10	0.99					
58	Did Impr	2	10	6	1	1	0	2.45	0.92					
59	Picked On *	5	3	3	3	6	0	2.90	1.58					
60	Do For All	7	8	1	3	1	0	2.15	1.19					

					Sex	-->
	Male	7				
	Female	13				

					Eth	nic	-->
	European	1					
	Maori	0					
	Pacific Is	19					
	Asian	0					
	Other	0					

					Abi	lity	-->
	Top	4					
	Middle	15					
	Bottom	0					

		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	2	7	7	1	0	0	2.41	0.77	2.63	0.81	-0.22	-0.27	1.00	1.63
2	Discuss.n	5	9	2	1	0	0	1.94	0.80	2.00	0.56	-0.06	-0.10	1.00	1.00
3	Handling *	0	0	3	8	6	0	1.82	0.71	1.95	0.89	-0.12	-0.14	5.00	-3.05
4	Clarity	3	6	4	2	2	0	2.65	1.23	2.56	1.26	0.09	0.07	2.00	0.56
5	Prompt	0	5	3	3	6	0	3.59	1.24	3.32	0.98	0.27	0.28	2.00	1.32
6	Happy	5	10	2	0	0	0	1.82	0.62	1.63	0.48	0.19	0.40	1.00	0.63
7	Control *	0	0	7	6	4	0	2.18	0.78	2.05	0.83	0.12	0.15	5.00	-2.95
8	Challenge	0	13	2	2	0	0	2.35	0.68	2.37	0.48	-0.02	-0.03	2.00	0.37
9	Approachable	0	5	6	3	3	0	3.24	1.06	3.11	1.15	0.12	0.11	6.00	-2.89
10	Related	0	10	6	1	0	0	2.47	0.61	3.16	0.74	-0.69	-0.92	2.00	1.16
11	Enthuses	7	6	2	2	0	0	1.94	1.00	2.26	0.96	-0.32	-0.33	2.00	0.26
12	Humour	4	5	2	6	0	0	2.59	1.19	2.47	1.31	0.11	0.09	2.00	0.47
13	Impt Pts	3	9	4	1	0	0	2.18	0.78	2.21	1.20	-0.03	-0.03	2.00	0.21
14	Strict *	0	3	6	8	0	0	2.71	0.75	3.16	0.74	-0.45	-0.61	3.00	0.16
15	Disrupt. Std	0	4	7	4	2	0	3.24	0.94	3.17	1.07	0.07	0.06	4.00	-0.83
16	Asmt Fair	1	9	5	1	1	0	2.53	0.92	2.42	0.75	0.11	0.14	1.00	1.42
17	Mat Prep	0	9	5	2	1	0	2.71	0.89	2.79	0.89	-0.08	-0.09	2.00	0.79
18	Know Wld	5	9	2	1	0	0	1.94	0.80	1.95	0.69	-0.01	-0.01	1.00	0.95
19	Chall/Intrst	0	8	9	0	0	0	2.53	0.50	2.63	0.93	-0.10	-0.11	2.00	0.63
20	Improve	0	7	7	3	0	0	2.76	0.73	3.05	0.83	-0.29	-0.35	6.00	-2.95
21	Asmt Crit	6	10	0	1	0	0	1.76	0.73	2.16	1.14	-0.39	-0.35	1.00	1.16
22	Hmwk Ovld *	1	1	10	3	2	0	2.76	0.94	2.26	0.96	0.50	0.52	4.00	-1.74
23	Try Hard	2	11	4	0	0	0	2.12	0.58	2.05	0.89	0.07	0.07	6.00	-3.95
24	Boys Qns *	2	2	3	6	3	0	2.63	1.27	2.68	1.17	-0.06	-0.05	5.00	-2.32
25	Ignore Tch *	1	2	5	8	1	0	2.65	0.97	2.68	0.86	-0.04	-0.04	5.00	-2.32
26	Feel Gd	0	4	8	4	1	0	3.12	0.83	3.05	0.89	0.07	0.07	1.00	2.05
27	Org/Expct	7	9	1	0	0	0	1.65	0.59	1.89	0.91	-0.25	-0.27	1.00	0.89
28	Too Much	0	8	8	0	1	0	2.65	0.76	2.58	0.75	0.07	0.09	4.00	-1.42
29	Incr Think	0	7	7	3	0	0	2.76	0.73	2.58	0.67	0.19	0.28	6.00	-3.42
30	Esy Undrst	0	8	7	2	0	0	2.65	0.68	2.63	0.93	0.02	0.02	2.00	0.63
31	Gen Conc	2	4	7	4	0	0	2.76	0.94	2.74	0.96	0.03	0.03	2.00	0.74
32	No Hm Hlp *	0	2	5	5	4	1	2.31	0.98	2.37	1.27	-0.06	-0.04	5.00	-2.63
33	Stdy Skls	0	5	4	4	4	0	3.41	1.14	3.26	1.12	0.15	0.13	2.00	1.26
34	Team	7	8	1	0	1	0	1.82	0.98	1.67	0.58	0.16	0.27	1.00	0.67
35	Adpts Ablty	2	10	5	0	0	0	2.18	0.62	2.53	0.99	-0.35	-0.35	2.00	0.53
36	Hmk Cpes	2	6	6	2	0	1	2.50	0.87	2.37	0.87	0.13	0.15	6.00	-3.63
37	St Disply	1	6	7	2	1	0	2.76	0.94	3.47	1.19	-0.71	-0.59	6.00	-2.53
38	Inclusive	6	5	4	2	0	0	2.12	1.02	2.37	0.93	-0.25	-0.27	2.00	0.37
39	Get Job	5	5	3	2	1	1	2.31	1.21	2.42	0.94	-0.11	-0.12	6.00	-3.58
40	S Ask Qns	5	7	5	0	0	0	2.00	0.77	2.63	1.09	-0.63	-0.58	1.00	1.63
41	Motivates	2	6	7	2	0	0	2.53	0.85	2.84	1.18	-0.31	-0.26	6.00	-3.16
42	Difficulty *	0	2	12	0	3	0	2.76	0.88	2.53	1.04	0.24	0.23	6.00	-3.47
43	Cmnts Hlp	1	10	4	2	0	0	2.41	0.77	2.42	0.88	-0.01	-0.01	2.00	0.42
44	On Time	1	5	5	4	2	0	3.06	1.11	2.47	0.68	0.59	0.86	1.00	1.47
45	Speed *	1	6	7	2	1	0	3.24	0.94	3.16	0.99	0.08	0.08	4.00	-0.84
46	Pronounce	6	6	3	1	0	1	1.94	0.90	1.84	0.74	0.10	0.13	1.00	0.84
47	Cont Nxt Yr	4	4	6	2	1	0	2.53	1.14	2.79	1.20	-0.26	-0.22	6.00	-3.21
48	Examples	1	7	7	1	0	0	2.50	0.71	2.26	0.85	0.24	0.28	2.00	0.26
49	Fin Early *	0	1	2	10	4	0	2.00	0.77	2.11	0.79	-0.11	-0.13	5.00	-2.89
50	Talk A Lot *	1	9	3	3	1	0	3.35	1.03	3.16	1.04	0.20	0.19	6.00	-2.84
51	Bst Tchr	1	8	6	2	0	0	2.53	0.78	2.68	1.17	-0.15	-0.13	6.00	-3.32
52	Bst Class	1	7	5	2	2	0	2.82	1.10	2.74	1.12	0.09	0.08	6.00	-3.26
53	Worthwhile	9	6	2	0	0	0	1.59	0.69						

54	Felt Good	9	7	0	1	0	0	1.59	0.77
55	Thanked	5	10	1	1	0	0	1.88	0.76
56	Discussed	5	9	2	0	1	0	2.00	0.97
57	Tried Impr	4	7	5	1	0	0	2.18	0.86
58	Did Impr	0	11	6	0	0	0	2.35	0.48
59	Picked On *	2	2	2	7	4	0	2.47	1.29
60	Do For All	14	2	0	0	1	0	1.35	0.97

Sex -->

Male	7
Female	10
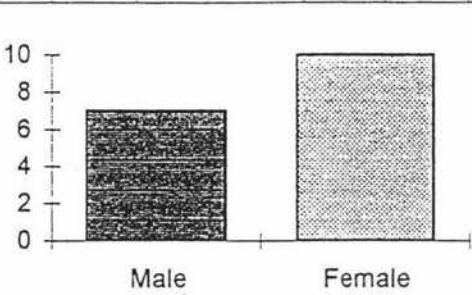
Ethnic -->

European	1
Maori	1
Pacific Is	13
Asian	2
Other	0
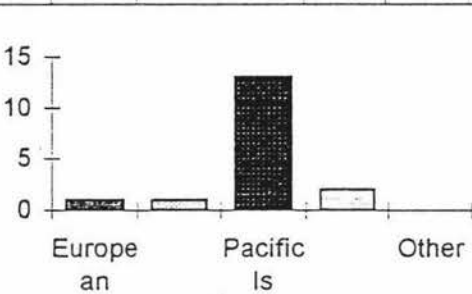
Ability -->

Top	3
Middle	13
Bottom	1
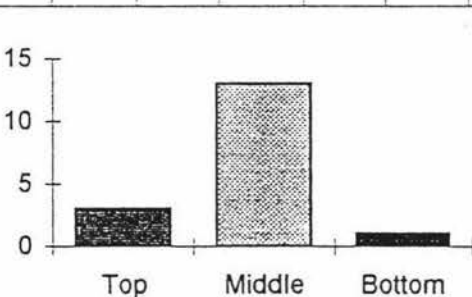
		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	1	14	6	2	0	0	2.39	0.71	2.38	0.79	0.01	0.01	2.00	0.38
2	Discuss.n	0	5	6	9	1	2	3.29	0.88	3.44	0.90	-0.15	-0.17	3.00	0.44
3	Handling *	1	3	4	8	7	0	2.26	1.15	1.81	0.96	0.45	0.47	4.00	-2.19
4	Clarity	5	12	2	3	1	0	2.26	1.07	2.12	0.93	0.15	0.16	1.00	1.12
5	Prompt	9	7	3	4	0	0	2.09	1.10	1.88	0.97	0.20	0.21	2.00	-0.12
6	Happy	8	11	3	1	0	0	1.87	0.80	1.92	0.83	-0.05	-0.06	2.00	-0.08
7	Control *	2	0	2	2	17	0	1.61	1.21	1.28	0.83	0.33	0.40	5.00	-3.72
8	Challenge	2	15	4	2	0	0	2.26	0.74	2.42	0.97	-0.16	-0.17	3.00	-0.58
9	Approachable	2	6	10	4	1	0	2.83	0.96	2.96	0.85	-0.14	-0.16	2.00	0.96
10	Related	0	2	14	5	1	1	3.23	0.67	3.54	0.84	-0.31	-0.37	4.00	-0.46
11	Enthuses	2	8	10	3	0	0	2.61	0.82	2.54	0.93	0.07	0.08	2.00	0.54
12	Humour	2	5	9	5	2	0	3.00	1.06	2.69	0.99	0.31	0.31	3.00	-0.31
13	Impt Pts	5	16	0	2	0	0	1.96	0.75	1.81	0.56	0.15	0.27	2.00	-0.19
14	Strict *	2	7	10	4	0	0	3.30	0.86	3.19	0.92	0.11	0.12	2.00	1.19
15	Disrupt. Std	1	9	1	8	4	0	3.22	1.25	3.76	0.95	-0.54	-0.57	4.00	-0.24
16	Asmt Fair	4	16	2	1	0	0	2.00	0.66	1.92	0.87	0.08	0.09	2.00	-0.08
17	Mat Prep	5	10	7	1	0	0	2.17	0.82	2.08	0.67	0.10	0.14	2.00	0.08
18	Know Wld	1	8	8	4	1	1	2.82	0.94	3.12	1.03	-0.30	-0.29	4.00	-0.88
19	Chall/Intrst	1	8	10	4	0	0	2.74	0.79	2.69	1.03	0.05	0.05	3.00	-0.31
20	Improve	3	11	6	3	0	0	2.39	0.87	2.19	0.62	0.20	0.32	3.00	-0.81
21	Asmt Crit	4	10	5	4	0	0	2.39	0.97	2.63	0.90	-0.23	-0.26	2.00	0.63
22	Hmwk Ovld *	0	5	9	7	2	0	2.74	0.90	2.76	0.81	-0.02	-0.03	3.00	-0.24
23	Try Hard	8	10	5	0	0	0	1.87	0.74	1.92	1.11	-0.05	-0.05	2.00	-0.08
24	Boys Qns *	0	0	8	4	8	3	2.00	0.89	2.08	0.93	-0.08	-0.09	4.00	-1.92
25	Ignore Tch *	0	0	2	11	9	1	1.68	0.63	1.68	0.68	0.00	0.00	4.00	-2.32
26	Feel Gd	0	5	10	6	2	0	3.22	0.88	3.12	0.97	0.10	0.10	3.00	0.12
27	Org/Expct	3	10	5	5	0	0	2.52	0.97	2.48	0.90	0.04	0.05	2.00	0.48
28	Too Much	2	7	12	1	1	0	2.65	0.87	2.96	1.09	-0.31	-0.28	4.00	-1.04
29	Incr Think	2	15	5	1	0	0	2.22	0.66	2.38	0.92	-0.17	-0.18	2.00	0.38
30	Esy Undrst	2	12	4	5	0	0	2.52	0.93	2.00	0.62	0.52	0.84	2.00	0.00
31	Gen Conc	5	8	7	3	0	0	2.35	0.96	2.54	0.97	-0.19	-0.20	2.00	0.54
32	No Hm Hlp *	1	0	2	11	9	0	1.83	0.92	1.85	1.20	-0.02	-0.02	5.00	-3.15
33	Stdy Skls	1	6	6	9	1	0	3.13	0.99	3.04	1.08	0.09	0.08	3.00	0.04
34	Team	0	4	8	8	2	1	3.36	0.88	3.28	0.78	0.08	0.11	2.00	1.28
35	Adpts Ablty	2	11	6	2	1	1	2.50	0.94	2.81	1.18	-0.31	-0.26	2.00	0.81
36	Hmk Cpes	8	14	0	1	0	0	1.74	0.67	2.19	0.92	-0.45	-0.49	3.00	-0.81
37	St Disply	0	2	5	4	8	4	3.95	1.05	4.00	1.11	-0.05	-0.05	5.00	-1.00
38	Inclusive	4	10	4	2	1	2	2.33	1.04	2.04	1.06	0.29	0.28	2.00	0.04
39	Get Job	12	4	6	1	0	0	1.83	0.96	1.73	0.76	0.10	0.13	2.00	-0.27
40	S Ask Qns	2	7	12	1	1	0	2.65	0.87	2.19	0.96	0.46	0.48	2.00	0.19
41	Motivates	2	8	9	2	2	0	2.74	1.03	2.54	0.84	0.20	0.24	2.00	0.54
42	Difficulty *	5	4	4	7	3	0	3.04	1.37	2.50	1.12	0.54	0.49	2.00	0.50
43	Cmnts Hlp	0	10	5	6	1	1	2.91	0.95	2.81	0.88	0.10	0.12	3.00	-0.19
44	On Time	1	6	7	5	4	0	3.22	1.14	3.12	0.97	0.10	0.10	5.00	-1.88
45	Speed *	3	6	5	7	2	0	3.04	1.20	2.58	1.04	0.47	0.45	4.00	-1.42
46	Pronounce	8	11	2	2	0	0	1.91	0.88	1.88	0.95	0.03	0.03	2.00	-0.12
47	Cont Nxt Yr	5	5	7	2	4	0	2.78	1.35	2.81	1.27	-0.03	-0.02	2.00	0.81
48	Examples	3	10	8	0	2	0	2.48	1.02	2.35	0.92	0.13	0.14	2.00	0.35
49	Fin Early *	0	0	1	10	12	0	1.52	0.58	1.42	0.57	0.10	0.17	4.00	-2.58
50	Talk A Lot *	0	0	4	12	7	0	1.87	0.68	1.73	0.71	0.14	0.20	4.00	-2.27
51	Bst Tchr	7	8	7	0	1	0	2.13	0.99	2.19	1.14	-0.06	-0.05	3.00	-0.81
52	Bst Class	1	10	5	5	2	0	2.87	1.08	2.77	0.93	0.10	0.11	3.00	-0.23
53	Worthwhile	5	12	6	0	0	0	2.04	0.69						

54	Felt Good	6	14	3	0	0	0	1.87	0.61					
55	Thanked	1	6	12	4	0	0	2.83	0.76					
56	Discussed	1	5	4	11	2	0	3.35	1.05					
57	Tried Impr	2	7	7	4	3	0	2.96	1.16					
58	Did Impr	0	9	9	2	3	0	2.96	1.00					
59	Picked On *	1	3	3	6	10	0	2.09	1.21					
60	Do For All	11	6	3	2	0	0	1.82	0.98					

Sex -->

Male	10
Female	13
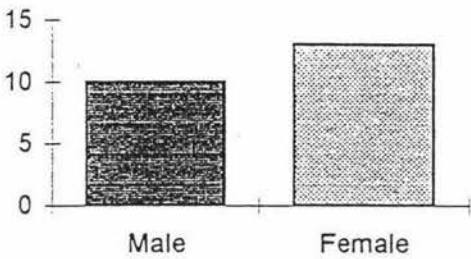
Eth nic -->

European	12
Maori	2
Pacific Is	5
Asian	4
Other	0
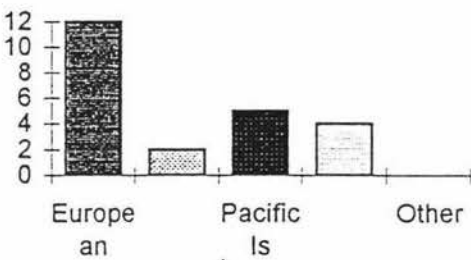
Abi lity -->

Top	3
Middle	18
Bottom	2
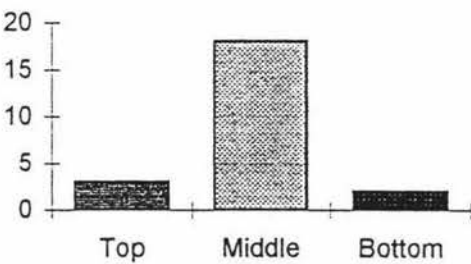
		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	1	6	9	1	0	0	2.59	0.69	2.91	0.73	-0.32	-0.44	2.00	0.91
2	Discuss.n	6	8	3	0	0	0	1.82	0.71	2.27	1.05	-0.45	-0.43	1.00	1.27
3	Handling *	0	2	2	7	6	0	2.00	0.97	2.00	1.02	0.00	0.00	5.00	-3.00
4	Clarity	2	8	3	3	0	1	2.44	0.93	2.55	0.94	-0.11	-0.11	2.00	0.55
5	Prompt	2	9	2	3	1	0	2.53	1.09	2.45	1.23	0.07	0.06	3.00	-0.55
6	Happy	4	10	3	0	0	0	1.94	0.64	1.91	0.90	0.03	0.04	1.00	0.91
7	Control *	0	2	3	9	3	0	2.24	0.88	2.14	0.92	0.10	0.11	4.00	-1.86
8	Challenge	6	8	2	0	1	0	1.94	1.00	2.18	0.65	-0.24	-0.37	2.00	0.18
9	Approachable	2	7	6	2	0	0	2.47	0.85	2.76	1.02	-0.29	-0.29	5.00	-2.24
10	Related	1	7	7	2	0	0	2.59	0.77	2.29	0.82	0.30	0.37	2.00	0.29
11	Enthuses	2	10	4	1	0	0	2.24	0.73	2.14	0.83	0.09	0.11	2.00	0.14
12	Humour	7	6	3	1	0	0	1.88	0.90	1.95	0.77	-0.07	-0.09	2.00	-0.05
13	Impt Pts	7	7	3	0	0	0	1.76	0.73	2.32	0.92	-0.55	-0.60	1.00	1.32
14	Strict *	0	0	4	10	3	0	2.06	0.64	2.05	0.84	0.01	0.01	3.00	-0.95
15	Disrupt. Std	0	3	6	4	4	0	3.53	1.04	3.27	1.14	0.26	0.23	2.00	1.27
16	Asmt Fair	1	11	3	2	0	0	2.35	0.76	2.14	0.77	0.21	0.27	1.00	1.14
17	Mat Prep	3	3	7	4	0	0	2.71	1.02	2.64	0.98	0.07	0.07	4.00	-1.36
18	Know Wld	5	9	2	1	0	0	1.94	0.80	2.14	0.89	-0.20	-0.23	2.00	0.14
19	Chall/Intrst	1	9	4	3	0	0	2.53	0.85	2.77	1.04	-0.24	-0.23	3.00	-0.23
20	Improve	0	11	4	2	0	0	2.47	0.70	2.41	0.83	0.06	0.07	1.00	1.41
21	Asmt Crit	1	10	4	1	0	0	2.31	0.68	2.90	1.11	-0.59	-0.53	2.00	0.90
22	Hmwk Ovld *	0	0	6	6	5	0	2.06	0.80	1.85	1.19	0.21	0.17	4.00	-2.15
23	Try Hard	3	5	8	1	0	0	2.41	0.84	2.41	1.19	0.00	0.00	3.00	-0.59
24	Boys Qns *	0	1	5	7	4	0	2.18	0.86	2.24	1.41	-0.06	-0.04	3.00	-0.76
25	Ignore Tch *	0	2	6	7	2	0	2.47	0.85	2.86	0.97	-0.39	-0.41	2.00	0.86
26	Feel Gd	1	3	9	3	1	0	3.00	0.91	3.09	1.16	-0.09	-0.08	2.00	1.09
27	Org/Expct	1	12	3	0	0	0	2.13	0.48	2.05	0.65	0.08	0.12	1.00	1.05
28	Too Much	0	4	3	9	1	0	3.41	0.91	3.50	0.99	-0.09	-0.09	2.00	1.50
29	Incr Think	1	8	7	1	0	0	2.47	0.70	3.00	0.93	-0.53	-0.57	3.00	0.00
30	Esy Undrst	2	9	5	1	0	0	2.29	0.75	2.55	1.08	-0.25	-0.23	2.00	0.55
31	Gen Conc	1	11	2	3	0	0	2.41	0.84	2.55	0.99	-0.13	-0.14	1.00	1.55
32	No Hm Hlp *	1	3	1	6	6	0	2.24	1.26	2.32	1.43	-0.08	-0.06	3.00	-0.68
33	Stdy Skls	3	5	3	5	0	1	2.63	1.11	3.15	0.91	-0.53	-0.58	2.00	1.15
34	Team	1	7	7	1	0	1	2.50	0.71	3.00	1.04	-0.50	-0.48	3.00	0.00
35	Adpts Ablty	1	9	5	2	0	0	2.47	0.78	2.62	1.05	-0.15	-0.14	2.00	0.62
36	Hmk Cpes	5	9	1	1	0	1	1.88	0.78	1.77	0.73	0.10	0.14	1.00	0.77
37	St Disply	2	1	3	7	2	1	3.40	1.20	3.35	1.31	0.05	0.04	4.00	-0.65
38	Inclusive	4	7	3	1	0	2	2.07	0.85	2.14	0.97	-0.07	-0.07	2.00	0.14
39	Get Job	4	8	3	1	1	0	2.24	1.06	2.41	1.23	-0.17	-0.14	2.00	0.41
40	S Ask Qns	6	8	2	1	0	0	1.88	0.83	1.95	1.09	-0.07	-0.06	2.00	-0.05
41	Motivates	1	4	9	3	0	0	2.82	0.78	2.90	1.06	-0.08	-0.08	3.00	-0.10
42	Difficulty *	1	2	4	3	7	0	2.24	1.26	2.82	1.37	-0.58	-0.43	4.00	-1.18
43	Cmnts Hlp	3	8	6	0	0	0	2.18	0.71	2.36	1.15	-0.19	-0.16	2.00	0.36
44	On Time	0	6	8	2	1	0	2.88	0.83	2.77	1.00	0.11	0.11	4.00	-1.23
45	Speed *	0	3	2	12	0	0	2.47	0.78	2.62	1.09	-0.15	-0.14	2.00	0.62
46	Pronounce	5	10	1	0	0	1	1.75	0.56	1.73	0.69	0.02	0.03	1.00	0.73
47	Cont Nxt Yr	3	1	7	2	3	1	3.06	1.30	2.82	1.37	0.24	0.18	3.00	-0.18
48	Examples	4	9	3	1	0	0	2.06	0.80	2.19	0.85	-0.13	-0.15	2.00	0.19
49	Fin Early *	2	6	6	2	1	0	3.35	1.03	3.14	0.83	0.21	0.25	2.00	1.14
50	Talk A Lot *	2	5	5	3	2	0	3.12	1.18	3.33	0.99	-0.22	-0.22	2.00	1.33
51	Bst Tchr	4	5	8	0	0	0	2.24	0.81	2.50	1.12	-0.26	-0.24	3.00	-0.50
52	Bst Class	3	4	7	2	1	0	2.65	1.08	2.95	1.33	-0.31	-0.23	3.00	-0.05
53	Worthwhile	4	8	5	0	0	0	2.06	0.73						

54	Felt Good	4	8	4	1	0	0	2.12	0.83					
55	Thanked	1	5	8	2	0	0	2.69	0.77					
56	Discussed	2	8	1	4	0	0	2.47	1.02					
57	Tried Impr	3	7	5	2	0	0	2.35	0.90					
58	Did Impr	1	8	6	2	0	0	2.53	0.78					
59	Picked On *	0	0	3	7	7	0	1.76	0.73					
60	Do For All	6	5	4	1	1	0	2.18	1.15					

Sex -->

Male	7
Female	10
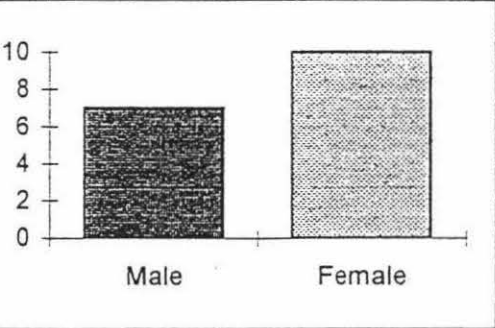
Eth nic -->

European	11
Maori	3
Pacific Is	1
Asian	1
Other	1
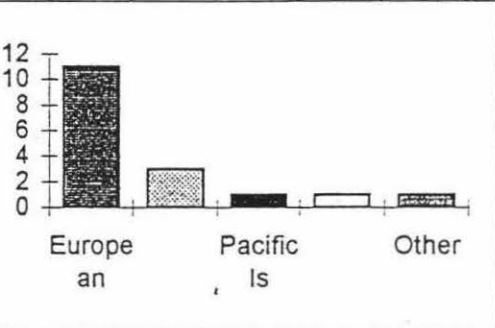
Abi lity -->

Top	7
Middle	9
Bottom	1
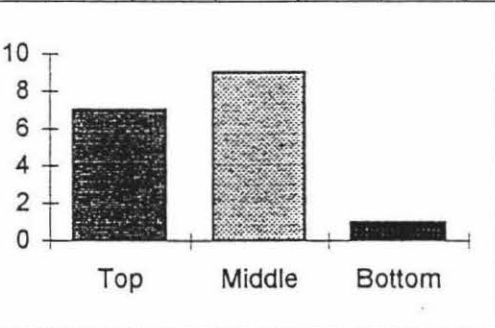
	1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1 Pride	1	7	5	0	0	0	2.31	0.61	2.42	0.76	-0.11	-0.14	2.00	0.42
2 Discuss.n	3	9	1	0	0	0	1.85	0.53	2.25	0.83	-0.40	-0.49	2.00	0.25
3 Handling *	0	1	1	9	2	0	2.08	0.73	2.17	0.90	-0.09	-0.10	1.00	1.17
4 Clarity	0	1	4	7	1	0	3.62	0.74	3.42	0.95	0.20	0.21	2.00	1.42
5 Prompt	0	4	5	4	0	0	3.00	0.78	3.33	1.18	-0.33	-0.28	4.00	-0.67
6 Happy	5	7	1	0	0	0	1.69	0.61	2.00	0.82	-0.31	-0.38	2.00	0.00
7 Control *	0	1	5	5	2	0	2.38	0.84	3.00	1.15	-0.62	-0.53	5.00	-2.00
8 Challenge	3	8	2	0	0	0	1.92	0.62	2.25	1.09	-0.33	-0.30	1.00	1.25
9 Approachable	1	4	3	4	1	0	3.00	1.11	3.92	1.26	-0.92	-0.73	4.00	-0.08
10 Related	0	2	2	8	1	0	3.62	0.84	4.17	0.69	-0.55	-0.80	5.00	-0.83
11 Enthuses	3	9	1	0	0	0	1.85	0.53	1.67	0.75	0.18	0.24	2.00	-0.33
12 Humour	1	8	3	1	0	0	2.31	0.72	2.58	0.95	-0.28	-0.29	2.00	0.58
13 Impt Pts	2	6	4	1	0	0	2.31	0.82	2.25	1.01	0.06	0.06	2.00	0.25
14 Strict *	0	0	4	8	1	0	2.23	0.58	2.33	0.85	-0.10	-0.12	3.00	-0.67
15 Disrupt. Std	0	1	4	5	3	0	3.77	0.89	3.22	0.79	0.55	0.70	2.00	1.22
16 Asmt Fair	1	10	1	1	0	0	2.15	0.66	2.17	0.55	-0.01	-0.02	1.00	1.17
17 Mat Prep	1	3	5	4	0	0	2.92	0.92	2.92	0.86	0.01	0.01	2.00	0.92
18 Know Wld	1	2	5	4	1	0	3.15	1.03	2.92	1.19	0.24	0.20	2.00	0.92
19 Chall/Intrst	0	8	4	1	0	0	2.46	0.63	2.50	1.12	-0.04	-0.03	2.00	0.50
20 Improve	0	7	5	1	0	0	2.54	0.63	2.92	0.95	-0.38	-0.40	2.00	0.92
21 Asmt Crit	3	6	2	1	0	1	2.08	0.86	2.00	0.60	0.08	0.14	2.00	0.00
22 Hmwk Ovld *	0	0	9	4	0	0	2.69	0.46	2.58	0.64	0.11	0.17	5.00	-2.42
23 Try Hard	0	6	7	0	0	0	2.54	0.50	2.58	0.86	-0.04	-0.05	2.00	0.58
24 Boys Qns *	2	1	0	7	3	0	2.38	1.33	2.09	0.90	0.29	0.33	5.00	-2.91
25 Ignore Tch *	0	1	6	6	0	0	2.62	0.62	2.50	0.65	0.12	0.18	4.00	-1.50
26 Feel Gd	0	1	9	3	0	0	3.15	0.53	3.17	0.80	-0.01	-0.02	1.00	2.17
27 Org/Expct	2	6	5	0	0	0	2.23	0.70	2.25	0.92	-0.02	-0.02	2.00	0.25
28 Too Much	1	2	0	10	0	0	3.46	1.01	3.08	1.04	0.38	0.36	2.00	1.08
29 Incr Think	0	7	2	4	0	0	2.77	0.89	2.50	1.04	0.27	0.26	2.00	0.50
30 Esy Undrst	0	4	7	1	1	0	2.92	0.83	3.18	1.03	-0.26	-0.25	2.00	1.18
31 Gen Conc	4	7	1	0	0	0	1.75	0.60	2.33	1.25	-0.58	-0.47	2.00	0.33
32 No Hm Hlp *	0	1	0	7	5	0	1.77	0.80	2.08	1.38	-0.31	-0.23	5.00	-2.92
33 Stdy Skls	0	5	2	4	2	0	3.23	1.12	2.92	1.11	0.31	0.28	3.00	-0.08
34 Team	5	6	2	0	0	0	1.77	0.70	2.50	0.96	-0.73	-0.76	2.00	0.50
35 Adpts Ablty	1	4	6	2	0	0	2.69	0.82	2.67	0.94	0.03	0.03	2.00	0.67
36 Hmk Cpes	2	8	2	1	0	0	2.15	0.77	2.42	0.76	-0.26	-0.35	2.00	0.42
37 St Disply	0	0	1	2	8	2	4.64	0.64	4.80	0.40	-0.16	-0.41	5.00	-0.20
38 Inclusive	4	7	1	1	0	0	1.92	0.83	2.67	1.18	-0.74	-0.63	1.00	1.67
39 Get Job	3	5	2	1	2	0	2.54	1.34	2.50	0.96	0.04	0.04	6.00	-3.50
40 S Ask Qns	2	6	3	2	0	0	2.38	0.92	3.08	1.11	-0.70	-0.63	2.00	1.08
41 Motivates	0	3	4	5	1	0	3.31	0.91	3.42	0.86	-0.11	-0.13	6.00	-2.58
42 Difficulty *	2	1	4	5	1	0	2.85	1.17	2.83	1.07	0.01	0.01	6.00	-3.17
43 Cmnts Hlp	4	8	0	1	0	0	1.85	0.77	1.83	0.90	0.01	0.01	1.00	0.83
44 On Time	0	3	1	5	4	0	3.77	1.12	3.25	1.09	0.52	0.48	4.00	-0.75
45 Speed *	0	2	4	7	0	0	2.62	0.74	2.75	0.83	-0.13	-0.16	4.00	-1.25
46 Pronounce	3	7	1	0	2	0	2.31	1.26	2.17	1.14	0.14	0.12	6.00	-3.83
47 Cont Nxt Yr	2	0	2	2	3	4	3.44	1.50	3.00	1.31	0.44	0.34	6.00	-3.00
48 Examples	0	11	0	1	1	0	2.38	0.92	2.50	0.87	-0.12	-0.13	2.00	0.50
49 Fin Early *	0	1	0	12	0	0	2.15	0.53	2.25	0.60	-0.10	-0.16	5.00	-2.75
50 Talk A Lot *	0	2	7	4	0	0	2.85	0.66	3.00	1.00	-0.15	-0.15	1.00	2.00
Bst Tchr	1	1	5	6	0	0	3.23	0.89	3.42	1.11	-0.19	-0.17	6.00	-2.58
Bst Class	2	1	4	4	2	0	3.23	1.25	3.25	1.42	-0.02	-0.01	6.00	-2.75
Worthwhile	4	7	2	0	0	0	1.85	0.66						

54	Felt Good	3	6	4	0	0	0	2.08	0.73
55	Thanked	0	2	7	1	3	0	3.38	1.00
56	Discussed	0	0	2	5	6	0	4.31	0.72
57	Tried Impr	0	2	6	4	1	0	3.31	0.82
58	Did Impr	0	1	8	3	1	0	3.31	0.72
59	Picked On *	0	0	5	3	5	0	2.00	0.88
60	Do For All	5	4	4	0	0	0	1.92	0.83

Sex -->

Male	8
Female	5

Ethnic -->

European	8
Maori	0
Pacific Is	3
Asian	0
Other	2

Ability -->

Top	2
Middle	10
Bottom	1

		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	4	14	4	0	0	0	2.00	0.60	1.64	0.71	0.36	0.51	2.00	-0.36
2	Discuss.n	8	11	2	1	0	0	1.82	0.78	1.73	0.69	0.09	0.13	2.00	-0.27
3	Handling *	1	4	3	7	7	0	2.32	1.22	2.19	1.22	0.13	0.10	4.00	-1.81
4	Clarity	10	7	4	1	0	0	1.82	0.89	1.73	0.69	0.09	0.13	1.00	0.73
5	Prompt	8	8	4	2	0	0	2.00	0.95	1.45	0.72	0.55	0.76	1.00	0.45
6	Happy	5	9	5	2	1	0	2.32	1.06	2.00	0.80	0.32	0.40	3.00	-1.00
7	Control *	0	1	2	8	11	0	1.68	0.82	1.45	0.80	0.23	0.29	5.00	-3.55
8	Challenge	4	9	6	3	0	0	2.36	0.93	2.45	0.99	-0.09	-0.09	3.00	-0.55
9	Approachable	2	3	7	3	7	0	3.45	1.30	3.41	0.89	0.05	0.05	6.00	-2.59
10	Related	0	7	9	5	1	0	3.00	0.85	3.23	1.04	-0.23	-0.22	3.00	0.23
11	Enthuses	4	12	5	1	0	0	2.14	0.76	2.10	0.81	0.04	0.05	2.00	0.10
12	Humour	2	5	10	3	1	0	2.81	0.96	2.82	1.03	-0.01	-0.01	4.00	-1.18
13	Impt Pts	7	11	2	1	1	0	2.00	1.00	1.86	0.56	0.14	0.26	3.00	-1.14
14	Strict *	1	6	10	3	2	0	3.05	0.98	3.36	0.98	-0.32	-0.32	2.00	1.36
15	Disrupt. Std	4	7	6	3	2	0	2.64	1.19	3.00	1.38	-0.36	-0.26	6.00	-3.00
16	Asmt Fair	13	8	0	1	0	0	1.50	0.72	1.36	0.57	0.14	0.24	2.00	-0.64
17	Mat Prep	6	15	1	0	0	0	1.77	0.52	2.00	0.62	-0.23	-0.37	3.00	-1.00
18	Know Wld	10	9	1	2	0	0	1.77	0.90	1.68	0.82	0.09	0.11	1.00	0.68
19	Chall/Intrst	5	9	6	2	0	0	2.23	0.90	2.18	0.98	0.05	0.05	2.00	0.18
20	Improve	5	7	7	1	2	0	2.45	1.16	2.09	0.79	0.36	0.46	6.00	-3.91
21	Asmt Crit	4	8	8	2	0	0	2.36	0.88	2.06	0.97	0.31	0.32	4.00	-1.94
22	Hmwk Ovld *	5	4	4	6	2	1	3.19	1.33	2.82	1.34	0.37	0.28	4.00	-1.18
23	Try Hard	9	9	4	0	0	0	1.77	0.73	1.50	0.50	0.27	0.55	2.00	-0.50
24	Boys Qns *	3	5	4	0	0	10	3.92	0.76	3.60	1.20	0.32	0.26	6.00	-2.40
25	Ignore Tch *	0	3	10	7	2	0	2.64	0.83	2.27	0.91	0.36	0.40	4.00	-1.73
26	Feel Gd	3	3	8	4	3	0	3.05	1.21	3.29	0.98	-0.24	-0.24	3.00	0.29
27	Org/Expct	5	7	6	2	1	1	2.38	1.09	2.33	1.36	0.05	0.04	2.00	0.33
28	Too Much	3	4	8	5	2	0	2.95	1.15	3.27	1.42	-0.32	-0.22	5.00	-1.73
29	Incr Think	3	16	2	1	0	0	2.05	0.64	2.10	0.68	-0.05	-0.07	6.00	-3.90
30	Esy Undrst	4	9	5	3	1	0	2.45	1.08	2.27	1.01	0.18	0.18	6.00	-3.73
31	Gen Conc	3	8	7	4	0	0	2.55	0.94	2.64	0.98	-0.09	-0.09	2.00	0.64
32	No Hm Hlp *	1	4	3	4	8	2	2.30	1.31	2.10	1.34	0.20	0.15	5.00	-2.90
33	Stdy Skls	5	6	6	3	2	0	2.59	1.23	2.45	1.12	0.14	0.13	4.00	-1.55
34	Team	5	9	5	1	2	0	2.36	1.15	2.14	1.10	0.23	0.21	1.00	1.14
35	Adpts Ablty	4	11	6	1	0	0	2.18	0.78	2.45	1.20	-0.27	-0.23	2.00	0.45
36	Hmk Cpes	7	7	4	1	3	0	2.36	1.33	1.73	1.14	0.64	0.56	6.00	-4.27
37	St Disply	9	9	1	2	1	0	1.95	1.11	1.68	0.70	0.27	0.39	1.00	0.68
38	Inclusive	9	9	4	0	0	0	1.77	0.73	1.90	0.70	-0.13	-0.18	1.00	0.90
39	Get Job	3	7	9	1	1	0	2.52	0.96	2.32	1.02	0.21	0.20	6.00	-3.68
40	S Ask Qns	8	6	5	2	1	0	2.18	1.15	2.45	1.20	-0.27	-0.23	2.00	0.45
41	Motivates	4	9	6	3	0	0	2.36	0.93	2.29	1.20	0.08	0.06	6.00	-3.71
42	Difficulty *	2	1	6	2	10	1	2.19	1.33	1.62	0.79	0.57	0.73	6.00	-4.38
43	Cmnts Hlp	5	11	3	2	1	0	2.23	1.04	2.09	1.20	0.14	0.11	3.00	-0.91
44	On Time	10	9	3	0	0	0	1.68	0.70	1.36	0.48	0.32	0.66	1.00	0.36
45	Speed *	4	3	6	5	3	1	3.00	1.31	2.27	1.29	0.73	0.57	4.00	-1.73
46	Pronounce	13	7	2	0	0	0	1.50	0.66	1.29	0.55	0.21	0.39	1.00	0.29
47	Cont Nxt Yr	10	3	3	3	3	0	2.36	1.49	2.09	1.41	0.27	0.19	6.00	-3.91
48	Examples	7	7	8	0	0	0	2.05	0.82	2.50	1.16	-0.45	-0.39	3.00	-0.50
49	Fin Early *	2	4	4	6	6	0	2.55	1.30	2.25	1.18	0.30	0.25	4.00	-1.75
50	Talk A Lot *	3	4	4	7	4	0	2.77	1.31	1.81	1.18	0.96	0.82	3.00	-1.19
51	Bst Tchr	9	4	3	3	2	0	2.29	1.39	2.36	1.19	-0.08	-0.07	6.00	-3.64
52	Bst Class	8	6	3	2	3	0	2.36	1.40	2.41	1.30	-0.05	-0.03	3.00	-0.59
53	Worthwhile	14	5	2	0	1	0	1.59	0.98						

54	Felt Good	15	7	0	0	0	0	1.32	0.47					
55	Thanked	1	7	12	0	2	0	2.77	0.90					
56	Discussed	4	7	4	4	3	0	2.77	1.31					
57	Tried Impr	6	7	6	1	2	0	2.36	1.19					
58	Did Impr	6	7	7	1	1	0	2.27	1.05					
59	Picked On *	1	0	6	1	14	0	1.77	1.13					
60	Do For All	14	3	4	0	1	0	1.68	1.06					

Sex -->

Male	22
Female	0
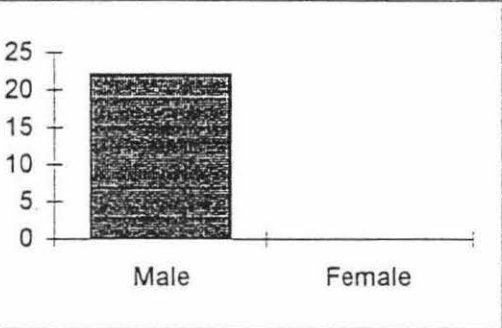
Ethnic -->

European	20
Maori	2
Pacific Is	0
Asian	0
Other	0
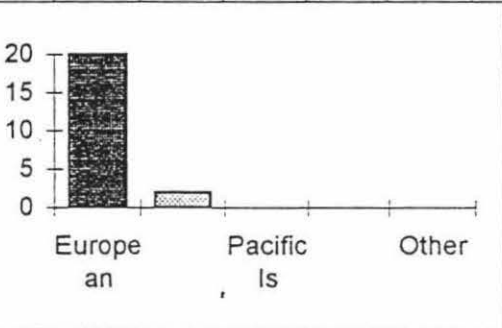
Ability -->

Top	15
Middle	6
Bottom	1
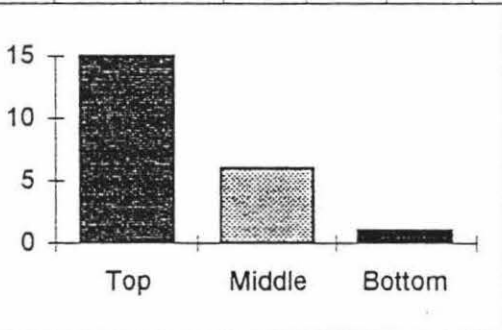
		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	6	9	4	3	0	0	2.18	0.98	2.17	0.75	0.02	0.02	1.00	1.17
2	Discuss.n	2	10	7	2	0	1	2.43	0.79	2.41	1.03	0.02	0.02	1.00	1.41
3	Handling *	0	1	6	6	9	0	1.95	0.93	1.75	0.88	0.20	0.23	5.00	-3.25
4	Clarity	2	12	3	3	1	0	2.48	1.01	2.21	1.08	0.27	0.25	1.00	1.21
5	Prompt	0	0	4	8	9	0	4.24	0.75	3.83	0.94	0.40	0.43	4.00	-0.17
6	Happy	3	14	3	0	1	0	2.14	0.83	1.71	0.73	0.43	0.59	1.00	0.71
7	Control *	1	1	2	11	7	0	2.00	1.00	1.63	1.07	0.38	0.35	5.00	-3.38
8	Challenge	8	10	3	1	0	0	1.86	0.81	1.46	0.58	0.41	0.70	1.00	0.46
9	Approachable	5	5	3	5	3	1	2.81	1.40	2.58	1.19	0.23	0.19	2.00	0.58
10	Related	2	7	3	7	2	1	3.00	1.20	3.04	1.12	-0.04	-0.04	3.00	0.04
11	Enthuses	8	8	5	0	1	0	2.00	1.00	1.96	0.73	0.04	0.06	1.00	0.96
12	Humour	3	10	4	3	2	0	2.59	1.15	2.29	0.93	0.30	0.32	2.00	0.29
13	Impt Pts	9	10	1	1	0	0	1.71	0.76	1.58	0.86	0.13	0.15	1.00	0.58
14	Strict *	0	4	14	3	1	0	2.95	0.71	3.00	0.91	-0.05	-0.05	2.00	1.00
15	Disrupt. Std	4	5	6	2	5	0	2.95	1.40	3.38	1.03	-0.42	-0.41	5.00	-1.63
16	Asmt Fair	10	9	3	0	0	0	1.68	0.70	1.63	0.70	0.06	0.08	1.00	0.63
17	Mat Prep	3	10	6	1	1	1	2.38	0.95	2.17	0.69	0.21	0.31	3.00	-0.83
18	Know Wld	3	3	8	2	4	1	3.05	1.28	2.80	0.93	0.25	0.27	3.00	-0.20
19	Chall/Intrst	9	11	1	1	0	0	1.73	0.75	1.71	0.79	0.02	0.02	1.00	0.71
20	Improve	8	10	0	2	1	0	1.95	1.09	1.83	0.99	0.12	0.12	1.00	0.83
21	Asmt Crit	4	7	9	1	1	0	2.45	0.99	2.61	0.97	-0.15	-0.16	2.00	0.61
22	Hmwk Ovld *	2	8	7	3	2	0	3.23	1.08	2.79	0.96	0.44	0.46	3.00	-0.21
23	Try Hard	10	6	6	0	0	0	1.82	0.83	1.75	0.97	0.07	0.07	3.00	-1.25
24	Boys Qns *	4	1	1	0	0	16	4.50	0.76	4.50	0.87	0.00	0.00	3.00	1.50
25	Ignore Tch *	0	2	7	9	3	0	2.38	0.84	2.13	1.05	0.26	0.24	5.00	-2.88
26	Feel Gd	2	4	10	2	4	0	3.09	1.16	2.92	0.86	0.17	0.20	1.00	1.92
27	Org/Expct	4	7	5	4	1	1	2.57	1.14	2.52	1.21	0.05	0.04	1.00	1.52
28	Too Much	0	7	8	5	2	0	3.09	0.95	3.00	0.87	0.09	0.10	3.00	0.00
29	Incr Think	8	13	1	0	0	0	1.68	0.55	1.75	0.66	-0.07	-0.10	2.00	-0.25
30	Esy Undrst	8	12	2	0	0	0	1.73	0.62	1.92	1.04	-0.19	-0.18	1.00	0.92
31	Gen Conc	6	8	4	3	1	0	2.32	1.14	2.00	1.08	0.32	0.29	1.00	1.00
32	No Hm Hlp *	2	0	2	8	10	0	1.91	1.16	1.71	1.17	0.20	0.17	4.00	-2.29
33	Stdy Skls	8	6	4	4	0	0	2.18	1.11	2.54	0.96	-0.36	-0.38	2.00	0.54
34	Team	0	2	8	8	3	1	3.57	0.85	3.35	0.96	0.22	0.23	3.00	0.35
35	Adpts Ablty	6	6	5	2	1	2	2.30	1.14	2.30	1.00	0.00	0.00	2.00	0.30
36	Hmk Cpes	6	9	2	5	0	0	2.27	1.09	2.08	0.81	0.19	0.23	6.00	-3.92
37	St Disply	1	2	1	5	13	0	4.23	1.17	3.21	1.08	1.02	0.94	4.00	-0.79
38	Inclusive	9	3	7	0	2	1	2.19	1.26	1.92	1.11	0.27	0.25	2.00	-0.08
39	Get Job	12	5	4	0	1	0	1.77	1.04	1.50	0.82	0.27	0.33	2.00	-0.50
40	S Ask Qns	14	4	3	1	0	0	1.59	0.89	1.63	0.75	-0.03	-0.05	1.00	0.63
41	Motivates	8	8	4	2	0	0	2.00	0.95	2.25	1.13	-0.25	-0.22	2.00	0.25
42	Difficulty *	2	4	4	6	6	0	2.55	1.30	2.38	0.99	0.17	0.17	6.00	-3.63
43	Cmnts Hlp	3	6	9	1	2	1	2.67	1.08	2.35	1.17	0.32	0.27	2.00	0.35
44	On Time	4	10	3	2	3	0	2.55	1.27	1.96	0.79	0.59	0.74	1.00	0.96
45	Speed *	1	5	9	5	2	0	2.91	1.00	2.54	1.15	0.37	0.32	3.00	-0.46
46	Pronounce	13	4	2	1	1	1	1.71	1.12	1.50	0.82	0.21	0.26	1.00	0.50
47	Cont Nxt Yr	10	5	5	2	0	0	1.95	1.02	1.71	1.02	0.25	0.24	6.00	-4.29
48	Examples	8	7	4	1	1	1	2.05	1.09	1.91	0.83	0.13	0.16	2.00	-0.09
49	Fin Early *	1	3	0	4	13	1	1.81	1.26	2.25	1.13	-0.44	-0.39	4.00	-1.75
50	Talk A Lot *	1	0	6	8	7	0	2.09	1.00	2.04	0.89	0.05	0.06	5.00	-2.96
51	Bst Tchr	13	4	4	0	1	0	1.73	1.05	1.67	1.07	0.06	0.06	6.00	-4.33
52	Bst Class	4	11	4	2	1	0	2.32	1.02	2.13	1.12	0.19	0.17	6.00	-3.87
53	Worthwhile	11	7	4	0	0	0	1.68	0.76						

54	Felt Good	7	11	4	0	0	0	1.86	0.69
55	Thanked	3	12	6	0	0	0	2.14	0.64
56	Discussed	11	9	1	0	1	0	1.68	0.92
57	Tried Impr	6	7	5	2	2	0	2.41	1.23
58	Did Impr	4	3	11	2	2	0	2.77	1.13
59	Picked On *	1	0	3	5	12	0	1.71	1.03
60	Do For All	14	4	2	1	1	0	1.68	1.10

Sex -->

| Male | 22 |
| Female | 0 |

Ethnic -->

European	17
Maori	2
Pacific Is	0
Asian	0
Other	3

Ability -->

Top	13
Middle	7
Bottom	2

		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	4	8	6	1	0	0	2.21	0.83	2.28	0.80	-0.07	-0.08	2.00	0.28
2	Discuss.n	9	7	3	0	0	0	1.68	0.73	1.89	0.66	-0.20	-0.31	1.00	0.89
3	Handling *	0	0	0	4	15	0	1.21	0.41	1.28	0.56	-0.07	-0.12	4.00	-2.72
4	Clarity	10	7	2	0	0	0	1.58	0.67	1.78	0.97	-0.20	-0.20	1.00	0.78
5	Prompt	14	5	0	0	0	0	1.26	0.44	1.22	0.42	0.04	0.10	1.00	0.22
6	Happy	7	10	1	1	0	0	1.79	0.77	1.72	0.65	0.07	0.10	1.00	0.72
7	Control *	0	0	0	9	10	0	1.47	0.50	1.22	0.53	0.25	0.47	5.00	-3.78
8	Challenge	4	9	4	2	0	0	2.21	0.89	2.28	0.73	-0.07	-0.09	2.00	0.28
9	Approachable	8	6	2	2	1	0	2.05	1.19	2.06	1.08	0.00	0.00	2.00	0.06
10	Related	6	5	5	3	0	0	2.26	1.07	2.00	0.88	0.26	0.30	2.00	0.00
11	Enthuses	16	3	0	0	0	0	1.16	0.36	1.17	0.37	-0.01	-0.02	1.00	0.17
12	Humour	15	4	0	0	0	0	1.21	0.41	1.33	0.67	-0.12	-0.18	1.00	0.33
13	Impt Pts	11	7	1	0	0	0	1.47	0.60	1.61	0.83	-0.14	-0.17	1.00	0.61
14	Strict *	0	2	8	6	3	0	2.47	0.88	2.61	0.76	-0.14	-0.18	3.00	-0.39
15	Disrupt. Std	0	5	2	8	4	0	3.58	1.09	3.89	0.87	-0.31	-0.35	4.00	-0.11
16	Asmt Fair	6	11	2	0	0	0	1.79	0.61	1.50	0.50	0.29	0.58	2.00	-0.50
17	Mat Prep	14	4	1	0	0	0	1.32	0.57	1.39	0.59	-0.07	-0.12	2.00	-0.61
18	Know Wld	4	6	9	0	0	0	2.26	0.78	2.44	1.12	-0.18	-0.16	2.00	0.44
19	Chall/Intrst	8	10	1	0	0	0	1.63	0.58	1.61	0.68	0.02	0.03	2.00	-0.39
20	Improve	8	10	1	0	0	0	1.63	0.58	1.94	0.85	-0.31	-0.37	2.00	-0.06
21	Asmt Crit	10	7	2	0	0	0	1.58	0.67	1.83	0.83	-0.25	-0.31	2.00	-0.17
22	Hmwk Ovld *	0	2	3	11	3	0	2.21	0.83	1.94	0.73	0.27	0.37	3.00	-1.06
23	Try Hard	4	10	3	2	0	0	2.16	0.87	2.17	0.90	-0.01	-0.01	2.00	0.17
24	Boys Qns *	0	0	0	0	0	19	#####	#####	#####	#####	#####	#####	6.00	#####
25	Ignore Tch *	0	0	1	8	10	0	1.53	0.60	1.94	1.27	-0.42	-0.33	4.00	-2.06
26	Feel Gd	3	11	3	2	0	0	2.21	0.83	2.22	0.71	-0.01	-0.02	1.00	1.22
27	Org/Expct	11	8	0	0	0	0	1.42	0.49	1.44	0.60	-0.02	-0.04	1.00	0.44
28	Too Much	1	4	7	7	0	0	3.05	0.89	3.22	0.97	-0.17	-0.17	1.00	2.22
29	Incr Think	3	10	5	1	0	0	2.21	0.77	2.22	0.53	-0.01	-0.02	2.00	0.22
30	Esy Undrst	4	13	2	0	0	0	1.89	0.55	1.89	0.81	0.01	0.01	1.00	0.89
31	Gen Conc	10	8	1	0	0	0	1.53	0.60	1.28	0.56	0.25	0.45	1.00	0.28
32	No Hm Hlp *	1	1	0	1	16	0	1.42	1.09	1.44	1.01	-0.02	-0.02	4.00	-2.56
33	Stdy Skls	2	4	9	3	1	0	2.84	0.99	2.83	0.96	0.01	0.01	2.00	0.83
34	Team	1	12	4	2	0	0	2.37	0.74	2.61	0.83	-0.24	-0.29	2.00	0.61
35	Adpts Ablty	5	10	4	0	0	0	1.95	0.69	1.83	0.50	0.11	0.23	2.00	-0.17
36	Hmk Cpes	7	9	1	2	0	0	1.89	0.91	1.78	0.85	0.12	0.14	3.00	-1.22
37	St Disply	2	10	4	2	0	0	2.33	0.82	1.89	0.81	0.44	0.55	2.00	-0.11
38	Inclusive	9	6	3	1	0	0	1.79	0.89	1.67	0.88	0.12	0.14	3.00	-1.33
39	Get Job	1	4	10	3	1	0	2.95	0.89	2.67	0.94	0.28	0.30	4.00	-1.33
40	S Ask Qns	7	12	0	0	0	0	1.63	0.48	1.39	0.59	0.24	0.41	2.00	-0.61
41	Motivates	6	7	3	3	0	0	2.16	1.04	2.28	0.73	-0.12	-0.16	2.00	0.28
42	Difficulty *	1	2	5	6	5	0	2.37	1.13	2.56	1.17	-0.19	-0.16	2.00	0.56
43	Cmnts Hlp	7	10	2	0	0	0	1.74	0.64	1.83	0.76	-0.10	-0.13	2.00	-0.17
44	On Time	9	8	1	1	0	0	1.68	0.80	1.72	0.73	-0.04	-0.05	1.00	0.72
45	Speed *	0	4	5	7	3	0	2.53	0.99	2.00	0.75	0.53	0.71	2.00	0.00
46	Pronounce	15	3	1	0	0	0	1.26	0.55	1.22	0.53	0.04	0.08	1.00	0.22
47	Cont Nxt Yr	5	4	1	5	3	0	2.83	1.50	3.00	1.37	-0.17	-0.12	3.00	0.00
48	Examples	6	8	3	1	0	0	1.94	0.85	1.71	0.67	0.24	0.36	2.00	-0.29
49	Fin Early *	0	1	2	8	8	0	1.79	0.83	1.94	0.91	-0.15	-0.17	4.00	-2.06
50	Talk A Lot *	0	4	7	7	1	0	2.74	0.85	2.67	1.00	0.07	0.07	1.00	1.67
51	Bst Tchr	14	4	1	0	0	0	1.32	0.57	1.39	0.59	-0.07	-0.12		1.39
52	Bst Class	6	9	2	1	1	0	2.05	1.05	2.11	0.87	-0.06	-0.07		2.11
53	Worthwhile	7	10	0	2	0	0	1.84	0.87						

54	Felt Good	7	11	0	1	0	0	1.74	0.71					
55	Thanked	9	9	0	0	1	0	1.68	0.92					
56	Discussed	12	6	0	0	1	0	1.53	0.94					
57	Tried Impr	13	5	1	0	0	0	1.37	0.58					
58	Did Impr	7	8	3	1	0	0	1.89	0.85					
59	Picked On *	0	0	1	3	15	0	1.26	0.55					
60	Do For All	8	8	1	1	1	0	1.89	1.07					

Sex -->

| Male | 0 |
| Female | 19 |

Ethnic -->

European	19
Maori	0
Pacific Is	0
Asian	0
Other	0

Ability -->

Top	6
Middle	10
Bottom	3

		1	2	3	4	5	6	A Mn2	B SD2	C Mn1	D SD1	E A-C	F E/D	G Tch 1	H C-G
1	Pride	4	8	2	0	0	0	1.86	0.64	2.15	0.66	-0.30	-0.45	2.00	0.15
2	Discuss.n	5	9	0	0	0	0	1.64	0.48	1.23	0.42	0.41	0.98	1.00	0.23
3	Handling *	1	1	0	2	10	0	1.64	1.23	2.15	1.35	-0.51	-0.38	5.00	-2.85
4	Clarity	2	10	1	1	0	0	2.07	0.70	2.23	0.89	-0.16	-0.18	2.00	0.23
5	Prompt	3	6	5	0	0	0	2.14	0.74	2.23	0.70	-0.09	-0.13	2.00	0.23
6	Happy	10	4	0	0	0	0	1.29	0.45	1.46	0.50	-0.18	-0.35	2.00	-0.54
7	Control *	0	0	0	0	13	1	1.00	0.00	1.15	0.53	-0.15	-0.29	5.00	-3.85
8	Challenge	9	5	0	0	0	0	1.36	0.48	1.46	0.50	-0.10	-0.21	2.00	-0.54
9	Approachable	3	8	3	0	0	0	2.00	0.65	2.23	0.89	-0.23	-0.26	2.00	0.23
10	Related	3	6	4	1	0	0	2.21	0.86	2.69	0.72	-0.48	-0.66	3.00	-0.31
11	Enthuses	11	3	0	0	0	0	1.21	0.41	1.62	0.49	-0.40	-0.82	1.00	0.62
12	Humour	6	5	3	0	0	0	1.79	0.77	1.85	0.86	-0.06	-0.07	2.00	-0.15
13	Impt Pts	9	5	0	0	0	0	1.36	0.48	1.31	0.46	0.05	0.11	2.00	-0.69
14	Strict *	1	4	9	0	0	0	3.43	0.62	3.23	0.70	0.20	0.28	2.00	1.23
15	Disrupt. Std	0	0	2	8	4	0	4.14	0.64	3.92	0.92	0.22	0.24	5.00	-1.08
16	Asmt Fair	6	5	2	0	0	0	1.69	0.72	1.92	0.83	-0.23	-0.28	1.00	0.92
17	Mat Prep	11	3	0	0	0	0	1.21	0.41	1.31	0.61	-0.09	-0.15	1.00	0.31
18	Know Wld	10	3	1	0	0	0	1.36	0.61	1.62	0.74	-0.26	-0.35	1.00	0.62
19	Chall/Intrst	10	4	0	0	0	0	1.29	0.45	1.69	0.72	-0.41	-0.56	2.00	-0.31
20	Improve	9	4	1	0	0	0	1.43	0.62	1.54	0.75	-0.11	-0.15	2.00	-0.46
21	Asmt Crit	9	3	2	0	0	0	1.50	0.73	1.46	0.50	0.04	0.08	1.00	0.46
22	Hmwk Ovld *	0	0	4	9	0	0	2.31	0.46	2.38	0.74	-0.08	-0.10	6.00	-3.62
23	Try Hard	8	4	2	0	0	0	1.57	0.73	2.00	0.88	-0.43	-0.49	1.00	1.00
24	Boys Qns *	0	0	0	0	0	14	#####	#####	#####	#####	#####	#####	6.00	#####
25	Ignore Tch *	0	0	1	1	11	0	1.23	0.58	1.23	0.42	0.00	0.00	6.00	-4.77
26	Feel Gd	2	6	4	2	0	0	2.43	0.90	2.38	0.84	0.04	0.05	2.00	0.38
27	Org/Expct	5	7	2	0	0	0	1.79	0.67	1.62	0.62	0.17	0.27	1.00	0.62
28	Too Much	0	0	3	11	0	0	3.79	0.41	3.62	0.74	0.17	0.23	4.00	-0.38
29	Incr Think	8	5	1	0	0	0	1.50	0.63	1.38	0.49	0.12	0.24	2.00	-0.62
30	Esy Undrst	9	3	2	0	0	0	1.50	0.73	2.00	0.68	-0.50	-0.74	2.00	0.00
31	Gen Conc	8	5	1	0	0	0	1.50	0.63	1.69	0.61	-0.19	-0.32	1.00	0.69
32	No Hm Hlp *	0	0	0	1	13	0	1.07	0.26	1.31	0.61	-0.24	-0.39	1.00	0.31
33	Stdy Skls	6	7	1	0	0	0	1.64	0.61	2.08	0.83	-0.43	-0.52	1.00	1.08
34	Team	5	5	4	0	0	0	1.93	0.80	1.62	0.62	0.31	0.50	1.00	0.62
35	Adpts Ablty	3	6	4	1	0	0	2.21	0.86	2.08	0.83	0.14	0.17	2.00	0.08
36	Hmk Cpes	5	9	0	0	0	0	1.64	0.48	1.77	0.80	-0.13	-0.16	2.00	-0.23
37	St Disply	9	4	1	0	0	0	1.43	0.62	1.77	0.70	-0.34	-0.49	1.00	0.77
38	Inclusive	10	4	0	0	0	0	1.29	0.45	1.54	0.84	-0.25	-0.30	1.00	0.54
39	Get Job	2	6	4	1	1	0	2.50	1.05	2.46	1.01	0.04	0.04	2.00	0.46
40	S Ask Qns	11	3	0	0	0	0	1.21	0.41	1.38	0.74	-0.17	-0.23	1.00	0.38
41	Motivates	11	2	1	0	0	0	1.29	0.59	1.69	0.91	-0.41	-0.45	1.00	0.69
42	Difficulty *	1	0	3	4	6	0	2.00	1.13	1.92	0.83	0.08	0.09	2.00	-0.08
43	Cmnts Hlp	5	9	0	0	0	0	1.64	0.48	1.92	0.62	-0.28	-0.46	2.00	-0.08
44	On Time	11	3	0	0	0	0	1.21	0.41	1.46	0.50	-0.25	-0.50	2.00	-0.54
45	Speed *	0	0	7	5	2	0	2.36	0.72	2.38	0.84	-0.03	-0.03	4.00	-1.62
46	Pronounce	11	2	1	0	0	0	1.29	0.59	1.50	0.65	-0.21	-0.33	1.00	0.50
47	Cont Nxt Yr	5	3	2	0	1	3	2.00	1.21	1.90	0.83	0.10	0.12	2.00	-0.10
48	Examples	11	3	0	0	0	0	1.21	0.41	1.23	0.42	-0.02	-0.04	2.00	-0.77
49	Fin Early *	0	0	2	8	4	0	1.86	0.64	1.67	0.62	0.19	0.31	4.00	-2.33
50	Talk A Lot *	0	0	2	6	6	0	1.71	0.70	1.46	0.63	0.25	0.40	4.00	-2.54
51	Bst Tchr	11	2	0	1	0	0	1.36	0.81	1.38	0.74	-0.03	-0.04		1.38
52	Bst Class	9	3	1	1	0	0	1.57	0.90	1.69	0.99	-0.12	-0.12		1.69
53	Worthwhile	4	7	3	0	0	0	1.93	0.70						

54	Felt Good	5	9	0	0	0	0	1.64	0.48					
55	Thanked	10	4	0	0	0	0	1.29	0.45					
56	Discussed	13	1	0	0	0	0	1.07	0.26					
57	Tried Impr	5	4	5	0	0	0	2.00	0.85					
58	Did Impr	1	5	7	1	0	0	2.57	0.73					
59	Picked On *	0	0	1	4	9	0	1.43	0.62					
60	Do For All	9	2	1	1	1	0	1.79	1.26					

Sex -->

Male	0
Female	14
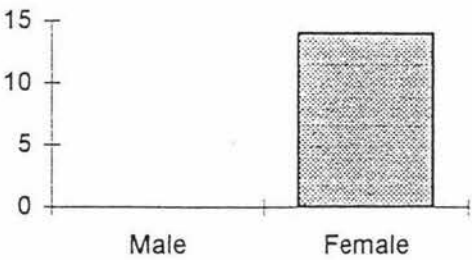
Ethnic -->

European	14
Maori	0
Pacific Is	0
Asian	0
Other	0
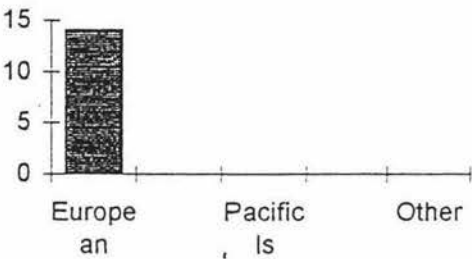
Ability -->

Top	3
Middle	10
Bottom	0
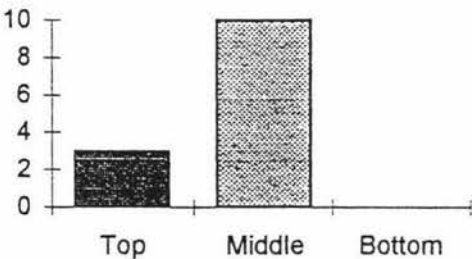
		A01			A02			B03			C05		
		t	d f	Sig	t	d f	Sig	t	d f	Sig	t	d f	Sig
1	Pride	0.12	42		-0.39	44		-0.50	36		0.67	44	
2	Discuss.n	-0.19	42		-2.29	44	*	-0.15	36		0.47	44	
3	Handling *	1.34	39		0.53	44		-0.25	36		2.42	44	*
4	Clarity	-1.53	42		-2.94	44	**	-2.06	35	*	-0.54	44	
5	Prompt	1.01	42		-0.10	44		-1.54	35		-2.37	44	*
6	Happy	0.52	42		-0.17	44		0.26	35		0.18	44	
7	Control *	0.74	41		0.33	44		-0.65	36		0.96	44	
8	Challenge	-1.63	42		-0.69	44		0.14	35		1.43	44	
9	Approachable	-1.28	42		-3.61	44	**	-0.31	36		-2.82	44	**
10	Related	-2.51	42	*	-1.86	44		-0.80	36		-1.82	44	
11	Enthuses	-2.17	42	*	-2.27	44	*	-0.63	36		-3.13	44	**
12	Humour	-1.08	42		-1.79	44		-0.37	36		-1.53	44	
13	Impt Pts	0.64	41		-1.91	43		-0.51	36		-0.49	44	
14	Strict *	-0.01	41		0.38	44		0.00	36		-0.61	44	
15	Disrupt. Std	0.71	42		0.73	44		-0.12	36		0.54	44	
16	Asmt Fair	0.60	41		-0.39	44		0.93	35		-0.79	44	
17	Mat Prep	1.33	42		-1.05	43		-1.12	35		-0.78	44	
18	Know Wld	-2.14	41	*	-0.80	43		-0.54	35		-1.51	44	
19	Chall/Intrst	-0.33	42		-0.59	44		-0.37	36		-0.43	44	
20	Improve	-1.70	42		-1.76	44		-0.85	35		-1.26	44	
21	Asmt Crit	-0.89	39		-2.84	43	**	0.22	35		-0.86	44	
22	Hmwk Ovld *	0.32	42		2.00	44		-0.12	36		1.60	44	
23	Try Hard	0.45	42		-1.14	44		-0.37	35		0.10	44	
24	Boys Qns *	1.28	40		0.45	43		-0.84	36		1.92	44	
25	Ignore Tch *	0.30	42		0.70	44		-1.62	36		1.37	44	
26	Feel Gd	0.69	42		-1.75	44		-1.82	36		-0.94	44	
27	Org/Expct	0.08	42		-1.84	42		-0.39	35		-0.18	44	
28	Too Much	2.09	42	*	-0.59	43		-0.59	36		1.09	44	
29	Incr Think	0.25	41		-3.11	43	**	-0.30	36		-1.69	44	
30	Esy Undrst	-1.03	42		-3.14	43	**	-0.53	36		-0.98	44	
31	Gen Conc	-1.62	41		-1.29	43		-1.18	35		-2.18	43	*
32	No Hm Hlp *	0.09	42		1.76	43		0.25	36		1.11	44	
33	Stdy Skls	-3.07	42	*	-2.39	42	*	0.77	36		-0.27	44	
34	Team	0.07	41		-4.38	43	**	-2.15	36	*	0.27	44	
35	Adpts Ablty	-0.74	41		-2.26	43	*	-0.57	36		-1.94	44	
36	Hmk Cpes	0.43	42		-0.60	43		-0.37	34		0.71	44	
37	St Disply	-6.25	41	**	-1.27	43		1.36	35		1.18	44	
38	Inclusive	-0.33	42		-1.06	43		-0.46	35		-0.11	44	
39	Get Job	-0.22	42		-0.86	43		-0.93	36		-0.80	44	
40	S Ask Qns	-0.67	42		-3.10	43	**	-1.45	36		-1.28	44	
41	Motivates	-2.02	42	*	-2.72	43	**	-0.44	36		-1.24	44	
42	Difficulty *	-0.55	42		1.60	43		-1.34	36		-0.81	44	
43	Cmnts Hlp	-0.42	42		-1.06	43		-1.14	36		-0.58	44	
44	On Time	-0.31	42		0.16	43		-0.44	36		-2.19	44	*
45	Speed *	0.27	42		-1.28	42		0.00	36		0.07	44	
46	Pronounce	-0.39	41		-1.25	43		0.30	36		0.11	44	
47	Cont Nxt Yr	0.48	42		-0.34	43		0.13	36		-0.53	44	
48	Examples	-0.98	42		-1.80	43		1.50	35		-0.75	44	
49	Fin Early *	0.19	42		0.11	43		-0.52	35		-1.17	44	
50	Talk A Lot *	-0.72	42		0.07	43		0.91	36		-0.52	44	
51	Bst Tchr	-2.39	42	*	-2.45	43	*	-1.08	36		-2.24	44	
52	Bst Class	-0.11	42		-2.10	43	*	-0.26	36		-1.35	44	
		* p<0.05			** p<0.01								

		C06			D07			D08			E09		
		t	d f	Sig	t	d f	Sig	t	d f	Sig	t	d f	Sig
1	Pride	-1.20	36		-0.65	27		1.15	29		-1.10	36	
2	Discuss.n	0.40	36		-1.88	27		0.21	29		0.47	36	
3	Handling *	2.71	35	*	-0.07	27		-1.41	29		0.38	36	
4	Clarity	-0.60	36		-0.70	26		2.89	29	**	-0.20	36	
5	Prompt	-0.59	36		-1.53	26		-2.04	29		-0.30	34	
6	Happy	0.69	35		0.43	27		-0.02	29		-0.29	36	
7	Control *	1.41	36		0.51	27		-2.52	29	*	0.38	36	
8	Challenge	1.16	36		-2.14	27	*	-0.94	29		-0.47	35	
9	Approachable	-0.96	36		0.00	27		2.02	28	*	-0.69	36	
10	Related	-1.68	36		0.13	27		-1.16	29		-1.40	36	
11	Enthuses	1.15	36		0.28	27		-0.75	29		-1.04	36	
12	Humour	-0.21	36		-0.25	27		-0.95	29		-1.68	36	
13	Impt Pts	0.12	36		0.12	27		-1.14	29		-0.90	36	
14	Strict *	2.95	36	**	-2.17	26	*	0.70	29		1.40	36	
15	Disrupt. Std	-0.87	36		-0.66	27		1.00	29		-2.16	36	*
16	Asmt Fair	1.31	36		-1.34	27		0.75	29		-1.65	36	
17	Mat Prep	-0.68	35		-1.05	27		-0.07	29		-0.69	36	
18	Know Wld	0.11	36		-1.43	27		-0.78	29		-1.33	36	
19	Chall/Intrst	0.25	36		-0.50	26		0.32	29		0.39	35	
20	Improve	-0.29	36		-0.94	27		1.17	29		-0.46	36	
21	Asmt Crit	1.56	36		-0.09	26		-2.12	28	*	-0.32	36	
22	Hmwk Ovld *	1.49	34		-0.79	26		-0.26	29		2.83	36	**
23	Try Hard	-0.80	36		1.45	27		0.97	29		-0.97	36	
24	Boys Qns *	-0.92	36		NA	NA		NA	NA		-1.64	36	
25	Ignore Tch *	-0.11	35		0.16	27		-1.85	29		-0.28	36	
26	Feel Gd	-0.57	36		-0.87	27		0.48	29		-1.08	36	
27	Org/Expct	-0.28	36		-1.18	27		-0.94	29		0.12	36	
28	Too Much	1.24	35		0.06	27		-1.30	29		-0.72	36	
29	Incr Think	-0.12	35		-1.66	27		-0.77	29		-0.94	36	
30	Esy Undrst	-0.35	35		-1.04	27		0.55	29		-1.38	36	
31	Gen Conc	-1.05	34		-1.34	27		0.29	29		-1.32	36	
32	No Hm Hlp *	0.52	34		0.93	27		-2.88	29	**	0.67	36	
33	Stdy Skls	-0.32	34		0.66	26		0.38	29		0.20	36	
34	Team	1.28	34		0.97	27		-1.33	29		0.05	36	
35	Adpts Ablty	-0.85	34		-0.04	27		-0.55	29		0.00	36	
36	Hmk Cpes	0.41	33		-1.55	27		0.92	29		0.38	36	
37	St Disply	-1.72	34		0.33	27		2.50	29	*	-0.78	36	
38	Inclusive	2.52	34	*	0.04	27		-0.78	29		0.42	36	
39	Get Job	0.84	33		0.85	27		-0.56	29		0.34	36	
40	S Ask Qns	0.05	34		-0.34	27		-0.23	29		-0.51	36	
41	Motivates	-0.38	34		-0.03	27		0.67	29		-0.42	36	
42	Difficulty *	0.29	33		0.66	27		-2.05	29	*	-0.61	36	
43	Cmnts Hlp	0.72	34		-0.51	27		-0.51	29		-0.80	36	
44	On Time	0.04	34		-0.55	27		0.63	29		1.02	36	
45	Speed *	0.46	32		1.39	26		-1.68	29		1.27	36	
46	Pronounce	-0.09	34		0.00	26		-0.64	29		-0.76	36	
47	Cont Nxt Yr	1.43	33		0.12	27		0.26	29		0.80	36	
48	Examples	0.85	34		-1.23	27		-0.03	29		-0.43	36	
49	Fin Early *	2.10	34	*	1.31	27		-0.97	29		-0.52	36	
50	Talk A Lot *	-0.07	34		-0.58	27		-1.17	29		-0.26	36	
51	Bst Tchr	-2.17	34		0.18	27		2.10	29		-0.24	35	
52	Bst Class	-1.41	34		-1.20	27		1.80	29		-1.70	36	
		* p<0.05			** p<0.01								

		E10			F11			F12			G14		
		t	d f	Sig	t	d f	Sig	t	d f	Sig	t	d f	Sig
1	Pride	-0.81	34		0.03	47		-1.35	37		-0.38	23	
2	Discuss.n	-0.25	34		-0.05	47		-1.48	37		-1.40	23	
3	Handling *	0.45	34		-1.47	47		-0.26	37		0.26	23	
4	Clarity	0.21	33		0.50	47		0.28	37		0.56	23	
5	Prompt	0.72	34		0.67	47		0.19	37		-0.80	23	
6	Happy	1.02	34		-0.22	47		0.12	37		-1.03	23	
7	Control *	-0.45	34		-1.25	47		-0.33	37		1.47	23	
8	Challenge	-0.08	34		-0.64	47		-0.89	37		-0.89	23	
9	Approachable	-0.07	34		-0.51	47		-1.24	37		-1.86	23	
10	Related	-2.93	34	**	-0.76	47		1.13	37		-1.72	23	
11	Enthuses	-0.96	34		0.27	47		0.35	36		0.67	23	
12	Humour	0.27	34		1.03	47		-0.26	37		-0.79	23	
13	Impt Pts	-0.10	34		0.78	47		-2.04	37	*	0.15	23	
14	Strict *	1.76	34		-0.43	47		-0.04	35		0.34	23	
15	Disrupt. Std	-0.21	34		-1.89	47		0.71	37		0.09	22	
16	Asmt Fair	0.38	34		0.34	47		-0.82	36		-0.05	23	
17	Mat Prep	-0.27	34		0.45	47		0.21	37		0.02	23	
18	Know Wld	-0.02	34		-0.82	47		-0.71	36		0.51	23	
19	Chall/Intrst	-0.39	34		0.17	47		-0.76	37		-0.10	23	
20	Improve	-1.07	34		0.91	47		0.24	37		-1.13	23	
21	Asmt Crit	-1.18	34		-1.50	47		-2.06	36	*	0.10	23	
22	Hmwk Ovld *	-1.53	34		-0.31	47		-0.60	35		-0.47	23	
23	Try Hard	0.25	34		-0.19	47		0.01	37		-0.15	23	
24	Boys Qns *	0.14	33		0.86	47		-0.10	37		-0.93	23	
25	Ignore Tch *	0.12	34		0.03	47		1.29	37		-0.44	23	
26	Feel Gd	0.22	34		0.37	47		-0.26	37		-0.05	23	
27	Org/Expct	-0.93	34		-0.31	47		-0.36	36		-0.06	23	
28	Too Much	0.26	34		-1.07	47		-0.28	37		0.89	23	
29	Incr Think	0.77	34		-0.71	47		-1.90	36		0.67	23	
30	Esy Undrst	0.05	34		2.29	47	*	-0.80	37		-0.65	22	
31	Gen Conc	0.09	34		-0.68	47		-0.43	37		-1.40	22	
32	No Hm Hlp *	0.47	34		0.06	47		0.18	37		0.67	23	
33	Stdy Skls	0.38	34		-0.07	47		-0.85	35		0.67	23	
34	Team	0.56	33		0.33	47		-0.84	37		-2.10	23	*
35	Adpts Ablty	-1.22	34		-0.45	47		-0.47	36		0.07	23	
36	Hmk Cpes	0.96	34		-1.90	47		1.06	37		-0.82	23	
37	St Disply	-2.38	34	*	-0.84	47		0.19	36		-0.54	23	
38	Inclusive	-0.75	34		1.68	47		0.91	37		-1.76	23	
39	Get Job	0.26	34		0.38	47		-0.45	37		0.08	23	
40	S Ask Qns	-1.94	34		1.71	47		-0.21	36		-1.64	23	
41	Motivates	-0.88	34		0.73	47		-0.26	36		-0.29	23	
42	Difficulty *	-0.72	34		-1.50	47		1.33	37		0.03	23	
43	Cmnts Hlp	-0.03	34		0.81	47		-0.58	37		0.04	23	
44	On Time	1.88	34		0.33	47		0.36	37		1.13	23	
45	Speed *	-0.23	34		-1.43	47		0.46	36		0.41	23	
46	Pronounce	0.94	34		0.12	47		0.90	37		0.28	23	
47	Cont Nxt Yr	-0.65	34		-0.07	47		0.90	37		-0.03	23	
48	Examples	0.86	33		0.47	47		-0.47	36		-0.31	23	
49	Fin Early *	0.39	34		-0.59	47		-0.68	36		0.41	23	
50	Talk A Lot *	-0.55	34		-0.68	47		0.16	36		0.44	23	
51	Bst Tchr	-0.45	34		-0.20	47		-0.80	37		-0.44	23	
52	Bst Class	0.23	34		0.34	47		-0.76	37		-0.03	23	
		* p<0.05			** p<0.01								

		I17			I18			J19			J20		
		t	d f	Sig	t	d f	Sig	t	d f	Sig	t	d f	Sig
1	Pride	1.79	42		0.06	44		-0.24	35		-1.14	25	
2	Discuss.n	0.40	42		-0.31	44		-0.87	35		2.28	25	*
3	Handling *	-0.34	41		-0.75	44		0.41	35		0.99	25	
4	Clarity	0.37	42		0.84	43		-0.70	35		-0.50	25	
5	Prompt	2.09	42	*	1.54	43		0.28	35		-0.30	25	
6	Happy	1.10	42		1.82	43		0.28	35		-0.93	25	
7	Control *	-1.16	41		-1.20	44		-1.44	35		1.00	24	
8	Challenge	-0.31	42		1.92	44		-0.24	35		-0.53	25	
9	Approachable	0.13	42		0.91	44		-0.01	35		-0.74	25	
10	Related	-0.77	42		-0.08	44		0.79	35		-1.50	25	
11	Enthuses	0.17	41		0.16	44		-0.07	35		-2.23	25	*
12	Humour	-0.03	41		0.95	44		-0.66	35		-0.18	25	
13	Impt Pts	0.56	41		0.52	43		-0.57	35		0.26	25	
14	Strict *	1.05	42		0.18	44		0.49	35		-0.75	25	
15	Disrupt. Std	-0.91	42		-1.14	44		-0.92	35		0.70	25	
16	Asmt Fair	0.68	42		0.27	44		1.52	35		-0.73	24	
17	Mat Prep	-1.28	41		1.30	44		-0.37	35		-0.45	25	
18	Know Wld	0.34	42		-0.33	43		-0.56	35		-0.96	25	
19	Chall/Intrst	0.16	42		0.08	44		0.10	35		-1.70	25	
20	Improve	1.19	42		0.38	43		-1.28	35		-0.40	25	
21	Asmt Crit	0.29	42		-0.52	43		-1.00	35		0.15	25	
22	Hmwk Ovld *	-0.53	42		-1.42	44		-1.00	34		0.31	24	
23	Try Hard	1.41	42		0.25	44		-0.03	35		-1.33	25	
24	Boys Qns *	NA	NA		NA	NA		NA	NA		NA	NA	
25	Ignore Tch *	-1.35	42		-0.87	43		1.26	35		0.00	24	
26	Feel Gd	-0.68	42		0.57	44		-0.04	35		0.13	25	
27	Org/Expct	0.10	42		0.53	43		-0.13	35		0.65	25	
28	Too Much	-0.80	42		0.33	44		-0.54	35		0.72	25	
29	Incr Think	-0.85	42		-0.37	44		-0.05	35		0.51	25	
30	Esy Undrst	0.57	42		-0.73	44		0.03	35		-1.77	25	
31	Gen Conc	-0.31	42		0.95	44		1.27	35		-0.78	25	
32	No Hm Hlp *	-0.21	42		-0.57	44		0.07	35		1.28	25	
33	Stdy Skls	0.38	40		-1.15	44		0.03	35		-1.50	25	
34	Team	0.65	42		0.72	44		-0.92	35		1.09	25	
35	Adpts Ablty	-0.88	42		-0.58	44		0.56	35		0.41	25	
36	Hmk Cpes	1.67	42		0.66	44		0.39	35		-0.48	25	
37	St Disply	0.95	42		3.01	44	**	1.59	34		-1.29	25	
38	Inclusive	-1.49	42		1.15	44		0.41	35		-0.94	25	
39	Get Job	0.67	41		0.97	44		0.91	35		0.09	25	
40	S Ask Qns	-0.75	42		-0.14	44		1.33	35		-0.72	25	
41	Motivates	0.23	41		-0.79	44		-0.39	35		-1.34	25	
42	Difficulty *	-1.55	42		-0.49	44		0.48	35		-0.19	25	
43	Cmnts Hlp	0.39	42		0.80	44		-0.41	35		-1.28	25	
44	On Time	1.72	42		1.86	44		-0.15	35		-1.36	25	
45	Speed *	-1.41	42		-1.13	44		-1.77	35		0.09	25	
46	Pronounce	0.00	42		1.19	44		0.22	35		-0.85	24	
47	Cont Nxt Yr	0.61	42		0.80	44		-0.34	34		0.21	19	
48	Examples	-1.47	42		0.93	43		0.90	33		-0.10	25	
49	Fin Early *	-1.25	42		1.44	44		0.53	35		-0.74	24	
50	Talk A Lot *	-2.47	41	*	-0.17	44		-0.22	35		-0.94	25	
51	Bst Tchr	-0.19	41		0.19	44		-0.37	35		-0.09	25	
52	Bst Class	-0.11	42		0.07	44		-0.18	35		-0.32	25	
		* p<0.05			** p<0.01								

			Positive	Negative	Dim Pos	Dim Neg	Pos %	Neg %
Attitude	1	Pride	7	0				
2.81%	23	Try Hard	3	2				
	26	Feel Gd	11	0				
	47	Cont Nxt Yr	0	0	21	2	4.42	0.58
Utility	10	Related	0	0				
0.12%	39	Get Job	1	0	1	0	0.21	0.00
Workld	28	Too Much	8	23				
10.74%	35	Adpts Ablty	0	5				
	42	Difficulty *	13	5				
	45	Speed *	11	23				
	49	Fin Early *	0	0	32	56	6.74	16.28
Org/Clar	4	Clarity	1	5				
8.30%	13	Impt Pts	1	5				
	17	Mat Prep	4	3				
	27	Org/Expct	1	1				
	30	Esy Undrst	26	19				
	44	On Time	2	0	35	33	7.37	9.59
Oth Stud	2	Discuss.n	8	5				
18.56%	6	Happy	90	11				
	34	Team	18	12				
	40	S Ask Qns	1	7	117	35	24.63	10.17
Rapport	3	Handling *	0	0				
1.83%	9	Approachable	4	1				
	24	Boys Qns *	0	0				
	31	Gen Conc	3	1				
	37	St Disply	2	4				
	38	Inclusive	0	0				
	46	Pronounce	0	0	9	6	1.89	1.74
Homewk	22	Hmwk Ovld *	6	19				
3.30%	32	No Hm Hlp *	0	0				
	36	Hmk Cpes	0	2	6	21	1.26	6.10
Enthusm	11	Enthuses	7	1				
6.47%	12	Humour	17	3				
	41	Motivates	15	10	39	14	8.21	4.07
Lrng/Val	8	Challenge	0	0				
18.44%	18	Know Wld	10	0				
	19	Chall/Intrst	74	33				
	29	Incr Think	5	0				
	33	Stdy Skls	0	1				
	48	Examples	5	7				
	50	Talk A Lot *	0	16	94	57	19.79	16.57
Control	7	Control *	9	3				
9.77%	14	Strict *	13	8				
	15	Disrupt. Std	4	41				
	25	Ignore Tch *	0	2	26	54	5.47	15.70
Ass.mt	5	Prompt	0	7				
5.37%	16	Asmt Fair	1	5				
	20	Improve	9	12				
	21	Asmt Crit	0	3				
	43	Cmnts Hlp	0	7	10	34	2.11	9.88
Overall	51	Bst Tchr	33	10				
14.30%	52	Bst Class	52	22	85	32	17.89	9.30
					475	344	100.00	100.00

			Positive	Negative	Tot Pos	Tot Neg	Pos %	Neg %
Factor 1	3	Handling *	0	0				
14.50%	4	Clarity	1	5				
	9	Approachable	4	1				
	11	Enthuses	7	1				
	12	Humour	17	3				
	13	Impt Pts	1	5				
	30	Esy Undrst	26	19				
	31	Gen Conc	3	1				
	35	Adpts Ablty	0	5				
	38	Inclusive	0	0				
	40	S Ask Qns	1	7				
	48	Examples	5	7	65	54	13.7	15.7
Factor 2	8	Challenge	0	0				
20.60%	18	Know Wld	10	0				
	19	Chall/Intrst	74	33				
	20	Improve	9	12				
	29	Incr Think	5	0				
	33	Stdy Skls	0	1				
	41	Motivates	15	10	113	56	23.8	16.3
Factor 3	7	Control *	9	3				
3.70%	25	Ignore Tch *	0	2				
	49	Fin Early *	0	0				
	50	Talk A Lot *	0	16	9	21	1.9	6.1
Factor 4	5	Prompt	0	7				
2.00%	17	Mat Prep	4	3				
	44	On Time	2	0				
	46	Pronounce	0	0	6	10	1.3	2.9
Factor 5	28	Too Much	8	23				
10.10%	42	Difficulty *	13	5				
	45	Speed *	11	23	32	51	6.7	14.8
Factor 6	2	Discuss.n	8	5				
5.30%	34	Team	18	12	26	17	5.5	4.9
Factor 7	6	Happy	90	11				
18.40%	15	Disrupt. Std	4	41				
	21	Asmt Crit	0	3				
	27	Org/Expct	1	1	95	56	20.0	16.3
Not in	1	Pride	7	0				
Factors	10	Related	0	0				
11.10%	14	Strict *	13	8				
	16	Asmt Fair	1	5				
	22	Hmwk Ovld *	6	19				
	23	Try Hard	3	2				
	24	Boys Qns *	0	0				
	26	Feel Gd	11	0				
	32	No Hm Hlp *	0	0				
	36	Hmk Cpes	0	2				
	37	St Disply	2	4				
	39	Get Job	1	0				
	43	Cmnts Hlp	0	7				
	47	Cont Nxt Yr	0	0	44	47	9.3	13.7
Overall	51	Bst Tchr	33	10				
14.30%	52	Bst Class	52	22	85	32	17.9	9.3
			475	344	475	344	100	100

			Good	Cmnt	Worry	Good	Cmnt	Worry	Gd %	Cm %	Wr %
Attitude	1	Pride	8	1	2						
7.99	23	Try Hard	11	0	0						
	26	Feel Gd	4	0	3						
	47	Cont Nxt Yr	7	0	1	30	1	6	10.6	1.6	5.1
Utility	10	Related	2	2	3						
3.46	39	Get Job	9	0	0	11	2	3	3.9	3.2	2.6
Workld	28	Too Much	4	2	3						
9.07	35	Adpts Ablty	3	2	4						
	42	Difficulty *	3	1	2						
	45	Speed *	0	4	5						
	49	Fin Early *	6	0	3	16	9	17	5.7	14.3	14.5
Org/Clar	4	Clarity	7	0	5						
11.45	13	Impt Pts	8	0	0						
	17	Mat Prep	7	1	1						
	27	Org/Expct	8	1	1						
	30	Esy Undrst	6	1	2						
	44	On Time	2	2	1	38	5	10	13.4	7.9	8.5
Oth Stud	2	Discuss.n	7	2	2						
8.21	6	Happy	10	0	1						
	34	Team	5	2	0						
	40	S Ask Qns	5	2	2	27	6	5	9.5	9.5	4.3
Rapport	3	Handling *	5	0	1						
13.61	9	Approachable	6	2	4						
	24	Boys Qns *	3	0	0						
	31	Gen Conc	5	1	4						
	37	St Disply	4	1	4						
	38	Inclusive	9	1	1						
	46	Pronounce	10	1	1	42	6	15	14.8	9.5	12.8
Homewk	22	Hmwk Ovld *	0	6	4						
5.18	32	No Hm Hlp *	3	2	2						
	36	Hmk Cpes	6	0	1	9	8	7	3.2	12.7	6.0
Enthusm	11	Enthuses	8	0	4						
6.26	12	Humour	7	2	1						
	41	Motivates	2	0	5	17	2	10	6.0	3.2	8.5
Lrng/Val	8	Challenge	8	0	3						
13.39	18	Know Wld	6	1	0						
	19	Chall/Intrst	6	0	2						
	29	Incr Think	7	1	0						
	33	Stdy Skls	1	3	4						
	48	Examples	6	1	2						
	50	Talk A Lot *	6	2	3	40	8	14	14.1	12.7	12.0
Control	7	Control *	5	1	1						
5.83	14	Strict *	1	4	2						
	15	Disrupt. Std	1	1	5						
	25	Ignore Tch *	2	2	2	9	8	10	3.2	12.7	8.5
Ass.mt	5	Prompt	4	4	4						
12.10	16	Asmt Fair	12	0	0						
	20	Improve	5	1	4						
	21	Asmt Crit	7	1	4						
	43	Cmnts Hlp	4	2	4	32	8	16	11.3	12.7	13.7
Overall	51	Bst Tchr	7	0	2	7	0	2	2.5	0.0	1.7
3.46	52	Bst Class	5	0	2	5	0	2	1.8	0.0	1.7
						283	63	117	100.0	100.0	100.0

			Good	Cmnt	Worry	Good	Cmnt	Worry	Good%	Cmnt%	Worry%
Factor 1	3	Handling *	5	0	1						
25.30%	4	Clarity	7	0	5						
	9	Approachable	6	2	4						
	11	Enthuses	8	0	4						
	12	Humour	7	2	1						
	13	Impt Pts	8	0	0						
	30	Esy Undrst	6	1	2						
	31	Gen Conc	5	1	4						
	35	Adpts Ablty	3	2	4						
	38	Inclusive	9	1	1						
	40	S Ask Qns	5	2	2						
	48	Examples	6	1	2	75	12	30	26.5	19.0	25.6
Factor 2	8	Challenge	8	0	3						
12.70%	18	Know Wld	6	1	0						
	19	Chall/Intrst	6	0	2						
	20	Improve	5	1	4						
	29	Incr Think	7	1	0						
	33	Stdy Skls	1	3	4						
	41	Motivates	2	0	5	35	6	18	12.4	9.5	15.4
Factor 3	7	Control *	5	1	1						
7.10%	25	Ignore Tch *	2	2	2						
	49	Fin Early *	6	0	3						
	50	Talk A Lot *	6	2	3	19	5	9	6.7	7.9	7.7
Factor 4	5	Prompt	4	4	4						
8.20%	17	Mat Prep	7	1	1						
	44	On Time	2	2	1						
	46	Pronounce	10	1	1	23	8	7	8.1	12.7	6.0
Factor 5	28	Too Much	4	2	3						
5.20%	42	Difficulty *	3	1	2						
	45	Speed *	0	4	5	7	7	10	2.5	11.1	8.5
Factor 6	2	Discuss.n	7	2	2						
3.90%	34	Team	5	2	0	12	4	2	4.2	6.3	1.7
Factor 7	6	Happy	10	0	1						
8.60%	15	Disrupt. Std	1	1	5						
	21	Asmt Crit	7	1	4						
	27	Org/Expct	8	1	1	26	3	11	9.2	4.8	9.4
Not in	1	Pride	8	1	2						
Factors	10	Related	2	2	3						
25.50%	14	Strict *	1	4	2						
	16	Asmt Fair	12	0	0						
	22	Hmwk Ovld *	0	6	4						
	23	Try Hard	11	0	0						
	24	Boys Qns *	3	0	0						
	26	Feel Gd	4	0	3						
	32	No Hm Hlp *	3	2	2						
	36	Hmk Cpes	6	0	1						
	37	St Disply	4	1	4						
	39	Get Job	9	0	0						
	43	Cmnts Hlp	4	2	4						
	47	Cont Nxt Yr	7	0	1	74	18	26	26.1	28.6	22.2
Overall	51	Bst Tchr	7	0	2						
3.50%	52	Bst Class	5	0	2	12	0	4	4.2	0.0	3.4
						283	63	117	100	100	100

Responses of Teachers - Post Evaluation

1 The Student Questionnaire

a To what extent do you think that the student evaluation form covered the teaching job that you do in the classroom?

Comprehensively	*To some extent*	*Not at all*
81%	19%	0%

b Do you think that the student evaluation form was a valid way of evaluating you as a teacher?

Extremely valid	*Of some validity*	*Invalid*
41%	59%	0%

c What features of the student evaluation questionnaire did you like?

d What features of the student evaluation questionnaire did you dislike?

e What questions did you think were the most useful? (Use the numbers on the questionnaire form.)

f What questions did you think were the least useful? (Use the numbers on the questionnaire form.)

g "Students are not competent to evaluate certain aspects of teaching (such as the teacher's subject knowledge), and these matters should not be in a student evaluation questionnaire."

Fully agree	*Neither agree/disagree*	*Totally disagree*
27%	40%	33%

2 Written Feedback

a How useful was the written feedback you received in terms of planning for teacher development?

Very useful	*Of some use*	*No use*
41%	59%	0%

b How much credibility do you think student evaluation data has?

Very credible	*Some credibility*	*Not credible*
38%	56%	6%

From which class would you give more credibility to feedback?

Senior class	*Junior class*	*Any class*	*No class level*
33%	0%	67%	0%

c Which data did you find the more useful?

Numerical tables	*Verbatim written comments*	*Both*
25%	44%	31%

d To what extent do you think that the feedback reflected your teaching of the class concerned?

Accurately	*With some accuracy*	*Inaccurately*
47%	53%	0%

e Did you discuss the results of the evaluation with your class?

Yes	*No*
75%	25%

Do you think that this discussion helped?

Yes	*No*
100%	0%

f Did the information you received reinforce any particular approaches you have to teaching?

Considerably	*To some extent*	*Not at all*
37%	50%	13%

3 The Consultation

a   How useful was the consultation in helping you to benefit from the student evaluation data?

Very useful	*Of some use*	*Not useful*
66%	28%	6%

b How useful was the consultation in terms of receiving support?

Very useful	*Of some use*	*Not useful*
7%	72%	21%

c How useful was the consultation in terms of planning your action for teacher development?

| Very useful | Of some use | Not useful |
| 44% | 38% | 19% |

d The researcher was deliberately non-judgmental in the consultation, and sought to help you interpret the data and plan for improvement. To what extent do you think this approach is helpful.

| Very helpful | Of some help | Not helpful |
| 67% | 33% | 0% |

e In almost all cases you received the student evaluation data prior to the consultation so that you had time to consider it. How helpful was this?

| Very helpful | Of some help | Not helpful |
| 64% | 29% | 7% |

f At the conclusion of the consultation you received a written report covering the main points covered in the discussion. To what extent do you think that this was helpful?

| Very helpful | Of some help | Not helpful |
| 56% | 44% | 0% |

g · How important is it to complement student feedback with some form of collegial consultation?

| Very important | Of some importance | · Not important |
| 42% | 39% | 19% |

4 Teacher Development

a The purpose of the process was to help you improve some aspect(s) of your teaching using information from students about your teaching. To what extent do you think this purpose is worthwhile?

| Very worthwhile | Some worthiness | Not worthwhile |
| 75% | 25% | 0% |

b To what extent do you think that the student feedback has been a factor leading to improvements in your teaching?

| Very useful | Some use | Not useful |
| 33% | 60% | 7% |

c What factors influenced you to change your teaching practice and why?

d To what extent did you **try to improve** the identified aspects of your teaching?

| Very hard | To some extent | Not at all |
| 39% | 61% | 0% |

e To what extent do you think that you **actually did improve** those aspects of your teaching?

Considerably	*To some extent*	*Not at all*
23%	77%	0%

f To what extent do you think that the learning of your students has been enhanced by the student feedback process?

Greatly	*To some extent*	*Not at all*
5%	86%	9%

5 Your reactions

a Describe your reaction to having participated in this research project?

b Describe your reaction when you received the results of the student evaluation.

c To what extent did participation in this project enhance or constrain your teaching?

Greatly enhanced	*Enhanced a little*	*Constrained*
15%	85%	0%

d What impact has this appraisal had on your job satisfaction and morale?

Great impact	*Some impact*	*No impact*
25%	62%	13%

6 Evaluation of this project

a What are the benefits and disadvantages you think you received as a result of this project?

b What suggestions would you make to improve the process.

c To what extent do you think that this type of teacher evaluation has a part to play in the development of better teachers in schools?

Considerable	*Some*	*None*
78%	22%	0%

d To what extent do you think that the process would have been more effective if it had been supported by a training component?

Considerably	*Somewhat*	*Not at all*
44%	44%	12%

7 Teacher Appraisal

This research project sits within the context of teacher appraisal.

a What do you think is the purpose of teacher appraisal?

Teacher Development (Formative evaluation)   **50%**

Personnel Decisions (Summative evaluation)
       - hiring/firing, promotion, competency etc   **0%**

Both of these.   **50%**

b Does your school have an teacher appraisal scheme?

Yes   *No*
69%   **31%**

Is participation compulsory?

Yes   *No*
82%   **18%**

What is the purpose of that scheme?

Teacher Development (Formative evaluation)   **92%**

Personnel Decisions (Summative evaluation)
       - hiring/firing, promotion, competency etc   **0%**

Both of these.   **8%**

c As a result of your experience in this project, to what extent should student evaluations be incorporated in an appraisal scheme?

To a large extent   *To some extent*   *Not at all*
25%   **62%**   **13%**

Have you changed your mind about this since entering into this project?

Yes   *No*
19%   **81%**

d Considering other teacher development strategies you have been involved in, what features of this project do you like/prefer or dislike compared with the others?

e Teacher appraisal is a time consuming business.   To what extent do you think that the time demands of this form of teacher appraisal (ie student evaluations) are reasonable given the potential outcomes?

Very reasonable	Reasonable	Unreasonable
20%	80%	0%

f As part of the process you were offered the right to have a support person with you at the consultation. Do you think that this is a desirable feature of an appraisal system?

Very desirable	Desirable	Undesirable
7%	64%	29%

g   Imagine that your school has decided to introduce an appraisal scheme which uses student evaluations. List three features that you would want to include in, and three features you would want to exclude from the new scheme.

h To what extent should the results of student evaluations be available to course supervisors/HODs/senior management team?

Freely available	Available at your discretion	Unavailable
18%	64%	18%

i For the purpose of teacher development, how often do you think that a teacher should collect and act on student evaluations?

Every term	Every year	Every two years	Never
14%	64%	21%	0%

j From whom should this data be obtained?

All classes	A sample of classes	One class
37%	50%	13%

8 General comments

Feel free to make any comments that you wish about this project or about teacher appraisal in general.

In conclusion

**"Student feedback is the most important source
of feedback for improving teaching."**

To what extent do you agree with this statement?

Totally agree	19%
Agree	56%
Neither agree/disagree	13%
Disagree	13%
Totally disagree	0%