



Centre for Higher Education Research and Development

The University of Manitoba  
220 Sinnott Bldg., 70 Dysart Rd.  
Winnipeg, MB R3T 2N2

Phone: (204) 474-7025  
Fax: (204) 474-7607

email: [uts@UManitoba.CA](mailto:uts@UManitoba.CA)  
[http://www.umanitoba.ca/academic\\_support/uts](http://www.umanitoba.ca/academic_support/uts)

© UTS, March 1998



The University of Manitoba

# *Students Rating Teaching*



*How student feedback  
can inform your teaching*

Mark L. Lawall, Ph.D.  
Associate Director  
University Teaching Services



***Students Rating Teaching:***  
***How student feedback can inform your teaching***

---

<b>Introduction .....</b>	<b>2</b>
<b>1. Development and study of SEEQ .....</b>	<b>4</b>
Development of SEEQ .....	4
SEEQ research .....	6
<b>2. Research on your teaching .....</b>	<b>12</b>
Uses of student ratings: summative and formative evaluation .....	12
SEEQ and formative evaluation .....	13
Universal use of SEEQ, customized use of the results .....	16
<b>3. The case of Jacob Deere .....</b>	<b>18</b>
Introduction .....	18
The Scenario .....	19
Summary of SEEQ results .....	22
Comments from evaluations .....	25
Responding to Prof. Deere’s feedback .....	27
<b>4. In your own case .....</b>	<b>29</b>
Self Evaluation as a baseline .....	29
Analyzing student feedback .....	30
Choosing a focus and strategies for improvement .....	31
Assessing the success of your strategy .....	33
Recording and presenting your efforts .....	34
<b>5. Annotated list of further readings .....</b>	<b>35</b>

**Introduction**

---

Student ratings of instruction remain a flashpoint for faculty, students, administrators, and, not surprisingly, for researchers in higher education. The relevant bibliography is long, sometimes polemical, and never easy to sort out for the non-specialist. The dual aims of this booklet are, first, to provide faculty with a summary of the background history of research on student ratings and, second, to suggest strategies for using student feedback to improve teaching. Practice towards this second goal is offered by the case study in Chapter 3. For faculty who are interested in further research on student ratings, there is a very selective, annotated bibliography of major works in the field. Many details of the research have been omitted from the text of this booklet in the interest of providing a concise and accessible summary.

One guiding principle for this booklet is that feedback on teaching is helpful for any faculty member. Feedback from students, colleagues, and other sources lets you know how a class went and where changes might be needed or attempted. Exploration of teaching and efforts towards improvement are not the sole concern of professors in trouble. Indeed, the most dedicated and respected teachers may benefit most from the uses of student feedback described in this booklet. Even an award-winning teacher, revered by students and colleagues alike, can still find areas to improve. This booklet, therefore, should not be seen as applicable only to cases where remedial attention to teaching practices seems needed.

The fundamental practice described in this booklet is the development of a research design for studying one’s teaching that uses SEEQ feedback as *one of several* sources of evidence of teaching effectiveness. While SEEQ results can be the catalyst for developing this personal research, these results cannot entirely guide teaching development. This booklet suggests other possible sources of feedback and how to coordinate those sources into a coherent program of personal development. SEEQ feedback, with its proven reliability and validity, provides a useful platform for beginning these faculty development efforts.

One point should be clarified concerning the role of University Teaching Services in the adoption, implementation, and further uses of SEEQ at the University.

Although UTS advised the Senate Committee on Teaching in their deliberations, UTS is not involved in the implementation of SEEQ at the University. Furthermore, UTS is not involved in any way in the analysis of SEEQ data for summative evaluation. This booklet was written to answer the following question:

**Given the use of SEEQ on campus, how can SEEQ be used to improve teaching at The University of Manitoba?**

Collection of student ratings alone *will not result in any significant change*. Instead, consultation is needed; consistent attention to teaching practice is required; and, perhaps most importantly, faculty (and students) must become aware of the quality, usefulness, and potential difficulties of SEEQ data. This booklet is intended as a contribution to making more effective use of students' evaluations of teaching.

---

## Acknowledgments

---

Funds for the publication and free distribution of this booklet came from UTS and the Faculty Development Initiative Fund. The design and layout of the booklet are the result of countless hours of work by Linda Rzeszutek, Program Assistant, UTS and for her work, I am extremely grateful.

I am also very grateful to Prof. Raymond Perry, the director of Research for the Center for Higher Education Research and Development, for his comments on an earlier draft of this booklet. Any lingering errors are entirely my own responsibility. Finally, I thank the faculty members who worked through the case study and discussed the various issues of student ratings with me in the Fall of 1997.

---

## Development and Study of SEEQ

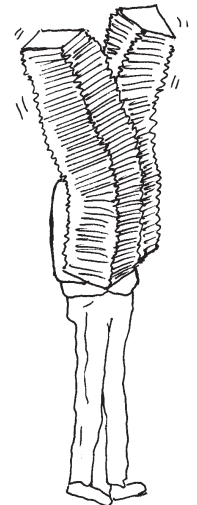
---

The Students' Evaluation of Educational Quality (SEEQ), recently adopted by The University of Manitoba, was developed by Herbert Marsh in the late 1970s and unveiled in 1982 in the *British Journal of Educational Psychology*. Marsh and others have used SEEQ data to explore the many issues that have characterized the past decades of student ratings research: reliability, validity, and stability of results; sources of bias in the responses; the utility of ratings in administrative decisions (summative evaluation); and the usefulness of the ratings for improving teaching (formative evaluation). Over the course of these studies, as in any academic field, debates and disagreements have surfaced concerning SEEQ. Some understanding of the development of SEEQ and these subsequent debates is needed for the most reasoned and effective use of student ratings feedback.

### *Development of SEEQ*

The development of SEEQ began with the question, "What is effective teaching?" Marsh and his colleagues (as well as their predecessors, contemporaries and successors) found considerable agreement in the idea that effective teaching is comprised of a definable set of independent elements. These elements include specific behaviours, such as "expressive speaking" or "highlights important points," and the general characteristics perceived by students from these behaviours, such as "instructor was dynamic and energetic" or "lectures facilitated notetaking." These elements are independent of one another in the sense that a student's judgment of a professor's organizational skills might differ greatly from that student's judgment of the professor's rapport with students.

Marsh began by trying to identify the various attributes or factors of good teaching and to define specific questions about those factors. Elements of



OK MR. MARSH, HERE ARE THE ANSWERS TO "WHAT IS GOOD TEACHING?"

effective teaching were culled from reviews of teaching research in higher education, existing faculty evaluation forms, interviews with students and with faculty, and student comments on open-ended questionnaires. The model of good teaching that resulted draws its strength from the range of sources used in its construction.

Nine general factors emerged from extensive statistical studies of how students rated the importance of items for evaluating teaching and how professors rated items' usefulness as feedback:

1. Learning/Value
2. Enthusiasm
3. Organization
4. Group Interaction
5. Individual Rapport
6. Breadth
7. Exams
8. Assignments
9. Workload

Subsequent studies (factor analyses) of thousands of SEEQ results — both from students and from professors' self-evaluations — show that these groupings of the specific questions' results best account for the variation in responses.

But surely teaching is too complex to be summarized by these nine variables! Perhaps, but studies of the sort used to develop SEEQ continually return to very similar elements of effective teaching. There are some attributes that have appeared in such studies that do not appear in SEEQ (e.g., instructors' pride in their positions at the given university; *Teaching Professor* 4(1) 1990, p. 1). Different groups may emphasize different attributes (e.g., students might emphasize a professor's helpfulness, while professors might place more emphasis on encouraging self-guidance; Feldman, *Research in Higher Education* 28(4) 1988, pp. 291-344). The areas, and even specific elements, of agreement among SEEQ and other studies suggest a strong consensus across the various constituent groups in higher education as to the major variables describing effective teaching.

## SEEQ Research

The bulk of research on student ratings of instruction has focused on determining the reliability, validity, and inherent biases in student ratings. Before reviewing some of these results, however, it is necessary to emphasize that the field of student ratings research is no different from any long-studied area of scholarly inquiry. As with any research topic, new ways of looking at the data are constantly developed and explored. The quality of research designs varies considerably, and the results are constantly reviewed and challenged by other practitioners in the field. The many years and pages of research results, however, do allow some certainty on a number of fundamental issues.

The *reliability* of a research tool may be defined as the consistency of responses to the same question on the same subject by different respondents. Studies of SEEQ results are particularly positive in this area. Correlations for individual instructors between class-average responses to particular items range from .95 for classes of 50 or more, .90 for 25 students, to .60 for five students. As found for other ratings tools, there is very high correlation between SEEQ class-average responses at end of term and alumni responses (.83). The belief that students only come to appreciate the quality of a professor's teaching long after graduation has been refuted by many studies of the reliability of student ratings. Study of individual professors' ratings over a 13-year time frame found no systematic relationship between year and results. Each instructor's ratings profiles (class average responses per item each year) showed consistency over time and were distinct from those of other instructors.

The *validity* of a student ratings tool requires a correlation between ratings and other indicators of effective teaching. Marsh focuses on the idea of "construct validity," which he defines as follows:

[Student Evaluations of Teaching (SETs)] are posited to be positively related to a wide variety of other indicators of effective teaching, and specific rating factors are required to be most highly correlated with variables to which they are most logically and theoretically related. (Marsh and Dunkin, p. 263)

Demonstration of the validity of responses to a given SEEQ item, therefore, depends on similar responses being gained from some other measure of teaching effectiveness related closely to that SEEQ item. Validity measures include tests of student learning,

professors' self-evaluations, and ratings from other classroom observers. Studies comparing student achievement test scores with teaching evaluations found significant correlations between amount of learning and quality of teaching in factors such as "teacher preparation and organization" and "clarity and understandableness." Factors such as "workload and difficulty" were unrelated to student achievement. Harry Murray, from The University of Western Ontario, studying the correlation between students' ratings of specific teaching behaviours and students' learning, found significant, positive correlations between specific teaching behaviours (echoing those included in SEEQ) and amount of studying, further courses taken in the field, perception of amount learned, and performance on final exams. Not every teaching behaviour equally affected each of these outcomes, but these research results show that defined factors of good teaching do result in, and are proven causes of, student learning (broadly defined and variously measured).

Other approaches to validity studies have provided more ambiguous results. This ambiguity, however, is often attributable to the questionable reliability or validity of the source being used for comparison. Comparisons of faculty self-evaluations and student ratings did result in significant correlations. Untrained peer observers' ratings do not tend to be correlated with students' ratings, probably due to infrequency of the peer observation and lack of training in observation. Much better matches were found between trained observers and student ratings for specific teaching behaviours. Peer ratings without actual observation have been found to echo student ratings. In this case, the peer raters probably form their opinion from informal conversations with students — with no guarantee that those students represent the consensus of class opinion. One other possible point of comparison is between research quality or academic rank and student ratings of teaching, but studies show no correlation in this regard.

The third major area of research has been the identification of sources of biases in student ratings. Often, this research considers demonstration of *bias* as a demonstration of correlation between student ratings and some external variable over which the instructor has no control (e.g., class size, student or instructor gender or age, time of day of the class, etc.). Marsh refines this definition as follows: "If a variable X legitimately influences the effectiveness of instruction and this influence is validly represented in SETs, then the influence of X should not be interpreted as a bias." (Marsh and Dunkin, pp. 285-286) For example, Marsh found that class size

correlated negatively with Group Interaction and Individual Rapport SEEQ factors. Class size is not considered by Marsh to be a bias because:

The specificity of the class size effect to dimensions most logically related to this variable, and the similarity of findings based on SETs and faculty self-evaluations argue that this effect is not a "bias" to SETs; rather, class size does have moderate effects on the aspects of effective teaching . . . to which it is most logically related and these effects are accurately reflected in the SETs. (Marsh and Dunkin, p. 288)



LARGE CLASSES CAN RESULT IN LOWER RATINGS FOR "GROUP INTERACTION" AND "INDIVIDUAL RAPPORT"!

Marsh does argue that it would then be unfair to compare ratings between instructors who taught classes of significantly different size. Similarly, students' prior interest in a topic affected the dimension of Learning/Value, again the expected dimension for this "external condition," so courses with different prior interest characteristics should be compared only with caution.

There are many potential biases that have been studied in past research. Many of these biases have crept into faculty lore as arguments against the use of student ratings. Although the refutation of these "myths" has been repeated in numerous publications, it seems worthwhile to review some common misconceptions here. (based in part on Aleamoni, 1987)

*Students are too immature, capricious, and inexperienced to give reliable feedback on teaching:* This idea is refuted by the reliability studies described above.

*Excellent records of research and publication predict teaching effectiveness and qualify professors to evaluate one another's teaching:* There is no consistent relationship between research and teaching quality, and some studies have argued that an inverse relationship might be expected. Peer evaluations are less reliable than student evaluations and are of unproven validity.

*Student ratings are just popularity contests:* Multidimensional surveys of teaching effectiveness clearly distinguish between enthusiasm, rapport, group interaction, etc., with the result that a high score in one category will not result in a high score in other categories. While certain personality types may be more naturally adapted to the conditions of certain classes, effective teaching in any situation is composed of learnable skills and behaviours.

*Students require time away from a course to appreciate the professor's quality:* Consistently strong correlations between alumni and current student ratings belie this myth. There may be elements of course content that students only appreciate later, but SEEQ primarily evaluates instructors, not courses.

*Age and gender of students and/or instructors influences ratings:* On this issue, the research shows inconsistent results suggesting that the effects are not systematic.

*GPA and academic level of students/rank of instructor affect ratings:* The research does not support this idea. Studies built on the expectation that academic rank is a condition against which to test student ratings validity are based on two unfounded assumptions: 1) that teaching quality is a primary factor in promotion and 2) that teaching quality necessarily increases with experience.

*Time of day the course is offered affects ratings:* The research does not support this hypothesis.

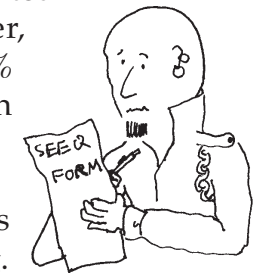
*Required courses are rated lower than electives:* Most research suggests this is true, though some found no significant differences. Some of the effects here may be a result of differences in "prior interest." Marsh found that such differences have an impact on those teaching dimensions where an effect is expected.

*Whether students are majors or non-majors affects ratings:* The research does not support this hypothesis.

*Level of course affects ratings:* The research shows mixed results. It seems likely in this case that other factors, such as class size, may affect the findings.

*Expected course grades, or lenient grading, influence ratings:* This is a very controversial topic; however, the majority opinion sees no significant biasing affect. A central difficulty in those studies that do see a significant effect is the lack of proof of "lenient grading" as opposed to widespread mastery of content in courses (the expected and legitimate result of good teaching).

Some *discipline-specific effects* appear to have resulted from extraneous variables (class-size, gender, required/elective, etc.), but these account for 1% to 14% of variation in overall instructor rating, with most disciplines showing 1 - 4% effect.



HMMM... I GOT AN 'A' ON THAT PAPER AND ON THE TEST... BUT I HAVEN'T REALLY LEARNED ANYTHING...

Research on the impact of these and other extraneous variables on student ratings of instructors is ongoing. Many new articles appear every year in journals on higher education in all fields. One difficulty in critically reviewing many such studies is the lack of consistency of the survey tool used in the research. If the tool is poorly constructed and of untested reliability and validity, then any research results should be considered with skepticism.

An issue that seems most affected by this ongoing research is whether it is valid to compare ratings among instructors in very different teaching contexts. This is an issue that affects any form of summative evaluation. Is a difference of .5 between two professors' average ratings significant? Should professors be compared against departmental or faculty norms? What are the proper weighting of responses to different factors of effectiveness in a final "score"? Summative use of multidimensional surveys is problematic, particularly in light of this last question. Marsh, himself, readily admits that there is little systematic study of the best use of SEEQ results for summative evaluation, but he adheres to his guiding principle that teaching is multidimensional and it should not be evaluated in other terms. These issues of summative evaluation cannot be resolved here, but they should be

taken up in informed discussion in the appropriate venues within the University.

One point is certain, despite the inevitable fluctuations in this field of research:

*For any individual instructor, teaching roughly the same sorts of courses from year to year, student ratings are an extremely reliable and valid measure of that instructor's teaching quality.* The ratings will not be changed significantly by a shift in the gender- or age-demographics of the class, a schedule change, or whether the instructor receives a promotion. For the purposes of formative evaluation such reliability, and hence comparability from year-to-year, is precisely what is needed. Student ratings of instruction, especially when collected using a factor-based tool such as SEEQ, provide a very useful source of feedback on one's teaching. They are not, however, the only available source. The combination of student ratings with other forms of feedback and assessment allows informed adjustments to one's teaching strategies.

## Research on your teaching

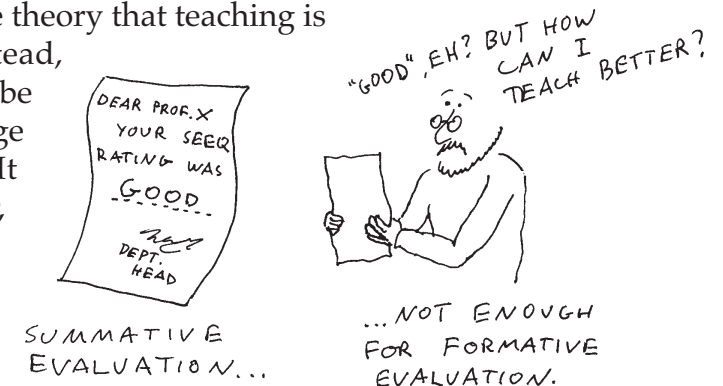
---

The preceding section summarized major trends in opinion about the quality of student evaluations as evidence of effective teaching. Now, we turn to consider the appropriate uses of student ratings in faculty evaluations and to identify other sources of evidence for developing an accurate evaluation.

### *Uses of student ratings: summative and formative evaluation*

There are two common uses for the information provided by student ratings. First, *summative evaluation* is aimed at making administrative decisions such as merit pay increases, teaching assignments, and tenure/promotion reviews. This booklet is more intended for the use of ratings for *formative evaluation*, in which the data collected are used to make improvements in teaching. Such improvements might eventually affect questions of tenure or promotion, but the primary intent in formative evaluation is faculty development.

These two different kinds of evaluation require different kinds of evidence. A summative evaluation seeks to identify differences in quality between faculty, and these differences must be based on the most reliable data available. The more general or global questions (SEEQ questions 30-32) are more reliable and better correlated with measures of student learning (one method of determining validity) than are specific factors. For summative use of student ratings, class averages of responses to global questions are most appropriate. Marsh disagreed with this approach, however, since it contradicts the theory that teaching is multidimensional. Marsh, instead, advocates that an overall rating be determined by a weighted average of the different components. It remains undetermined, however, which components should receive what weight in the final average. Cashin, who is otherwise a supporter of the



multidimensional approach, compares summative evaluation of teaching with evaluation of student success in a course: “If summative evaluation indicates that the course is effective, no further data are needed.” (Cashin & Downey 1992, pp. 564-5) A summative evaluation of teaching, therefore, should aim to determine whether (not necessarily how) the teaching was effective.

Global questions, however, are of little use in formative evaluation. Here, the “how” of teaching effectiveness is the central issue. Student feedback on specific teaching behaviours is a necessary component of formative evaluation. Although the reliability scores of such questions do not match those of global questions, and although there has been debate about the correlation between certain teaching dimensions and student learning quality; nevertheless, these specific questions are needed to highlight foci for further study. The possible problems with the student ratings data are, in this context, offset by their value in guiding development. If the data include errors, these errors will be illuminated and corrected by the further study. In other words, the student ratings provide an excellent guide for designing the teaching development process, but they are not used alone — other sources of feedback are also considered.

### *SEEQ and formative evaluation*

There are many characteristics of SEEQ feedback that correspond with characteristics of effective formative feedback as defined by the research of Robert Menges and Kathleen Brinko. The focus of their research was to determine the traits that best typified effective feedback on teaching for faculty members. This feedback was not limited to student ratings; videotapes, peer observers, consultations, and self-evaluation were all studied. In a recent article in *The Journal of Higher Education*, Brinko lists the following feedback characteristics with particular relevance to SEEQ results:

1) *the source should be seen as credible and knowledgeable*: One aim in the design of SEEQ was to gather student perceptions of instructors’ traits or behaviours the students observed every day. Students are rarely asked to judge those other characteristics of the instructor with which they have little or no experience.

2) *effective feedback includes specific, concrete data, focused on behaviours rather than personality*: In the SEEQ form, each teaching component is broken down into specific, behavioural segments for clarity.

3) *the feedback is relevant and meaningful*: Given the method described earlier used to develop SEEQ, it is likely that most instructors will agree that the components listed in the survey are related to effective teaching. As a result, the feedback on those questions should be considered a meaningful measure of teaching effectiveness.

4) *feedback is provided often but not too often*: Studies of SEEQ results’ stability over time for any individual professor (regardless of the class being taught) support the argument that ratings tools need not be implemented every year in every class. The most effective use of SEEQ is as part of a carefully designed program of teaching research.

By these criteria defined by Brinko, SEEQ is well-suited to playing a role in an effective program of feedback on teaching. And yet, the collection of student ratings alone rarely results in improved teaching.

The research on whether feedback from student ratings improves teaching consistently concludes that additional consultation about the results is a prerequisite for improvement. The consultation may come from various sources: colleague from within or outside your department, professional faculty development consultant (e.g., from UTS), or a trained faculty peer consultant. (*N.B.* Research in this area suggests that persons of higher authority (department heads or deans) are not effective consultants. They bring too much of a perceived threat, regardless of their intentions.) The consultant’s role is not to prescribe or dictate specific corrective action. Instead, the consultant guides the choice of strategies of improvement, ensures the implementation of selected strategies, and guides the process of gathering further feedback on those strategies. In other words, the consultant acts as a neutral assistant in the on-going process of researching your teaching using a wide range of sources.

There are many other sources of feedback on teaching to supplement student ratings. You might ask a colleague or member of the Peer Consultants group to observe a series of your classes. Regardless of who is doing the observing, you and your observer should plan to meet before and after each session to discuss goals, context, methods, assessment of effectiveness, etc. At many Universities, programs exist in which trained student observers can provide a student perspective on your teaching without the possible bias of being in your class. Your own structured reflection on your teaching, perhaps in the form of a teaching diary or development of a teaching portfolio, can



also help you identify areas to improve or areas where you see that specific, targeted assessment strategies are needed. Videotaping your teaching followed by discussion with a colleague can give you a chance to observe your own teaching. Finally, you can turn again to your students with less formal assessment methods aimed at determining whether a particular change in your teaching strategy was successful. These “Classroom Assessment Techniques” (CATs) can be simple surveys of opinion, brief written assignments, shared journals — there are many options listed in Angelo and Cross’ compendium (see Chapter 4, p. 33). CATs should be implemented as part of a planned cycle of developing new strategies, testing them, getting feedback, revising the strategy. More will be said about all of these strategies in Chapter 4’s discussion of specific ways to use student feedback.

It bears emphasis that these other sources of evidence on teaching have not been so thoroughly researched as have student ratings. They may be less reliable, more open to influence from extraneous variables, and less valid. And yet, these other sources can provide more specific, very useful feedback on issues that cannot be covered by SEEQ or other student ratings tools alone. Consider, for example, the SEEQ question “Instructor gave lectures that facilitated taking notes.” An instructor might make various changes in lecturing methods to improve this aspect of teaching. And yet, given the marginal variation in responses attributable to extraneous factors, possible changes in student learning-style demographics, etc., the instructor might be uncertain whether ratings in the next year indicate a substantive improvement in student note-taking. Therefore, to augment the data gathered from SEEQ, the instructor might ask to review students’ notes, might ask a colleague or student (from outside the class) to take notes as if they were students in the class, or might use simple questionnaires in class to learn whether main points were emerging from lectures. With careful use, in consultation with a colleague, this further evidence can indicate whether attempts at improvement are having a positive impact. These additional data may not be dependable enough for use in personnel decisions, but they are very useful in the process of formative feedback.

Furthermore, the additional sources of data provide an environment that is more conducive to effective feedback than the use of SEEQ results alone. The use of additional sources matches Brinko’s suggestion that information be gathered from a variety of sources and be considered as part of a process. In other words, a single attempt to gather feedback, at the end of a course, when there is no longer an opportunity to change that particular course, is not likely to engender attempts at

change. When the feedback is seen as a longer-term research project, the idea of experimentation and change becomes more familiar and desirable. Such a process-based view also satisfies Brinko’s suggestion that any process of evaluation include an opportunity for acknowledgment of, and response to, any criticisms. The use of SEEQ along with other forms of feedback also focuses the study of one’s teaching on specific issues, behaviours, or goals. While such targets are easily identified by the SEEQ format, refinement of those goals may require further modes of investigation. Finally, multiple sources of feedback can highlight points of disagreement or difference between the instructor’s view of a class’s goals or meaning and the students’ views. With too great a gap between these perspectives, the learning process suffers unnecessarily. A program of classroom research, based on multiple sources of evidence, can bring the perspectives of students and faculty closer together.

There is widespread agreement that the focus on specific teaching behaviours in SEEQ feedback is very beneficial for formative evaluation. The process of effective formative evaluation is certainly assisted by the data gathered by SEEQ. These data are focused on specific teaching behaviours and are explicitly tied to generally accepted elements of effective teaching. SEEQ can serve as an excellent starting point for formative evaluation. Other sources of data are useful additions to the overall cycle of data collection, analysis, and determination of strategies for improvement. Once formative evaluation of teaching becomes part of a continual process, instead of an end-of-term event, students should receive a clearer message, too, that efforts are constantly being made to improve the quality of their education.

#### *Universal use of SEEQ, customized use of the results*

Two points should be noted, in particular, from the foregoing advice on the use of SEEQ feedback as part of a broader study of teaching. First, teaching is multidimensional, so many sources are needed for a thorough study of one’s teaching. SEEQ can provide excellent data — at least for a starting point — but the time spent by an instructor and students in implementing the survey, and the time spent processing the data, will be time wasted if the data are not then incorporated into a program of classroom research. The second point to be drawn from the preceding comments is that both formative and summative uses of SEEQ data require a full consideration of the teaching context being evaluated. Such consideration may not be possible from SEEQ data alone, but there are many other sources available. An explicit consideration of evaluative criteria is a

prerequisite for any use of these sources in teaching evaluation. This general process of consulting various sources creates an individualized or customized process of teaching evaluation even in the context of universal use of SEEQ.

---

## The Case of Jacob Deere

---

### *Introduction*

The following case study is designed to give practice defining faculty development strategies in light of SEEQ numerical ratings and student comments. The case scenario and the student ratings are, of course, entirely fictional. The case was developed as follows. Characteristics of the professor's teaching methods — both inside the classroom and outside the classroom — were imagined first. These included characteristics that he, himself, would be likely to recognize as significant (these are included in the scenario), as well as those that he might not notice without prompting (these are reflected more in the student comments). Possible student reactions to his methods were then listed, based on student characteristics of a diverse population similar to that encountered in most 100-level courses at the U of M. Patterns of responses within the SEEQ data are based both on these fictional teacher and student characteristics and on patterns commonly noted in studies of SEEQ results.

The case centers on a predominantly lecture-based classroom in the Humanities. The process of consultation and feedback suggested in the case, however, is fully applicable to other teaching contexts and other disciplines. The more general process is described in the last chapter. Here, the aim is to provide practice working with student feedback, even if you are not at all familiar with the disciplinary context of the instructor. That said, there may be questions you would like to have answered by Prof. Deere. You should keep a list of these questions, but at the same time, you will want to be able to bring a series of preliminary observations and suggestions to your consultation session. Of course, the final decisions about which strategies to implement will emerge from that discussion (based, in part, on answers to your further questions). This case study provides a model for only the first stage of the process of classroom research.

One method for work on this case is as follows. First, read through the scenario, the SEEQ results (numbers and comments), and questions at the end. The scenario includes various details of Prof. Deere's teaching methods and style that do not emerge explicitly in the SEEQ data. You might want to make note of such clues from the scenario, noting especially those that relate to problem areas you identify

in the numerical ratings and comments. Then work through the questions at the end of the case, paying particular attention to the idea that your goal is to develop some specific strategies to address selected problems.

One point to bear in mind is that this exercise is primarily about the instructor, not his students. You might feel that the students, themselves, are not taking sufficient responsibility for their own learning. This may be true, but individual faculty members are likely to have very different views on this issue. In order to develop a constructive strategy for teaching development, it is best to focus on the behaviours of the instructor. This context is where you and the instructor have the most direct influence and control. Furthermore, changes in teaching on the part of the instructor might encourage change or development of students' learning behaviours.

### *The Scenario*

Prof. Jacob Deere is a casually dressed, white, male, 50-ish history professor. He is sitting in his dimly-lit office, tilted back in his well-worn, hardwood office chair, his sneakered feet perched precariously on the corner of his desk.

He has just received the course evaluation summaries and forms from his course "World History to 1500." He is a bit hesitant about all of this since this is the first year a new form is being used. It is the same form that he used in his upper level historiography seminar; his racquetball partner Max used the same form in his biology lab course. Even so, he is interested to see what the students thought of the course. This was the first year he required more reading of original historical texts. He had set aside "discussion days" to consider these primary sources in class at the end of each unit. Did they like the new format?

Jacob looked over the summary. The usual flood of numbers meant little to him except that in some categories the ratings were not very high. He noticed some disturbing trends. Many students seemed to think that he was poorly organized! That he did not give them ample time to participate in the class! That he was



inaccessible! How could this be? He had set aside time for discussions. But they never had anything to say; he always had to end up providing the answers to his questions. Those sessions had seemed so slow and tedious to him. He held three hours aside for office hours every week for that class alone. He would sit working away on grading or reading, waiting patiently for a knock on his office door. Was it his fault if the students could not find his office? A few of the older students in the class visited fairly regularly, and he thought the conversations helped them . . . but it was always the older students, rarely the first-year students. Clearly this new form was no more accurate and no more enlightening than the past versions.

He flipped through the original forms and scanned the comments. "I loved reading Aristophanes!", "There should be more primary sources", "Prof. Deer [sic] made history fun . . . [same paper] I had a hard time hearing the lecture 'cause of the airconditioner", "I came to University to learn from profs not listen to that guy in the front row ramble on about his own ideas!" That last comment clinched it: the new class format was a failure. Wait . . . that earlier comment did want more primary sources. It's always the same — one student contradicts another — this feedback form is no better than the old ones.

Somewhat disgruntled, he wandered out of his office and down the hall to the faculty lounge. Deere's colleague in the history department, Prof. Geoffrey Smythe, was drinking his usual noon can of Jolt. "You can tell he has five children," thought Jacob grimly. But Jacob could tell Geoffrey was troubled by something else too. "Did you get your student evaluations, too, Geoff?" he asked, filling his dirty mug with dirtier coffee.

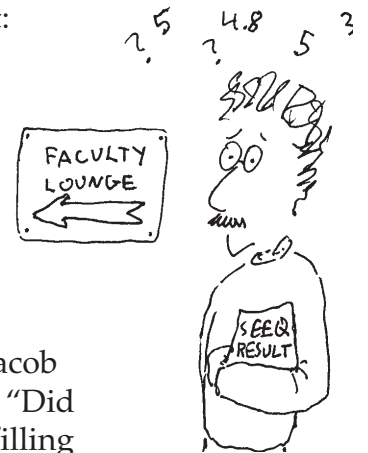
"Yes, the usual . . ."

"But these are published for the students!" interjected Jacob.

"Oh well, they don't read anything else we give them, why should they read these?"

Jacob felt moderately relieved. "OK, true enough," he agreed, "but what good is all this feedback anyhow, it's all a bunch of numbers . . . and the comments are all contradictory!"

"Hmmm . . . someone told me that this new form would help our teaching. I can't see how though . . ." Geoffrey scrambled through his mind to find a more



comfortable topic, "Say, I hear you just finished checking the proofs on your new book . . . that's great!"

"Yeah, funny how I had to send it to three different publishers before Oxford finally took it. Each reviewer contradicted the others! . . . just like these student comments. I mean how can we learn anything from these forms?"

Smythe could tell that Jacob was not going to let the topic drop so easily. "OK," he replied, "why don't you let me have a look over your forms — maybe a fresh pair of eyes will help sort things out — and we can get together early next week to discuss them."



Jacob finished his coffee, left the cup with its residue next to the sink, and wandered back to his office to retrieve the forms. The very idea of doing something with them was comforting, though he was unsure whether talking with Geoffrey would really help that much . . .

### Summary of SEEQ results

Faculty: **ARTS**  
 Department: **History**  
 Course: **World History to 1500**

Students enrolled: **60**  
 Students responding: **50**  
 % responding: **83%**

#### LEARNING

1. I have found the course intellectually challenging and stimulating. ....
2. I have learned something which I consider valuable. ....
3. My interest in the subject has increased as a consequence of this course. ....
4. I have learned and understood the subject materials of this course. ....

#### ENTHUSIASM

5. Instructor was enthusiastic about teaching the course. ....
6. Instructor was dynamic and energetic in conducting the course. ....
7. Instructor enhanced presentations with the use of humour. ....
8. Instructor's style of presentation held my interest during class. ....

#### ORGANIZATION

9. Instructor's explanations were clear. ....
10. Instructor's materials were well prepared and carefully explained. ....
11. Proposed objectives agreed with those actually taught so I knew where the course was going. ....
12. Instructor gave lectures that facilitated taking notes. ....

#### GROUP INTERACTION

13. Students were encouraged to participate in class discussions. ....
14. Students were invited to share their ideas and knowledge. ....
15. Students were encouraged to ask questions and were given meaningful answers. ....
16. Students were encouraged to express their own ideas and/or question the instructor. ....

	strongly disagree	disagree	neutral	agree	strongly agree	n/a
<b>LEARNING</b>						
1. I have found the course intellectually challenging and stimulating. ....		1	2	29	18	
2. I have learned something which I consider valuable. ....			12	30	8	
3. My interest in the subject has increased as a consequence of this course. ....		2	3	20	25	
4. I have learned and understood the subject materials of this course. ....		2	5	28	15	
<b>ENTHUSIASM</b>						
5. Instructor was enthusiastic about teaching the course. ....				17	33	
6. Instructor was dynamic and energetic in conducting the course. ....			4	16	30	
7. Instructor enhanced presentations with the use of humour. ....	10	5		15	20	
8. Instructor's style of presentation held my interest during class. ....			1	19	30	
<b>ORGANIZATION</b>						
9. Instructor's explanations were clear. ....		7	10	20	13	
10. Instructor's materials were well prepared and carefully explained. ....	7	5	15	15	8	
11. Proposed objectives agreed with those actually taught so I knew where the course was going. ....		2	16	12	8	12
12. Instructor gave lectures that facilitated taking notes. ....	9	18	3	13	7	
<b>GROUP INTERACTION</b>						
13. Students were encouraged to participate in class discussions. ....	7	10	10	18	5	
14. Students were invited to share their ideas and knowledge. ....	10	8	8	20	4	
15. Students were encouraged to ask questions and were given meaningful answers. ....	10	12	7	18	3	
16. Students were encouraged to express their own ideas and/or question the instructor. ....	10	5	8	22	5	

	<i>strongly disagree</i>	<i>disagree</i>	<i>neutral</i>	<i>agree</i>	<i>strongly agree</i>	<i>n/a</i>
<b>INDIVIDUAL RAPPORT</b>						
17. Instructor was friendly towards individual students. ....			10	17	23	
18. Instructor made students feel welcome in seeking help/ advice in or outside of class. ....	7	5	3	20	15	
19. Instructor had a genuine interest in individual students. ....		3	10	12	25	
20. Instructor was adequately accessible to students during office hours or after class. ....	5	10	2	23	10	
<b>BREADTH</b>						
21. Instructor contrasted the implications of various theories. ....			5	32	13	
22. Instructor presented the background or origin of ideas/ concepts developed in class. ....				12	38	
23. Instructor presented points of view other than his/her own when appropriate. ....				31	19	
24. Instructor adequately discussed current developments in the field. ....		1	23	16	10	
<b>EXAMINATIONS</b>						
25. Feedback on examinations/graded materials was valuable. ....	4	4	8	26	8	
26. Methods of evaluating student work were fair and appropriate. ....	6	8	10	20	6	
27. Examinations/graded materials tested course content as emphasized by the instructor. ....	7	6	12	20	5	
<b>ASSIGNMENTS</b>						
28. Required readings/texts were valuable. ....			4	16	30	
29. Readings, homework, laboratories contributed to appreciation and understanding of subject. ....		2	5	20	23	

	<i>very poor</i>	<i>poor</i>	<i>average</i>	<i>good</i>	<i>very good</i>
<b>OVERALL</b>					
30. Compared with other courses I have had at the U of M, I would say this course is: ....			5	28	17
31. Compared with other instructors I have had at the U of M, I would say this instructor is: ....	1	4	9	19	17
32. As an overall rating, I would say this instructor is: ....	1	4	7	20	18
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>STUDENT AND COURSE CHARACTERISTICS</b> (leave blank if no response applied)					
33. Course difficulty, relative to other courses, (1=very light, 2=light, 3=average, 4=difficult, 5=very difficult) ....		4	14	12	20
34. Course workload, relative to other courses was, (1=very light, 2=light, 3=average, 4=heavy, 5=very heavy) ....			16	24	10
35. Course pace was: (1=too slow, 2=slow, 3=about right, 4=fast, 5=too fast) ....		3	19	13	15
36. Hours per week required outside of class: (1=0 to 2, 2=2 to 5, 3=5 to 7, 4= 8 to 12, 5= over 12) ....		6	16	20	8
37. Level of interest in subject prior to this course: (1=very low, 2=low, 3=medium, 4=high, 5=very high). ....	3	4	10	17	16
38. Overall GPA at U of M. Leave blank if not yet established: (1=below 2.5, 2=2.5 to 3.0, 3=3.0 to 3.4, 4=3.4 to 3.7, 5=above 3.7) ....		5	3	3	1
39. Expected grade in the course: (1=F, 2=D, 3=C or C+, 4=B or B+, 5=A or A+) ....		2	12	23	13
40. Reason for taking this course. Select the one which is best: (1=required for major, 2=elective for major, 3=faculty requirement, 4=minor or related field, 5=general interest only) ....	9	4	4	3	30
41. Year in program: (1=first, 2=second, 3=third, 4=fourth, 5=pre-masters/graduate) ....	38	7	5		

### *Comments from evaluations*

(N.B. With these comments, the parenthetical notations indicate features of the accompanying numerical scores. If the notation does not mention a category, you should assume that the ratings were positive.)

They should get that A/C fixed! I could hardly hear anything! (low marks in organization and group interaction)

I got fairly lost pretty early in the class . . . I tried to get in to Prof. Deere's office hours, but the door was always shut. I just couldn't connect with this class. (low marks in organization and individual rapport, 1st yr. with low prior interest)

I loved reading Aristophanes!

There should be more primary sources. I really liked the way Prof. tied the readings of primary sources into the class in general. (neutral to low in group interaction)

Prof. Deer made history fun. I loved his jokes! But I had a hard time hearing the lecture 'cause of the airconditioner.

I came to University to learn from profs, not listen to that guy in the front row ramble on about his own opinions! (low in group interaction, one of those who wrote n/a for Organization/objectives)

Those guys in the front never let anyone else talk! Why should I listen to them? I pay to be taught! (low marks for group interaction)

I liked the discussion sessions — the source readings really made history come alive and I learned so much about writing from discussing my assignments with Prof. Deere. He really made my return to college easier!

Prof. Deere's lectures — especially about the primary sources were really full of information. I wish he'd slow down a bit though . . . All those new terms made it hard to get everything down. (gave generally low marks in organization and group interaction)

I liked the teacher's discussions of the primary readings. (gave generally low marks in organization and neutral for group interaction)

Very funny!! (neutral on learning and organization, low workload, low expected grade)

Dr. Deere joked around a lot, but I don't think they really added anything to the class . . . they kinda wasted time. And he would never let me talk . . . he only called on the people in the front. (low marks on enthusiasm/humor, group interaction and individual rapport)

I got pretty lost in the lectures . . . I just couldn't see where things were heading, or what was especially important. (low organization, neutral in breadth/various theories, low in examinations and assignments)

I really liked prof. Dear, his lectures really kept me awake . . . especially his lectures on the primary sources. (low in group interaction)

I was never really sure what Prof. Deere was looking for in our assignments. He gave us a LOT of stuff in the lectures, but he'd never say who was really right. (low in notetaking, but neutral in other organizational aspects, high in breadth, low in examinations and assignments)

This class was really hard. Why couldn't we have multiple choice tests like in my other courses. This was way harder than a first year course should be. (high in difficulty, workload, but only average in hours per week; low in organization, examinations, assignments)

I liked the fact that Dr. Deere told us what a lot of different people thought about different issues. Every topic turned into a mini-debate in my notes! (3rd year, history major)

I had a hard time taking notes in class and so I never knew what to study for the exams. It would be good if he'd pass out his lecture notes. (low in organization, examinations, assignments)

There was a lot in the lectures that wasn't in the book. How do we know what's important? I wish it was easier to meet with the prof. outside class, but his door was always closed! (low in organization, neutral in individual rapport)

The class discussions were great — it was like having your own seminar with the prof.!!

I liked the class discussions but I wish we could have spent more time discussing the readings that took so much of our time. (low in examinations and assignments)

More discussion would be great! It was great to debate issues with my classmates. That's what history's all about, right?

### *Responding to Prof. Deere's feedback*

The following series of questions are meant to guide you, as Prof. Deere's colleague, in helping him to identify issues for focus in his development as a teacher. Clearly, he is *not* a bad teacher, but he is puzzled and confused. He sees more negative feedback than he would like to have seen.

The following questions guide you to identify specific strengths and selected weaknesses. The term "selected weaknesses" is used to emphasize that you are seeking specific, achievable goals for teaching improvement. From these strengths and weaknesses, you can develop, implement, and assess specific strategies. Identification of every possible flaw suggested by the SEEQ feedback is not conducive for improvement. Problems that are ignored in this round may reappear next term, at which time they can become the focus of attention. On the other hand, these other problems may be corrected next term as a side effect of the adopted strategies.

Please sketch out answers to the following questions.

1. Numerical Ratings: In which SEEQ categories does Deere seem to do particularly well? Which categories present more problems? Identify two categories for each question.
2. Student Comments: What are some major themes that appear in the student comments? In other words, can you group the comments according to shared subject or underlying concept?
3. Combining the Numbers and Comments: How do tendencies in the numerical ratings correspond to themes in the comments?
  - 3a. Do the comments elaborate on the good and bad points listed in (1)?
  - 3b. Do the student comments and ratings correspond to Deere's goals for the class (from the text scenario)?
4. Having identified the two problematic points in the numerical ratings and the further information provided from the comments, choose the problem about which you feel you have the best data so far and that you feel is most "fixable." What is a strategy he might implement to alleviate this problem? If appropriate, you might build on an element of Deere's teaching that is already clearly successful.

*As you work through these questions, please keep a separate list of questions you have for Prof. Deere or other kinds of "teaching evidence" that might help your diagnosis.*

## In Your Own Case

Research on the effective use of student ratings in the improvement of teaching has often emphasized the need for one-on-one consultation, or mentoring, to guide the analyses of the results. The process you just followed — learning a bit about what Prof. Deere was trying to accomplish in his class, reviewing his results with a view toward selecting targets and strategies for improvement, and then discussing the results with the professor — can serve as a general model for your own self-study of your results. A better approach, however, would be to share this process with a colleague. In this way, you receive an “outside” perspective on your teaching.

### Self Evaluation as a baseline

The first step in analyzing your SEEQ results is to establish your own standards. In the given class, how do you believe you performed in each of the SEEQ questions? On a blank form, or on a form you mock-up yourself, mark the pair of responses that you expect will be the most common. For example:

	strongly disagree	disagree	neutral	agree	strongly agree	n/a
<b>LEARNING</b>						
1. I have found the course intellectually challenging and stimulating. ....	✓	✓				

This is your opportunity for careful self-evaluation. You were in that same classroom as the students. Were there aspects of your teaching that you thought did not go as well as you hoped? This self-evaluation should be carried out before you read the students’ responses. By comparing your self-ratings with your students’ ratings, you may quickly see points of divergence.

## Analyzing student feedback

The next step, as in Question 1 in the case study, is to identify specific areas of strength and weakness as indicated in the student ratings. Points of weakness that differ from your self-evaluation deserve particular attention.

Remember, as you go over the numbers, that you should not expect full student agreement on each topic. Reliability or consistency of response studies of SEEQ data show high levels of consistency only on a class-average basis, *not between individual students*. In any class, there are bound to be many student characteristics that legitimately influence how students perceive the teaching process and their own learning. For example, some students may be very adept at listening and processing information and can easily create coherent, organized, and useful lecture notes. Other students may still see the note-taking process as transcription. They inevitably will have more difficulty in a lecture course with few handouts, overheads, or other organizational hints. They will have a hard time deciding what is most important in the lecture, what is likely to be on the exam, etc. As a result of such a difference among the students in your class, you are likely to see a range of responses to questions of organization and/or examinations.

GOOD CLASS!  
FUN LECTURE!  
LOVED THE LABS!  
YUCK!!  
FUN CLASS 😊  
GREAT TEACHER!

YUCK? THEY HATED IT!  
I TRIED TO MAKE IT SO FUN!

Should you then ignore the negative responses in such areas that could be explained in part by student characteristics? Probably not.

But if such areas are among your chosen targets for improvement, you should be aware of how the effectiveness of your teaching is influenced by the characteristics of your students. Negative feedback might be saying as much about their difficulties as students as about your own difficulties as a teacher. While you might have some ability to correct their difficulties, you have far more control over adjustments to your own teaching so that you provide the most help to the widest proportion of your class.

... SOME COMMENTS STAND OUT

Student comments have always been a source of specific evidence that is too idiosyncratic to include in the general questionnaire. Even when SEEQ was being



developed, student comments added more items to the pool of survey questions for testing in the development process. One method of organizing these comments into a helpful framework is to determine themes that connect comments (as in Question 2 above). These themes can then be compared with the feedback from the numerical ratings.

The two forms of feedback – numbers and comments – serve complementary purposes. There is no reason to expect students to comment, uncoached, on each SEEQ factor and its various components. Their numerical feedback on these factors provides a general view of the students' perceptions of the major elements of teaching effectiveness. On their own, however, students are likely to elaborate on one or two factors in which the professor's teaching did or did not match the students' views of "good teaching." The comments section provides a forum for these views and, in so doing, provides further details on specific SEEQ factors.

### *Choosing a focus and strategies for improvement*

You now have three views of your teaching:

1. your self evaluation,
2. the good points and bad points in the student feedback and how they correspond to your self-study, and
3. the themes in the student comments and how they correspond to the numerical data.

The next stage in this process is to use these views of your teaching to select a target for your efforts at teaching improvement. You might choose the factor about which you have the most feedback (e.g., divergence between your self-evaluation and your students' views, and both numerical ratings and written feedback), or you might choose an area where you already have some ideas of how to improve. Remember that the changes you implement with a view toward one SEEQ factor might well affect others, so there is no need for strategies aimed at every factor at once.

Once you have picked the focus for your efforts, the next step is to select a strategy or set of strategies for improvement. There are many sources of such strategies.

First and foremost, if you are (as is advisable!) working with a colleague to decipher your SEEQ results, use your colleague as a sounding board and source for possibilities. Together, the two of you have many hours of classroom experience, both as teachers and students. Or ask others in your department; ask your squash partner (in Jacob Deere's case). The printed and electronic sources for tips or strategies are voluminous, so the following is only a partial list to get you started:

1. *Tips to Improve Academic Teaching* series: this is a series of teaching tips gathered for use with SEEQ factors. The paper version or electronic version (Macintosh or IBM) are available from UTS. Since this series is tied explicitly to SEEQ, this is the best place to start.
2. <http://uga.berkeley.edu/sled/compendium/sections.menu.html> — this is an excellent collection of teaching tips compiled by Barbara Gross Davis and grouped by topics that correspond quite well to SEEQ factors.
3. *Teaching Tips*, by Wilbert McKeachie: This is a well-indexed book, now in its 9th edition, which provides both the research bases and very practical advice on a wide range of teaching and learning issues. This book has, at times, been carried by the campus bookstore and could be ordered.
4. *New Directions for Teaching and Learning* series: These volumes are collected essays on particular themes such as *Assessing Students' Learning*, *Using Active Learning*, and *Fostering Critical Thinking*. Although it may be more difficult to find specific strategies in these volumes, they provide easy access to the state of higher education research in the given topic or issue. They are currently housed in the D.S. Woods Education Library.

There are many other sources, and the staff at UTS can help guide your search. For example, UTS has published *Teaching at The University of Manitoba: A Handbook*. This book covers many topics related to teaching improvement in the various contexts found in the University. UTS also maintains a large, indexed collection of books and articles on teaching practices.

### *Assessing the success of your strategy*

Once you have decided upon a possible strategy, it is also necessary to decide how you will assess that strategy's effectiveness. Waiting for the next SEEQ round at the end of the term will be neither timely nor particularly effective. SEEQ, alone, may not "detect" the specific effects of precise changes in teaching method. A more focused assessment tool is needed. Tom Angelo and Patricia Cross have gathered numerous possible assessment tools in their book, *Classroom Assessment Techniques: A Handbook for College Teachers* (San Francisco: Jossey-Bass 1993). Each technique in the book is indexed according to the type of teaching or learning that is being measured, and each is described in terms of the amount of effort needed for implementation and evaluation. Cross and Angelo advocate making classroom assessment a consistent and systematic part of classroom activity.

Imagine, for example, that you were concerned that many of your students were not understanding how lecture material fit the objectives of the course. At the end of a series of lectures on a given topic you might ask the students to write, in a "syllabus journal," at least two course objectives related to the lectures with a brief commentary. You could then collect the journals and skim them to be sure that students' perceptions matched your own. If some did not, you could devote a few minutes at the beginning of the next class to reviewing how you thought the course objectives had been achieved. After class, you might invite those students who were unclear on the objectives (you now know who they are!) to come to your office hours individually or as a group. The "syllabus journal" would have helped you check that you were being clear in your lectures, and your students would have a personal study guide tied directly to the syllabus and their lecture notes. [The "syllabus journal" is not one of Angelo and Cross's suggestions, but they offer many similar ideas].

There are other sources for assessment of your improvement strategies. One would be to continue working with your colleague to institute a cycle of classroom observations. Select a series of class sessions through the term. Before each session, meet with your observer to review the context for that particular class and what in particular you would like your observer to look for. After class, your observer should review the notes taken and give you feedback. Do not put off this debriefing session too long after the class, lest one of you forget important details. Some improvement strategies are best evaluated by a series of videotapes of your teaching. Again,

discussion of these videos with a colleague or Peer Consultant would be most beneficial. Such observation techniques can be quite effective in conjunction with Classroom Assessment Techniques as described above. The more sources of evidence, the better!

### *Recording and presenting your efforts*

As you can see, the method advocated here for effective use of SEEQ feedback moves the process of teaching evaluation from being a one-time, end of term ritual, to being a consistent and necessary part of effective course planning. And yet, the added efforts should not only improve your students' learning but should also increase your SEEQ ratings on the chosen factors. Nothing is guaranteed, of course, when working with real humans instead of class averages. But your chances of success are much improved with deliberate, careful implementation of strategies that have proven effective elsewhere.

How then will you keep track of, and receive credit for, this effort on behalf of your students? The most efficient method of documenting your teaching is through the use of a Teaching Portfolio (Dossier). The Portfolio is your opportunity to present your teaching in the specific context of your own view of the goals and purposes of teaching. UTS can provide you with many different resources and models for Portfolio construction. For the present purposes, however, your portfolio could contain the following items: summaries of your SEEQ scores, per year, per class; brief notes on how you used those results to guide your development process; and descriptions and assessment summaries of your strategies for improvement. All of these items would be grouped in a section of your Portfolio dedicated to "Efforts at Teaching Improvement."

UTS offers many workshops and discussion groups on the development of Teaching Portfolios.

## Annotated list of further readings

---

Aleamoni, L. (ed.) *Techniques for Evaluating and Improving Instruction*, New Directions for Teaching and Learning 31 (San Francisco 1987). Topics in this collection include the improvement of teaching through student ratings, typical faculty misconceptions of ratings, and the potential role of student governments in the evaluation of teaching.

Braskamp, L., Brandenburg, D. and Ory, J. *Evaluating Teaching Effectiveness: A Practical Guide*, (Beverly Hills 1984). This book provides an overview of methods of evaluating teaching, the different goals of evaluation, and the sources of data appropriate to those goals. This is a good source for thinking about adding different sources of data to your study of your teaching.

Brinko, K.T. "The Practice of Giving Feedback to Improve Teaching: What is Effective?" *Journal of Higher Education* 64(5) 1993, 574-593. This study reviewed previous studies of effective teaching improvement practices with the aim of defining general practices that have a high likelihood of success. As a result, the article provides a good starting point for defining the roles and effective behaviours for someone who has agreed to assist a colleague in reviewing evaluation data.

Cashin, W. and Downey, R. "Using Global Student Rating Items for Summative Evaluation," *Journal of Educational Psychology* 84(4) 1992, 563-572. Cashin and Downey review many of the issues related to policy development in the area of summative use of student ratings. Their comments are particularly appropriate to The University of Manitoba, since Cashin generally advocates a multi-dimensional approach to teaching evaluation for formative purposes (similar to Marsh).

Costin, F., Greenough, W., and Menges, R. "Student Ratings of College Teaching, Reliability, Validity, and Usefulness," *Review of Educational Research* 41 1971, 511-535. A very thorough review of major developments and trends in student ratings research prior to 1970. Marsh developed SEEQ out of this general research context.

Greenwald, A. (ed.) *Current Issues: Student Ratings of Professors*, in *American Psychologist*, November 1997, 1182-1225. Includes the following contributions: A. Greenwald, "Validity concerns and usefulness of student ratings of instruction," H. Marsh and L. Roche, "Making students' evaluations of teaching effectiveness effective," S. d'Apollonia and P. Abrami, "Navigating student ratings of instruction," A. Greenwald and G. Gillmore, "Grading leniency is a removable contaminant of student ratings," W. McKeachie, "Student ratings: The validity of use." This collection provides a very convenient place to see the current state of student ratings research. Marsh, McKeachie, Abrami and d'Apollonia represent various perspectives dominating 1980s and 1990s research; Greenwald and Gillmore represent themselves as challengers to the orthodoxy (every field has them!). What is particularly interesting in this collection is that each contributor not only provides a standard article but also comments in direct response to the other contributors.

Marsh, H. and Dunkin, M. J. "Students' Evaluations of University Teaching: A Multidimensional Perspective," in R. Perry and J. Smart (eds.) *Effective Teaching in Higher Education: Research and Practice* (New York 1997), 241-320. Marsh and Dunkin provide a very thorough and detailed overview of research on the SEEQ ratings survey. This article includes consideration of many of the ongoing debates in student ratings research and is an excellent place to gain a working introduction to student ratings research.

Marsh, H. and Roche, L. "The Use of Students' Evaluations and an Individually Structured Intervention to Enhance University Teaching Effectiveness," *American Educational Research Journal* 30(1) 1993, 217-251. This work, along with the works by McKeachie et al. 1980 and Roseman and Kerwin 1991 (below), document the effectiveness of consultation in the process of using student ratings to improve teaching.

McKeachie, W. et al. "Using Student Ratings and Consultation to Improve Instruction," *British Journal of Educational Psychology* 50 1980, 168-174.

Murray, H. "Effective Teaching Behaviors in the College Classroom," in R. Perry and J. Smart (eds.) *Effective Teaching in Higher Education: Research and Practice* (New York 1997), 171-204. Murray uses research focused on his Teaching Behaviors Inventory (a survey tool that is even more focused on behaviours

than SEEQ but is also longer and hence more cumbersome to use), in conjunction with studies of student achievement and attitude development, to connect specific behaviours with student learning and development. As a direct result of the research design, this article provides many practical strategies for teaching improvement.

Overall, J. and Marsh, H. "Midterm Feedback from Students: Its Relationship to Instructional Improvement and Students' Cognitive and Affective Outcomes," *Journal of Educational Psychology* 71(6) 1979, 856-865. This work considers the benefits in terms of both student achievement and student attitudes toward academia that stem from the implementation of mid-term, as opposed to end-of-term, student ratings surveys. The benefits especially accrued in those situations where the instructor discussed the results and intended reforms with the class following the survey process.

Perry, R.( ed.) Special Section: Instruction in Higher Education, in *Journal of Educational Psychology* 82(2) 1990, 183-274. This collection of articles reviews many aspects of research in higher education, including the history of research; student preference for giving mid-term, interview-style feedback on teaching; issues of the unit of analysis in student ratings research; validity and reliability studies of peer and student ratings; the low impact of student feedback alone; and the influence of personality on teaching effectiveness.

Roseman, J. and Kerwin, M. "Evaluating the Effectiveness of a Teaching Consultation Program on Changing Student Ratings of Teaching Behaviors," *Journal of Staff, Program and Organizational Development* 9(4) 1991, 223-230.

Theall, M. and Franklin, J. (eds.) *Student Ratings of Instruction: Issues for Improving Practice*, New Directions for Teaching and Learning 43 (San Francisco 1990). This collection includes coverage of some essential issues of policy concerning the implementation of a student ratings system, such as the ethics of implementation, who should know the results of ratings and how results should be communicated, and other issues related to the summative use of student ratings.

Theall, M. and Franklin, J. (eds.) *Effective Practices for Improving Teaching*, New Directions for Teaching and Learning 48 (San Francisco 1991). This collection of articles highlights the connection (and the prerequisites for making the connection) between gathering data on teaching quality and improving the quality of teaching.