



EDB 7560

8888 University Drive, Burnaby, BC
Canada V5A 1S6

TEL 778.782.4130

FAX 778.782.4900

tlcdir@sfu.ca

MEMORANDUM

ATTENTION	Senate	DATE	December 9, 2013
FROM	Panayiotis Pappas, Chair, SCUTL	PAGES	1/1
RE:	TCEP Final Report		

Please find enclosed the Final Report for the Teaching and Course Evaluation Project, as well as all the necessary supporting documents. This project carried forward earlier work by the Senate Committee on University Teaching and Learning (SCUTL) and the Task Force on Teaching and Learning (TFTL). The focus of this work is to enable student evaluations of teaching and courses at SFU to provide more useful feedback to instructors for the purposes of self-evaluation, and higher quality information to those who evaluate performance for reappointment, salary review, tenure and promotion.

After careful deliberation and a fruitful discussion, the committee has passed the following motion:

That, based on the final draft which includes requested revisions, the Senate Committee on University Teaching and Learning endorses the recommendations in the Teaching and Course Evaluation Project (TCEP) Final Report and is in favour of adopting the corresponding principles as outlined in the TCEP Best Practices on Interpretation and Use of the Evaluation Data.

Therefore, we recommend that Senate pass the following motion

Motion: That Senate endorses the recommendations in the Teaching and Course Evaluation Project (TCEP) Final Report and is in favour of adopting the corresponding principles as outlined in the TCEP Best Practices on Interpretation and Use of the Evaluation Data.

In closing, I would like to thank, on behalf of the University, all the current and past members of SCUTL, the members of TFTL, and the TCEP team, for their efforts in this important project.

Panayiotis A. Pappas

Chair, Senate Committee on University Teaching and Learning
Associate Professor, Linguistics

Enclosures



SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

THE TEACHING AND COURSE EVALUATION PROJECT FINAL REPORT

NOVEMBER 27, 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

DOCUMENT CONTROL

DOCUMENT INFORMATION

	INFORMATION
Document Owner	<i>Corinne Pitre-Hayes</i>
Project /Organization Role	<i>Project Leader</i>

DOCUMENT HISTORY

VERSION	ISSUE DATE	CHANGES
<i>1.0</i>	<i>Sep 30, 2013</i>	<i>Preliminary outline</i>
<i>1.1</i>	<i>Nov 2, 2013</i>	<i>Revisions to outline per feedback from SCUTL</i>
<i>1.2</i>	<i>Nov 11, 2013</i>	<i>Rough draft for core team review</i>
<i>1.3</i>	<i>Nov 14, 2013</i>	<i>Revisions incorporating input from core team</i>
<i>1.4</i>	<i>Nov 17, 2013</i>	<i>Further revisions from core team input</i>
<i>1.5</i>	<i>Nov 18, 2013</i>	<i>Final revisions prior to distribution to SCUTL</i>
<i>1.6</i>	<i>Nov 26, 2013</i>	<i>Revisions from SCUTL review meeting on Nov 25, 2013</i>
<i>1.7</i>	<i>Nov 27, 2013</i>	<i>Final edits</i>

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	1
	RESEARCH	1
	CONSULTATION WITH OTHER INSTITUTIONS	2
	SFU COMMUNITY ENGAGEMENT	2
	PROOF OF CONCEPT	2
	KEY FINDINGS AND RECOMMENDATIONS	2
	CONCLUSIONS, HIGH LEVEL IMPLEMENTATION PLAN AND NEXT STEPS	3
2	SFU PROJECT ON STUDENT EVALUATION OF TEACHING AND COURSES	4
	2.1 BACKGROUND	4
	2.2 SCOPE	4
	2.3 GOALS, OBJECTIVES AND OUTCOME	5
	2.4 A PRAGMATIC APPROACH	5
	2.5 THE PROJECT TEAM	5
3	RESEARCH	6
	3.1 LITERATURE SEARCH	6
	3.2 INSTRUMENT	7
	3.3 OTHER INSTITUTIONS	7
4	SFU COMMUNITY ENGAGEMENT	7
	4.1 COMMUNICATION PLAN	7
	4.2 ADMINISTRATIVE INTERVIEWS	8
	4.3 STUDENT SURVEY	8
	4.4 FACULTY SURVEY	9
	4.5 COMPARISON OF STUDENT VS FACULTY RESPONSES	9
	4.6 FOCUS GROUPS	10
5	KEY FINDINGS	10
6	PROOF OF CONCEPT	11
	6.1 PROOF OF CONCEPT OVERVIEW	11
	6.2 VOLUNTEER INSTRUCTOR FEEDBACK	11
	6.3 STUDENT FEEDBACK	11
	6.3.1 EVALUATION INSTRUMENT	11
	6.3.2 REPORTS	12
	6.4 CONCLUSION	12
7	RECOMMENDATIONS	12
8	CONCLUSIONS, IMPLEMENTATION AND NEXT STEPS	14
	8.1 CONCLUSIONS	14
	8.2 HIGH LEVEL IMPLEMENTATION PLAN	15
	8.3 RECOMMENDED NEXT STEPS	16

APPENDIX I: PROJECT CHARTER

APPENDIX II: REPORT ON KEY RESEARCH FINDINGS

APPENDIX III: REPORT ON KEY RESEARCH FINDINGS - INSTRUMENTATION

APPENDIX IV: INSTITUTION CONSULTATION KEY FINDINGS SUMMARY

APPENDIX V: COMMUNICATION PLAN

APPENDIX VI: FACULTY ADMIN SURVEY KEY FINDINGS SUMMARY

APPENDIX VII: RESULTS FROM SURVEY OF STUDENTS: TEACHING AND COURSE EVALUATION PROCESS

APPENDIX VIII: RESULTS FROM SURVEY OF FACULTY AND INSTRUCTORS: TEACHING AND COURSE EVALUATION PROCESS

APPENDIX IX: FOCUS GROUPS KEY FINDINGS SUMMARY

APPENDIX X: PROOF OF CONCEPT

APPENDIX XI: PROOF OF CONCEPT, SAMPLE SUPPORT MATERIALS

APPENDIX XII: PROOF OF CONCEPT, RESULTS FROM STUDENT SURVEYS

APPENDIX XIII: INCORPORATING CONTEXT INTO INTERPRETATION OF EVALUATION DATA

APPENDIX XIV: REFERENCES

1 EXECUTIVE SUMMARY

The goal of the Teaching and Course Evaluation (TCE) project has been to make recommendations to govern the selection/development of an updated system for student evaluation of teaching and courses. It is important to stress that the project did not arise from dissatisfaction on the part of students or academic administrators about the quality of instruction or curriculum. Rather, the purpose of the project was to develop an updated instrument and process that reflects new understanding from educational research and is more focused on student learning. The project was based on earlier work by the Senate Committee on University Teaching and Learning (SCUTL) and the Task Force on Teaching and Learning (TFTL).

In particular, our goal has been to recommend key requirements for the replacement of the existing instrument and processes with an emphasis on:

- Improving the teaching and learning environment
- Ensuring efficient methods of data collection, storage and protection of privacy
- Adopting guidelines for best practices in the use of evaluation data.

The project utilized a pragmatic approach starting with a literature review in support of the goals of the project and consultation with other institutions. The heart of the approach was a strong focus on engaging with the SFU community. Lastly, a proof of concept was completed as a small scale practical application of the recommendations that had emerged.

SCUTL acted as the project steering committee and the core project team consisted of a project manager, the Chair of SCUTL, members of the Teaching and Learning Centre (TLC) as well as members of Institutional Research and Planning (IRP).

RESEARCH

Four main themes emerged in the broad literature review:

- 1) The complexity and multi-dimensionality of teaching
- 2) The recent shift to a more learner-centred pedagogy
- 3) The many challenges to validity that can be introduced in the design process, implementation, and interpretation of results
- 4) The importance of using a *set* of evaluative processes.

From a faculty perspective, the research suggested two main concerns with respect to student evaluation of teaching:

- Validity and the effect of student bias
- Use of data for tenure and promotion decisions by personnel without a sophisticated understanding of how to interpret the data.

For academic administrators and decision-makers, concerns centred more on:

- Perception by faculty that bias may exist in the use of the data for tenure and promotion
- Capacity and commitment of the university to provide training and development that supports faculty teaching
- Ensuring that staff undertaking administrative roles in the evaluation process are familiar with and sensitive to issues that can distort the results.

From the student point of view, the research indicated that “improvement in teaching” is the outcome they consider most valuable from participation in the evaluation process.

An additional review of the literature specifically related to evaluation instruments highlighted similar concerns to the broad review. The instrument review also indicated that flexibility is key to enable giving special attention to local definitions of effective teaching, priorities and standards at an institution.

Another strong theme was that numerous studies clearly demonstrate that online evaluations *can* produce reliable and valid results if they are designed and tested according to rigorous theoretical and psychometric principles. This can be accomplished by utilizing an existing databank of questions that have been rigorously tested and then continuing to add to this databank with the assistance of the appropriate expertise.

CONSULTATION WITH OTHER INSTITUTIONS

Interviews were conducted with 23 different institutions that were categorized as leaders, competitors, comparables and others. The emphasis was on learning from the *experiences* of the institutions.

Key best practices that emerged from discussions with those interviewed included:

- Defining governing principles and implementing policies that align with these principles
- Adopting a flexible framework capable of meeting the needs of stakeholders
- Providing structured support from a designated department in the institution
- Taking a careful, thoughtful approach.

SFU COMMUNITY ENGAGEMENT

A communications strategy was developed along with a community engagement strategy. The community engagement included:

- Interviews with 33 managers, advisors, secretaries, assistants and directors involved with administering student evaluations of teaching and courses; this was an update to the 2008 survey conducted for SCUTL
- A survey of undergraduate students
- A survey of all faculty and instructors who taught a course in 2012
- 11 focus groups conducted across 6 faculties and 3 important stakeholder groups.

PROOF OF CONCEPT

A small scale demonstration or "proof of concept" for the emerging recommended approach was conducted to obtain pragmatic feedback from SFU instructors and students. The proof of concept also:

- Demonstrated that the adapted multi-level framework appears to be sufficiently flexible
- Provided insight into the challenges of implementation
- Highlighted the importance of providing adequate support when moving to the new system
- Underscored the importance of clearly defining reporting requirements up front
- Confirmed that, with careful attention to encouraging student participation, it is possible to achieve a strong response rate with online evaluations at SFU.

KEY FINDINGS AND RECOMMENDATIONS

In general, our findings can be characterized in seven key areas:

- 1) Care must be taken in evaluation design, implementation and interpretation to produce reliable and **valid results**
- 2) The instrument must have sufficient **flexibility** to meet the needs of a wide range of stakeholders including the university as a whole, individual faculties, individual schools or departments, and instructors
- 3) **Responsible use of the data** is paramount and requires a keen focus on training and support of those undertaking administrative roles in the evaluation process
- 4) Proactive measures are needed to encourage the **use of evaluation data to improve teaching**
- 5) **Efficiency** is critical both in administering evaluations and in making the results easier to access and interpret
- 6) **Engagement** of faculty and students in the evaluation process is essential
- 7) Adequate **structured support** for design, implementation and interpretation of evaluations is a critical success factor.

A number of clear recommendations have been identified in each of the seven key areas outlined above.

CONCLUSIONS, HIGH LEVEL IMPLEMENTATION PLAN AND NEXT STEPS

A great deal of common ground and similarity was evident between our findings in the literature, the consultation with other institutions and our engagement with the SFU community. The framework that has emerged has been adapted slightly to meet the specific needs of SFU and was successfully demonstrated in a small scale trial or proof of concept. The SFU community has clear concerns that need to be addressed as part of the implementation, and have indicated a strong organizational readiness for change. The steps in a high level implementation plan have been developed and a rough timeline in three phases suggested.

Once Senate approval is received, it is recommended that a team be formed to undertake a formal Request for Proposal process. It is also recommended that a parallel process begin immediately to define key institution-wide policies and teaching and learning priorities for institution-wide questions.

2 SFU PROJECT ON STUDENT EVALUATION OF TEACHING AND COURSES

2.1 BACKGROUND

The Teaching and Course Evaluation (TCE) project carried forward earlier work by the Senate Committee on University Teaching and Learning (SCUTL) and the Task Force on Teaching and Learning (TFTL). The focus of this work is to enable student evaluations of teaching and courses at SFU to provide more useful feedback to instructors for the purposes of self-evaluation, and higher quality information to those who evaluate performance for reappointment, salary review, tenure and promotion.

It is important to stress that the project did not arise from dissatisfaction on the part of students or academic administrators about the quality of instruction or curriculum. Rather, the purpose of the project was to develop an updated instrument and process for student evaluation of teaching that reflects new understanding from educational research on such evaluations. It also provided an opportunity to consider how these evaluation processes could be used to gather information on the learning environment, and thus to focus less on student perceptions of instructors, and more on how students learn. The original project charter is included in Appendix I of this report.

2.2 SCOPE

The scope of the project was to make recommendations with respect to:

- 1) Replacing SFU's form(s) for student evaluation of teaching and courses with an emphasis on improving the teaching and learning environment for course instructors and students
- 2) Ensuring efficient methods of data collection and storage are used and that the privacy rights of instructors and students are protected
- 3) Adopting guidelines for best practices in the use of evaluation data.

It should be emphasized that developing methods for faculty evaluation was not in scope for the TCE project. Evaluation of the various instruments that are currently in place across the University was also not in scope. SCUTL previously completed a comprehensive evaluation of the University's current practices and tools related to student evaluations of courses and instructors. This project built on that earlier work.

Additionally, as the focus was on instructors, the evaluation instrument used for Teaching Assistants (TAs) and Tutor Markers (TMs) was not in scope. It should be noted that meetings were held with representatives of the Teaching Support Staff Union (TSSU) to get their input and feedback on the project. The scope of the project was discussed and feedback was noted regarding ensuring students had an opportunity to indicate evaluation questions were not applicable to them, and that sessional instructors were given an opportunity to respond to the instructor survey. It was also discussed that, once the new evaluation system is implemented, it can be offered to the TSSU to determine if they would recommend adoption for TAs and TMs.

The final outcome of the project is outlined in this report. The Senate, as the representative body that governs academic matters, is requested to approve the implementation of the recommendations contained herein.

2.3 GOALS, OBJECTIVES AND OUTCOME

The goals, objectives and outcome of the Teaching and Course Evaluation (TCE) Project are summarized in Table 1.

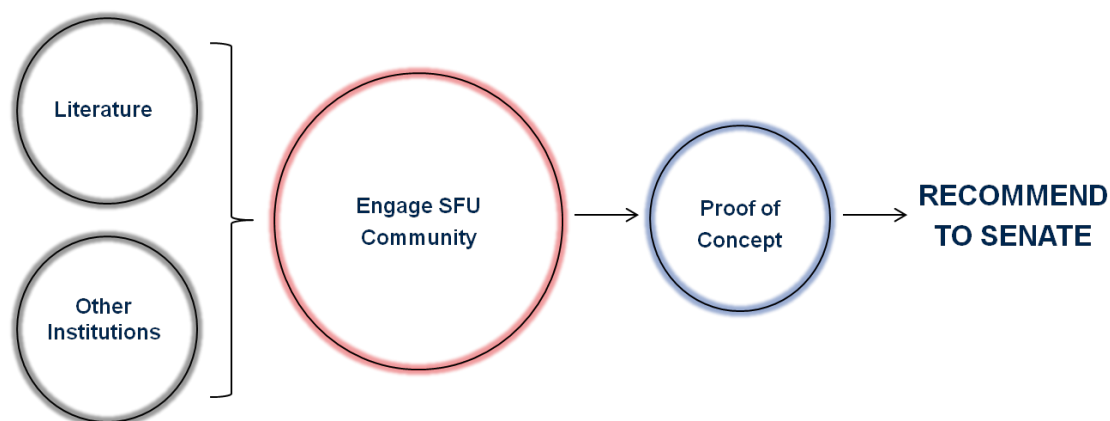
Table 1: TCE Project Goals, Objectives and Outcome

Goals	<ul style="list-style-type: none"> • Support teaching • Enhance student learning • Ensure effective and responsible use of evaluation data and information
Objectives	<ul style="list-style-type: none"> • Replace SFU's instrument and processes for student evaluation of instructors and courses • Engage the SFU community throughout the project • Develop a best-practices guide on interpretation and use of the data
Outcome	<ul style="list-style-type: none"> • Recommend for approval by Senate

2.4 A PRAGMATIC APPROACH

The project utilized a pragmatic approach to achieving its goals and objectives. The foundation of the approach was a literature review and consultation with other institutions. The heart of the approach was a strong focus on engaging with the SFU community. The importance of this engagement was such that the project duration was significantly lengthened to ensure comprehensive and meaningful input. Lastly, a proof of concept was completed as a small scale practical application of the recommendations that had emerged. Figure 1 illustrates the project approach.

Figure 1: TCE Project Approach



2.5 THE PROJECT TEAM

The project and commitment of resources for this initiative were approved by Senate at the [May 2011 Meeting](#). The Senate Committee on University Teaching and Learning (SCUTL) has overseen the project (for a current list of SCUTL representatives please see [SCUTL Standing committee](#)), with support provided by the Teaching and Learning Centre (TLC) and Institutional Research and Planning (IRP).

The project team members included:

- Corinne Pitre-Hayes, Project Leader – Managed the timelines for the activities of the project; coordinated the activities of the team members, SCUTL, and the broader participation by SFU instructors, administrators, and students; coordinated logistics for meetings and sessions for participation in the project; coordinated correspondence and deliverables for the project

- Stephen Spector, former Chair of SCUTL – Provided direction and guidance and contributed to overall coordination of the project
- Johanne Provençal, Researcher – Reviewed the research literature on student evaluation of teaching and courses; provided comprehensive summary of the literature for the SFU community
- Hui Niu, Researcher – Reviewed and summarized the research literature on student evaluation of teaching and courses from instrument development and methodological perspectives; reviewed instruments and processes at other universities, gathered feedback from SFU community
- Chris Groeneboer, Associate Director, TLC – Participated as core team member; developed the community engagement strategy; assisted in gathering and documenting input from the SFU community
- Mark Bachmann, Communications Officer, TLC – Participated as core team member; managed the communications requirements for the project; assisted in gathering and documenting input from the SFU community
- Vea Banana, Communications Assistant, TLC – Participated as core team member; assisted with developing project communications
- Daniel Ahadi, Coordinator, Applied Research in Teaching and Learning, TLC – Participated as core team member; assisted in gathering and documenting input from the SFU community
- Jessica Tilley, Intermediate Analyst, IRP – Participated as core team member; assisted with the development, delivery and analysis of student and faculty surveys
- Joanne Quan, Intermediate Analyst, IRP – Participated as core team member; assisted with the development, delivery and analysis of student and faculty surveys; assisted in gathering and documenting input from the SFU community.

3 RESEARCH

3.1 LITERATURE SEARCH

In the past 30 years, student evaluation of teaching has become widely adopted internationally. Significant research literature on student evaluation of teaching (and courses) has formed. The full overview of key issues affecting the discussion from the research literature is included in Appendix II.

In summary, four main themes that emerged in the literature review:

- 1) The complexity and multi-dimensionality of teaching
- 2) The recent shift to a more learner-centred pedagogy
- 3) The many challenges to validity that can be introduced in the design process, implementation, and interpretation of results
- 4) The importance of using a *set* of evaluative processes.

From a faculty perspective, the research suggested two main concerns:

- Validity and the effect of student bias
- Use of data for tenure and promotion decisions by personnel without a sophisticated understanding of how to interpret the data.

For academic administrators and decision-makers, concerns centred more on:

- Perception by faculty that bias may exist in the use of the data for tenure and promotion
- Capacity and commitment of the university to provide training and development that supports faculty teaching
- Ensuring that staff undertaking administrative roles in the evaluation process are familiar with and sensitive to issues that can distort the results.

From the student point of view, the research indicated that “improvement in teaching” is the outcome they consider most valuable from participation in the evaluation process.

3.2 INSTRUMENT

The project also conducted an additional literature review focused on evaluation instruments. The full overview of research findings on student evaluation of teaching and courses with a focus on instrument development and related issues is included in Appendix III.

Similar to the overall literature review, this additional review highlighted concerns regarding the complexity and multi-dimensionality of teaching. The research showed that what constitutes effective teaching may vary from institution to institution, discipline to discipline, and faculty to faculty. Therefore, special attention should be given to local preferences, priorities and standards pertinent to effective teaching at an institution.

Another strong theme in the instrument review was that numerous studies clearly demonstrate that online evaluations *can* produce reliable and valid results. With respect to ensuring validity and reliability, the literature suggests that this can be achieved by designing and testing instruments according to rigorous theoretical and psychometric principles. Psychometrics is the field of study concerned with the theory and technique of psychological measurement. The field is primarily concerned with the construction and validation of measurement instruments and includes the development of theoretical approaches to measurement. An efficient approach to designing in accordance with psychometric principles is to utilize an existing databank of tested and validated questions and then testing and validating new questions as they are added to the databank.

3.3 OTHER INSTITUTIONS

A key reason for reviewing and updating student evaluation of teaching and courses at SFU is the desire to take advantage of more recent knowledge and understanding in this area – both nationally and internationally. To that end, a key component of our research included consultation with other institutions. Interviews were conducted with 23 different institutions that were categorized as leaders, competitors, comparables and others.

The emphasis was on learning from the *experiences* of these institutions. Topics related to evaluation frameworks, processes used, types of instruments, use of evaluation data and best practices were explored.

Key best practices that emerged from discussions with those interviewed included:

- Defining governing principles and implementing policies that align with these principles
- Adopting a flexible framework capable of meeting the needs of stakeholders
- Providing structured support from a designated department in the institution
- Taking a careful, thoughtful approach.

Consultation with other institutions raised questions that the team explored further in the Proof of Concept phase of the project. The full summary of the findings of this phase of research can be found in Appendix IV.

4 SFU COMMUNITY ENGAGEMENT

4.1 COMMUNICATION PLAN

A two part communication plan was developed for the TCE project. It consisted of an overall communication strategy and a summary of communication and information resources for the Proof of Concept phase of the project. Communications for the project were managed from the perspective of those most affected by the evaluation process.

Critical communications issues were identified and addressed:

- For those who are evaluated (faculty members and instructors)
- For those who evaluate (students)

- For those who use the evaluation data (e.g. administrators)
- For those who coordinate the evaluation process (e.g. departmental managers).

The following communication and information resources were completed:

- A website containing information regarding the project, research reports, project reference documents and interesting articles on the subject of student evaluation of teaching and courses
- Articles in SFU News, The Peak and other SFU-related media
- General information, sample scripts, Q&As, how-to guides as support materials for the Proof of Concept.

The project Communications Plan is included in Appendix V.

4.2 ADMINISTRATIVE INTERVIEWS

In 2008, the Senate Committee on University Teaching and Learning (SCUTL) conducted a survey of how academic units use evaluations. The results of this survey can be found in the document "Evaluating How We Evaluate: Examining SFU's course and instructor evolution system" (Summer 2008). Building on this earlier work, our engagement with the SFU community included interviews with staff involved with the administration of student evaluation of teaching and courses. Interviews were conducted with 33 managers, advisors, secretaries, assistants and directors.

The objectives of these interviews were to:

- Update our information on the types of instruments currently in use at SFU
- Gain a deeper understanding of the current student evaluation business practices across the institution
- Obtain insight into the organizational readiness of the SFU community to embrace a new approach to student evaluation of teaching and courses; as assessment of readiness includes consideration of whether there is a clearly defined need for change, how open the community is to change, whether the resources are available to make the change and what might be needed to sustain the change.

Key themes that emerged from discussions with individuals across the institution involved in teaching and course evaluation processes include:

- Significant challenges with the existing largely paper based processes
- Opportunities for better faculty and student engagement with the evaluation process
- Concern regarding use and interpretation of evaluation results.

Almost everyone interviewed agreed that an online approach would help create efficiencies. In general, the top priorities for a new system are reducing time and cost while increasing interactivity between students and faculty. Overwhelmingly, all academic units interviewed would like to see a redesigned evaluation process that helps create a consistent approach to improving teaching and learning through effective course delivery. A full summary of the key findings of the faculty administration survey are included in Appendix VI.

4.3 STUDENT SURVEY

In order to invite wider student input on the teaching and course evaluation process, a few questions were included in the SFU Undergraduate Student Survey, administered in October 2012 by SFU's Office of Institutional Research and Planning. Of the 12,260 students invited to participate, 2,785 responded to these questions, yielding a response rate of 23%.

Respondents said that the following improvements are important:

- Having individual instructors use the findings to modify their teaching or the course (selected by 92% of respondents)
- Showing that departments are taking the findings seriously (90%)
- Clearly informing students about what's being done with the findings (90%)

- Providing training for instructors who score below a certain level (88%)
- Having students provide input mid-way through the term, so that instructors have time to make adjustments (87%).

The details of the survey results can be found in Appendix VII.

4.4 FACULTY SURVEY

Since not all faculty/instructors could be included in focus groups, faculty and instructors who taught a course in 2012 were invited to participate in a survey, to provide feedback on the teaching and course evaluation process. The survey, administered in Spring 2013 by SFU's Office of Institutional Research and Planning, was sent to 1,429 faculty and instructors. In total, 519 responded, yielding a response rate of 36%.

Faculty/instructors were asked how student evaluation results should be used. Among respondents:

- 77-78% indicated that results should be used to a great/moderate extent when:
 - Making changes to teaching methods or course designs
 - Identifying opportunities for teacher development
 - Reviewing candidates for teaching positions
- 29-34% said that tenure & promotion and merit increases should be impacted very little or not at all by student evaluation of instruction.

Respondents said that the following are important aspects of an evaluation process:

- Clearly informing faculty (selected by 87%) and students (81%) about what's being done with the findings
- Allowing customizable sections where departments (selected by 83%) and individual instructors (77%) may choose their own questions.

Faculty/instructors were also asked which aspects of teaching need to be included on an evaluation in order for it to be a representative assessment of teaching.

- 12 aspects of teaching were identified as important by more than half of respondents
- The most commonly selected aspects were:
 - Organization of teaching and course material (82%)
 - Clarity of presentation (81%)
 - Appropriate assessment and feedback (78%)
 - Stimulation of critical thinking (78%).

The details of the survey results can be found in Appendix VIII.

4.5 COMPARISON OF STUDENT VS FACULTY RESPONSES

The Fall 2012 student survey and the Spring 2013 faculty/instructor survey had several questions in common. Among respondents:

- Both students and faculty/instructors said it is important to clearly inform students about what is being done with the findings (92% of student respondents and 81% of faculty/instructor respondents indicated this was important)
- Student respondents were more likely to say that it is important for individual instructors to use the findings to modify their teaching or the course (87%, versus 69% of faculty/instructor respondents)
- Student respondents were nearly twice as likely to say that completing a mid-term evaluation is important (90%, versus 47% of faculty/instructor respondents).

The details of the survey results can be found in Appendix VII (student survey) and Appendix VIII (faculty/staff survey).

4.6 FOCUS GROUPS

In order to facilitate a more in-depth dialogue with members of the SFU community, 11 focus groups were conducted across 6 faculties and 3 important stakeholder groups. The focus groups included:

- Beedie School of Business
- Faculty of Applied Sciences
- Faculty of Arts and Social Sciences
- Faculty of Education
- Faculty of Health Sciences
- Faculty of Science
- SFU Faculty Association
- TPC Chairs
- Students.

While there were wide ranging and instructive comments from faculty, many of the main concerns related primarily to:

- Validity and the effect of student bias
- The fundamental need for a flexible instrument
- Inappropriate use of data for tenure and promotion.

Concerns of TPC Chairs centred around:

- Faculty attitudes and perception of bias in the use of the data
- Concerns about how evaluation data may sometimes be the only source of information available and may be used inappropriately for decision making
- Capacity and commitment of departments to provide training to TPC Chairs.

Students expressed concern regarding:

- Whether the data is used at all for “improvement in teaching”
- Student access to more reliable information to make course selection decisions.

Focus groups provided perspective, deep insights and many valuable suggestions for implementation. A full summary of the 11 focus groups is included in Appendix IX.

5 KEY FINDINGS

Research, consultation with other institutions and engagement with the SFU community have provided a strong foundation and a rich source of insight for recommendations to replace SFU's form(s) for student evaluation of teaching and courses.

In general, our findings can be characterized in seven key areas:

- 1) Care must be taken in evaluation design, implementation and interpretation to produce reliable and **valid results**
- 2) The instrument must have sufficient **flexibility** to meet the needs of a wide range of stakeholders including the university as a whole, individual faculties, individual schools or departments, and instructors
- 3) **Responsible use of the data** is paramount and requires a keen focus on training and support of those undertaking administrative roles in the evaluation process
- 4) Proactive measures are needed to encourage the **use of evaluation data to improve teaching**
- 5) **Efficiency** is critical both in administering evaluations and in making the results easier to access and interpret
- 6) **Engagement** of faculty and students in the evaluation process is essential
- 7) Adequate **structured support** for design, implementation and interpretation of evaluations is a critical success factor.

6 PROOF OF CONCEPT

6.1 PROOF OF CONCEPT OVERVIEW

A small scale demonstration of the emerging recommended approach was conducted to obtain pragmatic feedback from SFU instructors and students. This demonstration, or "proof of concept", was based on an adaptation of a flexible, multi-level framework from the University of Toronto which included a large set of psychometrically tested questions.

A total of 14 volunteer instructors from six different faculties participated. There were 14 lecture/seminar and 4 CODE courses ranging in size from 13 to 329 students for a potential total of 1,329 evaluations. Participants were provided with a number of support materials. A compendium of support materials provided to participants is included in Appendix XI.

Care was taken to ensure the privacy rights of instructors and students were protected. The proof of concept was in compliance with University Policy | 10.08, Collection and Disclosure of Instructor and Course Evaluations.

The response rate for this small scale effort was 72% overall, with lecture/seminar courses at 82% and CODE courses at 27%. For lecture/seminar courses, the response rate was considerably higher than the typical range of 30-60% reported for online evaluations in the literature (Rawn 2008). The main differences between lecture/seminar and CODE courses were that lecture/seminar courses included a direct appeal from the instructors and, in a number of cases, also included a small bonus mark incentive.

6.2 VOLUNTEER INSTRUCTOR FEEDBACK

On the whole, the feedback was very positive. Participants also provided valuable constructive feedback and suggestions.

Key themes in participant feedback included:

- The flexibility of the multi-level framework is great, but it will be very important and could be challenging to reach agreement on the right questions to ask at the different levels
- The support materials provided were good, but personal support is strongly recommended during the transition to the new system
- The reports were valuable and useful, but more sophisticated reporting capabilities would be very helpful
- With a strong focus on encouraging student participation, the online approach seemed to produce good response rates and the open-ended responses were longer and more thoughtful.

More details on the feedback from volunteer instructors can be found in Appendix X.

6.3 STUDENT FEEDBACK

6.3.1 EVALUATION INSTRUMENT

During the Summer 2013 proof of concept administered by eXplorance Inc., a few questions were included at the end of each evaluation, asking students about their experience with the evaluation process. All 1,297 students who received invitations to fill out the proof of concept course evaluations were also invited to participate in the survey. Of these students, 960 participated in the survey, yielding a response rate of 74%.

Among respondents:

- 95% agreed that the evaluation questions were clear, and that the evaluation interface was user-friendly
- Students also had the opportunity to provide open-ended feedback. Among those who gave feedback:
 - Most said that the process was a positive experience
 - Many expressed their preference for an online system over the current in-class paper evaluations
 - Many students provided constructive feedback to improve the evaluation process.

The details of the survey results can be found in Appendix XII.

6.3.2 REPORTS

At the beginning of the Fall 2013 term, all 1,297 students who were enrolled in courses that participated in the Summer 2013 proof of concept were invited to view reports with aggregated results from the online teaching and course evaluations, as well as a guide to help them interpret the reports. A survey, administered by eXplorance Inc., was then sent to these students, seeking their feedback on the reports. In total, 180 students responded to the survey, yielding a response rate of 14%.

Among respondents:

- 70% had read the guide, and of these 86% found it useful in helping them understand the evaluation report
- 89% were satisfied with the readability of the reports
- 60 - 69% said it was important to include means and bar charts to describe responses to evaluation questions
- Fewer than 50% said it was important to include median, mode, standard deviation, frequency tables, and a comparison of course scores to average faculty or university scores.

The details of the survey results can be found in Appendix XII.

6.4 CONCLUSION

Overall, the project team felt that the proof of concept was a valuable exercise. It demonstrated that the multi-level framework adapted from the University of Toronto appears to be sufficiently flexible and provided insight into some of the challenges that will need to be overcome for implementation. It highlighted the importance of providing adequate support during the transition to the new system. The importance of clearly defining reporting requirements up front and ensuring strong reporting capabilities during system selection was strongly underscored. Finally, it demonstrated that, with careful attention to encouraging student participation, it is possible to achieve a strong response rate with online evaluations at SFU.

7 RECOMMENDATIONS

Table 2 summarizes our recommendations specifically aimed at addressing our findings in the key areas outlined above.

Table 2: Summary of Recommendations in Key Areas

Key Areas	Recommendation
Valid Results	<ol style="list-style-type: none"> 1. Set clear evaluation goals, including clear definitions of what constitutes effective teaching at each level in the organization, and ensure that evaluation questions reflect these goals 2. Design and test instruments according to rigorous theoretical and psychometric standards; this can be accomplished by utilizing a databank of questions that have been rigorously tested by the University of Toronto (used in the proof of concept) as the basis for implementation and then continuing to develop and test additional questions for inclusion in the databank with the assistance of the TLC and IRP. 3. Include appropriate supplementary evidence with evaluation data 4. Establish appropriate and standardized policies and processes for the administration of course evaluations; namely, ensure that: <ul style="list-style-type: none"> • Policy and practice about the administration of evaluations is standardized at the administrative level at which comparison between instructors or courses (if employed) is made • An appropriate amount of data is distributed to appropriate populations; appropriate and consistent policies for access to and storage of data are developed • The privacy of instructors and students is protected
Flexibility	<ol style="list-style-type: none"> 5. Multiple-level approach to instrumentation is recommended including a core set of questions common to all forms (institutional level), faculty-wide question set if desired, department-level question set if desired, and instructor selected and/or created questions if desired 6. The results of instructor selected/created questions should be private to the instructor 7. Select an evaluation instrument with flexible, sophisticated reporting capabilities 8. The timing of evaluations will be flexible to meet the local needs of the faculties, schools, department and specific courses
Responsible use of the data	<ol style="list-style-type: none"> 9. Student Evaluation of Teaching (SET) should not be the sole source of data for decision-making around teaching performance; other sources might include peer evaluation (classroom observation) and/or teaching portfolios 10. One of the principles in the literature is that evaluation results should not be interpreted without incorporating contextual information about the course (such as class size, CGPA of students, year level of students, etc.); the recommended approach to address this issue is to use evaluation data to create a model that takes into account contextual factors that are beyond the instructor's control; this will indicate whether an achieved evaluation score is above, below, or within expectation; Institutional Research and Planning undertook an exercise to demonstrate this approach using the proof of concept data; the process is described in Appendix XIII
Use of evaluation data to improve teaching	<ol style="list-style-type: none"> 11. Request a brief summary from faculty about their teaching practices, goals, and challenges for each course that would accompany and help contextualize evaluation results; a set of questions can be provided to assist faculty with structuring the summary; the timing and audiences for these summaries may vary across faculties 12. Use evaluation data as a means of providing diagnostic or formative feedback 13. Use evaluation data for summative purposes 14. Offer students other means to provide feedback, such as mid-course evaluations 15. Test and review instruments when institutional priorities or teaching practices change 16. Conduct self-studies and internal research on evaluation validity and factors affecting evaluation responses

	17. Establish policy frameworks for the collection, administration and use of student course evaluation systems
Key Areas	Recommendation
Efficiency	18. Establish more automated processes for collecting, storing and distributing evaluation data 19. Present data so that it can be easily and accurately interpreted
Engagement	20. Articulate and broadly communicate evaluation goals and purpose 21. Provide sufficient information to instructors and students about the administration and use of evaluations 22. Provide instructors and students with access to appropriate evaluation results; it should be noted that there are concerns with regard to sharing evaluation data with students; the general principle of sharing with students is recommended, however, the details of which results are communicated to students may vary across faculties 23. Promote to instructors and students ways in which the data can be used to make improvements in teaching 24. Ensure that evaluation methodologies and approach continue to be relevant through ongoing engagement with the SFU community regarding student evaluation of teaching and courses
Structured support	25. Encourage and provide the infrastructure for consultation on teaching evaluations 26. Provide an opportunity for instructors to receive individualized assessment and advice 27. Provide faculty with information about evaluation data collection and use 28. Educate and train administrators 29. Develop educational materials and support and make these materials centrally available

8 CONCLUSIONS, IMPLEMENTATION AND NEXT STEPS

8.1 CONCLUSIONS

The project team has observed a great deal of common ground and similarity between our findings in the literature, the consultation with other institutions and our engagement with the SFU community. The framework that has emerged has been adapted slightly to meet the specific needs of SFU and was successfully demonstrated in a small scale trial or proof of concept. The SFU community has clear concerns that need to be addressed as part of the implementation, and have indicated a strong organizational readiness for change.

8.2 HIGH LEVEL IMPLEMENTATION PLAN

The steps that outline a high level plan to implement a new system for student evaluation of teaching and courses at SFU are summarized in Table 3.

Table 3: Outline of High Level Implementation Plan

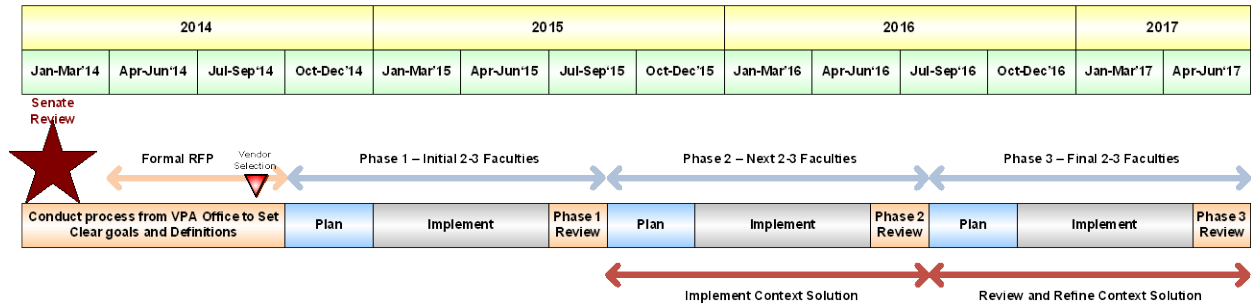
Implementation Step	Description
Step 1	Form initial implementation team
Step 2	Draft project charter and confirm steering committee (SCUTL is recommended)
Step 3	Draft formal Request for Proposal (RFP), conduct selection process
Step 4 (Note: Step 3 and 4 could be conducted in parallel)	Conduct process from VPA Office to set clear evaluation goals, including clear general definitions of what constitutes effective teaching, and develop institution-wide evaluation questions that reflect these goals; it is recommended to ensure both education and evaluation expertise are involved in this process
Step 5	Confirm and validate goals, definitions and questions set in Step 4
Step 6	Confirm and validate the adapted evaluation framework for adoption by SFU
Step 7	Define phases of implementation project in partnership with selected vendor and SFU community
Step 8	Begin phase 1 of the implementation
Step 9	Define, build and/or update structured support for student evaluations of teaching and courses in the Teaching and Learning Centre
Step 10	Complete phase 1 of the implementation
Step 11	Conduct brief review of lessons learned in phase 1 of the implementation
Step 12	Initial implementation of proposed solution for incorporating context into the interpretation of evaluation results (please refer to Appendix XIII for details)

Based on consultation with other institutions, it is recommended to tackle the implementation of a new system incrementally by starting with two to three faculties and then slowly rolling the system out to the rest of the community. Depending on the phases defined in Step 7, and the lessons reviewed in Step 11, it is recommended that the subsequent phases of the implementation project follow a similar pattern to Steps 8 through 11. Following the completion of the first phase of the project, it is recommended that the solution for incorporating context into the interpretation of evaluation results outlined in Appendix XIII be implemented and further explored.

Figure 2 illustrates a suggested high level timeline for rolling out the implementation in three phases of two to three faculties each. The initial rollout begins in the Fall of 2014 following the conclusion of the formal RFP process. Phase 2 and 3 would follow with the target for all faculties to be fully implemented by mid 2017.

Figure 2: Suggested High Level Timeline

Suggested High Level Timeline
 Updated: Nov 17, 2013



8.3 RECOMMENDED NEXT STEPS

Once Senate approval is received the following next steps are recommended:

- Form a team to draft a Request for Proposal for online evaluation system with participation from TLC, IRP, IT, SCUTL and faculty members
- Conduct a formal selection process for an evaluation system
- In parallel with the evaluation system selection process, begin a process of defining key institution-wide policies aligned with recommended principles including defining teaching and learning priorities for institution-wide questions.

APPENDIX I: PROJECT CHARTER

APPENDIX II: REPORT ON KEY RESEARCH FINDINGS

**APPENDIX III: REPORT ON KEY RESEARCH FINDINGS -
INSTRUMENTATION**

**APPENDIX IV: INSTITUTION CONSULTATION KEY FINDINGS
SUMMARY**

APPENDIX V: COMMUNICATION PLAN

APPENDIX VI: FACULTY ADMIN SURVEY KEY FINDINGS SUMMARY

**APPENDIX VII: RESULTS FROM SURVEY OF STUDENTS: TEACHING
AND COURSE EVALUATION PROCESS**

**APPENDIX VIII: RESULTS FROM SURVEY OF FACULTY AND
INSTRUCTORS: TEACHING AND COURSE EVALUATION PROCESS**

APPENDIX IX: FOCUS GROUPS KEY FINDINGS SUMMARY

APPENDIX X: PROOF OF CONCEPT

APPENDIX XI: PROOF OF CONCEPT, SAMPLE SUPPORT MATERIALS

**APPENDIX XII: PROOF OF CONCEPT, RESULTS FROM STUDENT
SURVEYS**

**APPENDIX XIII: INCORPORATING CONTEXT INTO INTERPRETATION
OF EVALUATION DATA**

APPENDIX XIV: REFERENCES



SIMON FRASER UNIVERSITY

TEACHING AND COURSE EVALUATION
PROJECT CHARTER

MARCH 5, 2012

Contact: Corinne Pitre-Hayes
604.765.4610
corinne_pitre-hayes@sfu.ca

DOCUMENT CONTROL

DOCUMENT INFORMATION

	INFORMATION
Document Owner	<i>Corinne Pitre-Hayes</i>
Project /Organization Role	<i>Project Manager</i>

DOCUMENT HISTORY

VERSION	ISSUE DATE	CHANGES
1.0	<i>Jan 5, 2012</i>	<i>Outline</i>
1.1	<i>Jan 10, 2012</i>	<i>Draft for core team review</i>
1.2	<i>Jan 11, 2012</i>	<i>Draft for SCUTL Committee review and approval in principle</i>
1.3	<i>Feb 5, 2012</i>	<i>Updated with SCUTL feedback for core team review</i>
1.4	<i>Feb 10, 2012</i>	<i>Core team revisions for distribution to SCUTL for feedback</i>
1.5	<i>Mar 5, 2012</i>	<i>Approved version with final wording changes from SCUTL</i>

DOCUMENT APPROVALS

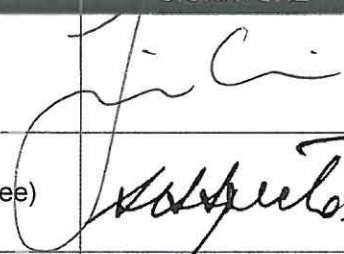
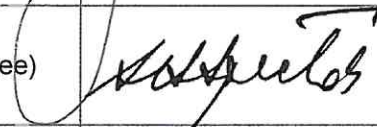

ROLE	NAME	SIGNATURE	DATE
Project Sponsor	Dr. Jon Driver		<i>30 March 2012</i>
SCUTL Committee	Chair, Stephen Spector (on behalf of the Committee)		March 28, 2012
Project Manager	Corinne Pitre-Hayes		March 28, 2012

TABLE OF CONTENTS

- 1 PROJECT CHARTER PURPOSE.....1
- 2 EXECUTIVE SUMMARY1
- 3 OVERVIEW OF THE TCP PROJECT1
- 4 SCOPE OF THE PROJECT2
 - 4.1 GOALS AND OBJECTIVES2
 - 4.2 ORGANIZATIONAL IMPACTS.....2
 - 4.3 PROJECT PHASES3
 - 4.4 RELATED INITIATIVES3
 - 4.5 OUT OF SCOPE3
 - 4.6 PROJECT TIMELINE.....4
- 5 PROJECT CONDITIONS5
 - 5.1 ASSUMPTIONS5
 - 5.2 ISSUES5
 - 5.3 RISKS5
- 6 STRUCTURE AND APPROACH6
- 7 PROJECT ORGANIZATION7
- 8 PROJECT REFERENCES7

1 PROJECT CHARTER PURPOSE

This project charter defines the scope, objectives, and overall approach for the Teaching and Course Evaluation Project (TCE). It is a critical element for initiating, planning, executing, controlling, and assessing the project. It is a single point of reference on the project for project goals and objectives, scope, organization, estimates and work plan. In addition, it serves as a contract between the core project team and the project sponsor, stating what will be delivered according to the budget, time constraints, risks, resources, and standards agreed upon for the project.

2 EXECUTIVE SUMMARY

- The project goals are to support teaching and enhance student learning through the responsible use of teaching and course evaluation data and information at SFU.
- The objectives include:
 - Replacing SFU's instruments and processes for student evaluation of teaching by instructors.
 - Replacing SFU's instruments and processes for student evaluation of courses.
 - Engaging the SFU community throughout the project e.g. open forum discussion sessions on each campus, meetings at the faculty and/or departmental level, focus groups, input via social media, and surveys.
 - Developing a best practices guide on interpretation and use of the data.
- The project will be in three phases:
 - I. Environmental scan and needs assessment
 - II. Instrument Development
 - III. Instrument Implementation Planning & Documentation
- Key assumptions include that there are proven course and teaching evaluation instruments and practices that will be well suited to SFU's needs and that it will be possible to achieve a reasonable consensus among all affected SFU constituencies.
- The project is expected to run from October 2011 through December 2012; the timeline may need to be extended based on the outcome of each phase.
- The approach is to use a robust project organization that includes a Core Team producing deliverables and the SCUTL Committee that will review and recommend deliverables on a timely basis. Project communication will be key and will be facilitated by a project website. Issue escalation will be swift with similarly rapid resolution.

3 OVERVIEW OF THE TCP PROJECT

The goal of this project is to support teaching and to enhance student learning at SFU. It carries forward the earlier work by the Senate Committee on University Teaching and Learning (SCUTL) and the Task Force on Teaching and Learning (TFTL), and responds to concerns that current evaluations by students could provide more useful feedback to instructors for the purposes of self-evaluation, and higher quality information to those who evaluate performance for reappointment, salary review, tenure and promotion.

It is important to stress that the project does not arise from significant dissatisfaction on the part of students or academic administrators about the quality of instruction or curriculum. Rather, the purpose of the project is to develop an updated instrument and process for student evaluation of teaching that reflects new understanding from educational research on such evaluations. It also provides an opportunity to consider how these evaluation processes could be used to gather information on student learning, and thus to focus less on student perceptions of instructors, and more on how students learn.

During the course of the project there will be opportunities to provide input and feedback. Ultimately, Senate, as the representative body that governs academic matters, will be asked to approve the final outcome of this project.

4 SCOPE OF THE PROJECT

4.1 GOALS AND OBJECTIVES

Goals	Objectives
Support teaching and enhance student learning at SFU	<ol style="list-style-type: none"> 1. Replace SFU's instruments and processes for student evaluation of teaching and courses 2. Engage the SFU community throughout the project e.g. open forum discussion sessions on each campus, meetings at the faculty and/or departmental level, focus groups, input via social media, and surveys
Responsible use of evaluation data	<ol style="list-style-type: none"> 3. Develop a best practices guide on interpretation and use of the evaluation data

4.2 ORGANIZATIONAL IMPACTS

Organizational Group	Impact to and Participation of Organization Group
Instructors at SFU responsible for teaching courses, including tenured and tenure-track faculty members, lab instructors, lecturers and senior lecturers, limited term appointments, and sessional instructors	Very significant impact personally and professionally; the project will invite input from the instructors throughout the project including: open forum discussion sessions on each campus, meetings at the faculty and/or departmental level, focus groups, input via social media, and surveys.
Students	Major impact on their ability to provide feedback on the SFU learning environment; the project will similarly invite input from students throughout the project.
SFUFA and the TSSU	Significant changes that affect SFUFA and TSSU members; the project will invite input from the SFUFA and TSSU throughout the project.
Tenure and Promotion Committee chairs	Significant changes that affect how Tenure and Promotion Committee chairs work with course and evaluation data at SFU; the project will invite input from the Tenure and Promotion Committee chairs throughout the project.
Academic departments	Significant changes that affect how academic departments work with course and evaluation data at SFU; the project will invite input from academic departments throughout the project.
Administrative staff involved in managing teaching and course evaluation processes	Potentially significant impact on the processes used to do their work; the project will similarly invite input from administrative staff throughout the project.
Office of the Vice President, Academic and Provost	Provide linkages to related and ongoing initiatives.

4.3 PROJECT PHASES

Phase	Tasks
I. Environmental scan and needs assessment	<ul style="list-style-type: none"> • Review SFU documents, professional and academic literature, work at other institutions and provide summary to support SFU community as project participants • Circulate summary of literature findings to SCUTL and the project team • Identify constituencies • Consult with SFU constituencies for input on the development of the instrument through meetings, info sessions, focus groups, surveys • Identify people/groups interested in being more actively involved in the project • Finalize details of project process and specifics of timelines based on feedback
II. Instrument Development	<ul style="list-style-type: none"> • Develop the instrument: construct mapping, clarity, validity, reliability • Provide constituencies with specifics of timelines for their input • Incorporate feedback to revise instrument and follow up with constituencies • Pilot the instrument to select group within SFU (guided by recommendations from SCUTL and the SFU community) and accompanied by the draft user guide • Provide status update to Senate
III. Instrument Implementation Planning & Documentation	<ul style="list-style-type: none"> • Circulate best practices/guidelines document for effective and responsible use of student evaluation of teaching and courses (based on literature and SFU input) • Circulate revised instrument user guide (after revisions to instrument completed) • Work with all stakeholders on implementation planning • Submit project report and report back to Senate

4.4 RELATED INITIATIVES

SFU is undertaking a number of initiatives that focus on teaching and learning. It is important to note that this project is separate from the consultation about implementation of a learning outcomes and assessment process.

4.5 OUT OF SCOPE

Evaluation of the various instruments that are currently in place across the University is not in the scope of this project. SCUTL previously completed a comprehensive evaluation of the University's current practices and tools related to student evaluations of courses and instructors. This project builds on that earlier work.

The project will make recommendations with respect to:

- 1) Replacing SFU's teaching and course evaluation form(s) in order to improve the teaching and learning environment for course instructors and students
- 2) Ensuring efficient methods of data collection and storage are used and that the privacy rights of instructors and students are protected, and
- 3) Adopting guidelines for best practices in the use of evaluation data.

As the focus is on instructors, the evaluation instrument used for TAs and TMs is not in the scope of this project. Additionally, developing methods for faculty evaluation is not in the scope of this project.

Senate, as the representative body that governs academic matters, will be asked to approve the final outcome of this project.

4.6 PROJECT TIMELINE

Project Phase	Date Estimate	Deliverable(s) Included	Confidence Level
Phase I: Environmental scan and needs assessment	October 2011	<ul style="list-style-type: none"> Summary points of findings presented to SCUTL 	High
	January 2012	<ul style="list-style-type: none"> Finalized content for info sheet 	High
	February-March 2012	<ul style="list-style-type: none"> Confirmed categories of participants Updated draft of process 	High
	March 2012	<ul style="list-style-type: none"> Summary of approaches used at other universities Status report on date and time of meetings with participants 	Medium
	March 2012	<ul style="list-style-type: none"> Updated list of participants and meeting dates/times Summary of approaches used in SFU departments Summary of relevant research findings 	Medium
	July 2012	<ul style="list-style-type: none"> Summary of preliminary input from participants 	Medium
	July 2012	<ul style="list-style-type: none"> Status report to Senate 	High
Phase II: Instrument Development	July 2012	<ul style="list-style-type: none"> Status report on development of instrument 	High
	August 2012	<ul style="list-style-type: none"> Status report on development of instrument 	High
	September-December 2012	<ul style="list-style-type: none"> Continue collecting and analyzing input from participants 	Medium
Phase III: Instrument Implementation Planning & Documentation	January-February 2013	<ul style="list-style-type: none"> Instrument pilot Draft user guide for pilot 	Medium
	March-June 2013	<ul style="list-style-type: none"> Revised instrument and revised user guide following pilot Best practices 	Medium

		guidelines document • Project report • Draft implementation plan • Report to Senate	
	July 2013	• Submit outcome of project to Senate	Medium

5 PROJECT CONDITIONS

5.1 ASSUMPTIONS

- Proven course and teaching evaluation instruments and practices exist that will be well suited to SFU's needs.
- It is possible to achieve a reasonable consensus among all affected SFU constituencies regarding the most appropriate course and teaching evaluation instrument and processes to replace the existing outdated ones.

5.2 ISSUES

Priority Criteria

- 1 – High-priority/critical-path issue; requires immediate follow-up and resolution.
- 2 – Medium-priority issue; requires follow-up before completion of next project milestone.
- 3 – Low-priority issue; to be resolved prior to project completion.
- 4 – Closed issue.

#	Date	Priority		Description	Status & Resolution
1	Jan 10, 2012	High		Overall project charter and plan not yet completed	Expedite approval in principle of charter by SCUTL and complete plan based on approved charter
2	Jan 10, 2012	Medium		Input and feedback from the SFU community behind schedule	Arrange for distribution of information document; develop and secure approval of communication plan; execute approved plan
3	Jan 10, 2012	Medium		Consultation with other institutions behind schedule	Develop and secure approval for consultation plan; execute approved plan

5.3 RISKS¹

#	Risk Area	Likelihood		Project Impact-Mitigation Plan
1	Misunderstandings	High		• Resistance to the project

¹ For the purposes of continuity, the Project Manager role will have overall responsibility for developing strategies to mitigate risks identified during the course of the project. The implementation of these strategies will also be conducted by the Project Manager in discussion with the SCUTL Committee Chair.

#	Risk Area	Likelihood		Project Impact-Mitigation Plan
	about the project			<ul style="list-style-type: none"> • Initiate multiple efforts to communicate the purpose and process for the project, and sources informing the work of the project
2	Low participation by constituencies	Medium		<ul style="list-style-type: none"> • Inadequate ownership of the project by constituencies • Provide multiple opportunities for participation throughout the project
3	Low level of support for the new instrument by constituencies	Medium		<ul style="list-style-type: none"> • Difficulties with implementation • Integrate into the project plan formal support for the project at multiple levels
4	Difficulty using the revised instrument effectively	High		<ul style="list-style-type: none"> • Data set (from evaluations) not helpful for constituencies • Develop and distribute: <ol style="list-style-type: none"> (1) User manual for instrument (2) Best practices guide to support effective and responsible interpretation and use of evaluation data
5	Low level of commitment and follow up by relevant SFU constituents	High		<ul style="list-style-type: none"> • Take an iterative approach to all aspects of the project to engage constituents on an ongoing basis; look for opportunities to increase commitment and buy-in

6 STRUCTURE AND APPROACH

- The plan to manage the challenges of the project is as follows:
 - Create robust project organization that includes a Core Team producing deliverables, the SCUTL Committee that reviews and recommends deliverables on a timely basis, Project Sponsor that reviews and approves deliverables on a similarly timely basis.
 - Regular meetings of all key groups: Core Team weekly and Steering Committee monthly.
 - Brief but complete documentation that enables everyone to be clear and on the same page.
 - A project website that provides a means of storing a “single version of the truth” and that enables effective collaboration on deliverables.
 - All meetings will include a review of how the team is tracking to the project plan, and discussion of how to correct any delays.
 - If it becomes apparent that there is an issue that will result in a significant delay, the issue will immediately be escalated to the Steering Committee and or the Project Sponsor for resolution.

7 PROJECT ORGANIZATION

Project Team Role	Project Team Member(s)	Responsibilities
Project Sponsor	Dr. Jon Driver	<ul style="list-style-type: none"> • Primary liaison with the project manager • Ensure adequate resources are applied to the project
SCUTL Committee	Stephen Spector (Chair) Diana Cukierman (Applied Sciences) Adrienne Burk (Arts and Social Sciences) Stephen Spector (Business Admin) Russell Taylor (Communication, Art & Technology) Roger Fries (Education) Bob Muir (Environment) Timothy Beischlag (Health Sciences) Chris Kennedy (Science) TBA, Undergraduate Student Marena Brinkhurst, Graduate Student Stephanie Chu, Director, Teaching and Learning Centre, Secretary, Ex-officio Elaine Fairey, Director, Student Learning Commons Ex-officio Nancy Johnston, Executive Director, Student Affairs (or designate) Ex-officio Maria Davis, Recording Secretary Ex-officio	<ul style="list-style-type: none"> • Approve and recommend project deliverables • Identify risks/issues, potential mitigation and resolution • Oversee issues resolution • Liaison with the community
Core Team	Chris Groeneboer Corinne Pitre-Hayes Hui Niu Johanne Provencal Stephen Spector	<ul style="list-style-type: none"> • Produce project deliverables • Subject matter expertise • Research risk mitigation and issue resolution
Project Manager	Corinne Pitre-Hayes	<ul style="list-style-type: none"> • Develop and maintain project charter and plan • Coordinate internal communications

8 PROJECT REFERENCES

Name	Description and/or Link
Project website	https://sharepoint.sfu.ca/sites/tcp/SitePages/Home.aspx



SFU

SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

TEACHING AND COURSE EVALUATION PROJECT FINAL REPORT

APPENDIX II: REPORT ON KEY RESEARCH FINDINGS

This research report was prepared for the Simon Fraser University Senate Committee on University Teaching and Learning (SCUTL) by Johanne Provençal, PhD, as part of the SFU Teaching and Course Evaluation Project. The report has been reformatted for inclusion in the project's final report.

JANUARY 15, 2012

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	1
1.1	BACKGROUND AND PURPOSE	1
1.2	CONTEXT: BROAD ISSUES IDENTIFIED IN THE RESEARCH.....	1
1.2.1	THE QUESTION OF “GOOD TEACHING”	1
1.2.2	THE SHIFT TO A LEARNER-CENTRED PEDAGOGY.....	1
1.2.3	VALIDITY AND RELIABILITY/STUDENT BIAS	2
1.2.4	NEED FOR A SET OF EVALUATIVE PROCESSES.....	2
1.3	KEY ISSUES AROUND STUDENT EVALUATION OF TEACHING AND COURSES	2
1.3.1	FACULTY PERSPECTIVE	2
1.3.2	ADMINISTRATOR PERSPECTIVE	3
1.3.3	STUDENT PERSPECTIVE	3
2	INTRODUCTION	4
2.1	BRIEF HISTORY OF THE PROJECT.....	4
2.2	INTERNATIONAL CONTEXT AND BROAD ISSUES IDENTIFIED IN THE RESEARCH.....	4
2.2.1	THE MULTI-DIMENSIONALITY OF TEACHING AND THE QUESTION OF “GOOD TEACHING”	6
2.2.2	UNDERSTANDINGS ABOUT STUDENT LEARNING AND THE SHIFT TO LEARNER- CENTRED CURRICULUM	7
2.2.3	ISSUES OF VALIDITY AND RELIABILITY IN STUDENT EVALUATION OF TEACHING ..	9
3	EVALUATION OF TEACHING AND COURSES	11
3.1	PEER EVALUATION OF TEACHING	12
3.2	TEACHING PORTFOLIOS	13
3.3	STUDENT EVALUATION OF TEACHING AND COURSES.....	14
4	STUDENT EVALUATION OF TEACHING AND COURSES: ISSUES FOR FACULTY, ADMINISTRATORS, AND STUDENTS	14
4.1	ISSUES FOR FACULTY	15
4.1.1	RECOGNIZING POSSIBLE STUDENT BIAS	15
4.2	ISSUES FOR ADMINISTRATORS AND THE UNIVERSITY ADMINISTRATION.....	19
4.3	ISSUES FOR STUDENTS	21
5	REFERENCE LIST	25

1 EXECUTIVE SUMMARY

1.1 BACKGROUND AND PURPOSE

This project was prompted by two reports:

- The “Evaluating How We Evaluate” report (Senate Committee on University Teaching and Learning (SCUTL, 2009), which called for a review of SFU’s process for evaluating teaching and courses; and
- The “Task Force on Teaching and Learning Recommendations Report” (2010), which repeated the call.

The purpose of the project is to review and update the instrument used for student evaluation of teaching and courses at SFU and to provide documentation to support faculty, administration, and students in their use of the new instrument and in the interpretation of instrument data.

This project did not arise from dissatisfaction with teaching and learning at SFU. Rather, it comes from a desire to take advantage of new understandings from educational research on the use, nationally and internationally, of student evaluation of teaching and courses.

1.2 CONTEXT: BROAD ISSUES IDENTIFIED IN THE RESEARCH

In the past 30 years, student evaluation of teaching has become widely adopted internationally, and evaluation of universities has become of increasing importance as accountability for post-secondary institutions and concerns regarding student learning have drawn the attention of researchers in the field of higher education, of governments providing funding to universities, of industry and community organizations employing university graduates, and of students (as well as parents) investing time and tuition dollars in postsecondary education.

An important part of quality assurance for universities is in the area of teaching or instruction, and student feedback is the most common and significant form of assessing this area. As a consequence, a significant research literature on student evaluation of teaching (and courses) has formed. An overview of key issues affecting the discussion from the research literature is presented in the report.

1.2.1 THE QUESTION OF “GOOD TEACHING”

The complexity and multi-dimensionality of teaching are widely recognized and there is not a universally accepted definition of what constitutes “teaching excellence” or a “good teacher.” Underlying the various characterizations of teaching are assumptions and perspectives not only about the purpose of higher education, but also, about the roles and responsibilities of both teacher and student. In recent years there has been a notable shift in higher education from a “teacher-centred” to a “learner-centred” classroom and curriculum.

1.2.2 THE SHIFT TO A LEARNER-CENTRED PEDAGOGY

A learner-centred approach requires teaching assessments to take into account the role of students, since students are understood to be active participants in the learning process. It also requires changes in post-secondary teaching practice. For example, students who have become dependent learners in teacher-centred learning environments may need to learn how to become independent, autonomous learners, and both teachers and students need to acquire an understanding of the responsibilities of the teacher for teaching and the learner for learning. Some level of student resistance can be expected in shifting from a teacher-centred to a learner-centred classroom.

1.2.3 VALIDITY AND RELIABILITY/STUDENT BIAS

Validity in student evaluation of teaching and courses can be defined as “the effectiveness of the test in representing, describing, or predicting the attribute that the user is interested in.”¹ The concept also encompasses the issue of student bias. Reliability studies generally address the question “Are student ratings consistent both over time and from rater to rater?”

Research demonstrates that tools used for faculty evaluation, particularly student rating forms, “can be designed to be valid and reliable,”² although challenges remain in terms of development and implementation. The issues of validity and reliability will be discussed in more detail in a separate report. However, it can be noted that SFU faculty, students, and administrators will need to determine what components of effective teaching to address in questions. Reliability will be a key consideration during the test phase of the development of the instrument, and use of studies identifying sources of student bias and showing what areas students are not equipped to evaluate will be essential.

Ultimately, it is important to be aware that in the absence of formal and careful evaluation, faculty “will be evaluated by hearsay evidence, gossip, and other shoddy means.”³ Although a perfect process may not be attainable, the SFU community is encouraged to hold in sight the purpose and promise of the project, which is to support faculty, administrators, and students in creating an environment that is rewarding and conducive to successful teaching and learning.

1.2.4 NEED FOR A SET OF EVALUATIVE PROCESSES

One of the key findings that surfaces repeatedly in the research literature is the importance of using a *set* of evaluative processes. Two approaches to teaching evaluation that are most often used to complement student evaluation of teaching and courses are peer evaluation of teaching, most commonly conducted through classroom observations by faculty in the same department as the faculty member being evaluated, and teaching portfolios, which contain a careful selection of material collected and developed over the course of a faculty member’s career.

1.3 KEY ISSUES AROUND STUDENT EVALUATION OF TEACHING AND COURSES

1.3.1 FACULTY PERSPECTIVE

The research literature suggests that faculty have two main concerns about student evaluations:

- Validity and the effect of student bias; and
- Use of data for tenure and promotion decisions by personnel without a sophisticated understanding of how to interpret the data.

Research indicates that some faculty members see student evaluations as a “popularity contest” that can lead to potentially unsound educational outcomes; for example, leniency in grading in order to obtain more positive evaluations. Such flaws and abuses can exist, but it must also be noted that careful instrument design can reduce “noise” and guidelines on interpreting the data can provide considerable support for responsible, effective and instructive use of student evaluation of teaching and courses.

Similarly, student bias can exist for many reasons, from class size to time of day to gender of the instructor or student. Again, carefully chosen questions and proper instructions for interpreting and using data can reduce the impact of such variables.

¹ See Arreola 2007, p. 99 (citing Thorndike and Hagen 1969, pp.163-177)

² See Arreola 2007, p. 98

³ See Seldin 2006, p. 7

1.3.2 ADMINISTRATOR PERSPECTIVE

Administrators need to be aware of the perception by faculty that bias may exist in the interpretation and use of data for tenure and promotion decisions.

A second concern involves the capacity and commitment of the university administration to using the data produced by student evaluation to provide training and development that supports faculty teaching.

Effective and responsible use of student evaluation of teaching requires that staff undertaking administrative roles in the process be familiar with and sensitive to a number of issues that can distort the results.

Beyond the responsible and informed collection and interpretation of data produced from student evaluation of teaching, it is important for administrators to use the data to identify where support is most needed by faculty and to provide faculty career development accordingly.

1.3.3 STUDENT PERSPECTIVE

One of the most significant findings from the research on student perceptions of teaching evaluation is that "improvement in teaching" is the outcome that students consider to be most valuable in their participation in such evaluations.

The shift to a learner-centred classroom and curriculum presents challenges to students, as it does for teachers. Although a shared understanding between faculty and students about their respective roles (and responsibilities) in teaching and learning is invaluable, knowledge of the reasons students give for their high and low evaluations of teaching is useful for identifying student views on effective teaching.

2 INTRODUCTION

The purpose of this report is to provide the Simon Fraser University (SFU) community – faculty,⁴ administrators, and students – with a summary of research findings in support of the project overseen by the Office of the Vice-President Academic and Provost to update the SFU process and form used for student evaluation of teaching and courses. In the interest of effective and responsible use of student evaluation of teaching and courses at SFU, wide use of this report is encouraged and therefore, the main text shall remain relatively brief with footnotes and further reading in the reference list for those interested in further details.⁵

2.1 BRIEF HISTORY OF THE PROJECT

The Senate Committee on University Teaching and Learning (SCUTL) report, “Evaluating How We Evaluate” (2009) and the 2010 “Task Force on Teaching and Learning Recommendations Report” (which followed university-wide consultations) both called for a review of the process used by SFU for the evaluation of teaching and courses. In 2011, the Vice-President Academic and Provost, Jonathan Driver, made a request to Senate to approve a project to be led by SCUTL to review and update the instrument used for student evaluation of teaching and courses at SFU and to provide documentation to support faculty, administration, and students in their use of the new instrument and in the interpretation of instrument data. After Senate approval of the project in the summer of 2011, SCUTL and the project team⁶ began work on the project in the first semester of the 2011/2012 academic year, with the administrative support of the SFU Teaching and Learning Centre.

From the outset of the project, the Vice-President Academic and Provost, as well as the project team, have aimed to communicate clearly that this project did not arise from dissatisfaction with teaching and learning at SFU. In addition, those involved in the project recognize the importance of a set of evaluative processes for teaching and courses in which student evaluation of teaching and courses is only one component. The purpose of the present project is to develop an updated form and process for student evaluation of teaching and courses at SFU that reflects new understandings from educational research on the widespread use, nationally and internationally, of student evaluation of teaching and courses.

2.2 INTERNATIONAL CONTEXT AND BROAD ISSUES IDENTIFIED IN THE RESEARCH

In the past 30 years student evaluation of teaching has become widely adopted internationally.⁷ The evaluation of universities, more broadly, has become of increasing importance as accountability for post-secondary institutions and concerns regarding student learning have drawn the attention of researchers in the field of higher education, of governments providing funding to universities, of industry and community organizations employing university graduates, and of students (as well as parents) investing time and tuition dollars in post-secondary education. Evaluation and quality assurance for universities in Europe, for example, is an important part of the Bologna Process.⁸ Quality assurance organizations in the

⁴ Throughout this report, the term “faculty” is used to represent tenured and tenure-track faculty members, lab instructors, lecturers and senior lecturers, limited term appointments, and sessional instructors.

⁵ While the main text of this report, admittedly, contains more than 8000 words, it is relatively brief, given the vastness of the literature on the relevant questions for SFU’s project on student evaluation of teaching and courses.

⁶ At the time of writing (January 2012), the project team members include: Corinne Pitre-Hayes (Project Manager), Stephen Spector (SCUTL Chair), Hui Niu (Researcher, Instrumentation), Johanne Provençal (Researcher and Writer), and Chris Groeneboer (Manager, Applied Research, Teaching and Learning Centre, SFU).

⁷ Chen and Hoshower (2003, p. 71) cite, for example, studies by Seldin (1985), Abrami (1989), Wagenaar (1995), Abrami *et al* (2001), and Hobson and Talbot (2001) on the international use of what is widely referred to in the research literature as “student evaluation of teaching” (SET) or “student evaluation of teaching effectiveness” (SETE).

⁸ See the Council of Europe (http://www.coe.int/t/dg4/highereducation/ehea2010/bolognapedestrians_en.asp): “The Bologna Process is a European reform process aiming at establishing a European Higher Education Area by 2010” and follows from the 1999 Bologna Declaration, which “states the following objectives:

- adoption of a system of easily readable and comparable degrees;

United Kingdom (UK)⁹ and Australia¹⁰ include in their mandates evaluation of teaching and learning, and the Council for Higher Education Accreditation in the United States (US) and the Association of Universities and Colleges of Canada also support quality assurance.¹¹ In Canada, there are post-secondary education quality assurance measures within the provincial jurisdiction for education in several provinces.¹² The following is a useful description of the context for universities in Canada:

Global competition necessitates that, as Canadian institutions, we maintain our competitive edge and remain cognizant of major international developments beyond our borders; in particular, those shepherded by the Bologna process in the European Union, the Programme on Institutional Management in Higher Education (IMHE), and the Forum on Higher Education of the Organization for Economic Co-operation and Economic Development (OECD). It also necessitates that we keep an open mind and ear to what competencies are required for our students to be successful as life-long learners and professionals anywhere they choose to fulfil their aspirations.¹³

An important part of quality assurance for universities, as can be expected, is in the area of teaching or instruction. As Algozzine *et al* report, “By far, the most common method of evaluating instruction in university classes is to have students provide feedback on instructional ‘effectiveness’ using rating scales”¹⁴ and citing Berk,¹⁵ the authors note that, “student ratings have had a greater impact as a source of evidence for assessing teaching than all of the remaining dozen sources combined.”¹⁶ That said, precisely because of the significance of student evaluation of teaching and courses, it has been subject to scrutiny and criticism (if not cynicism) both informally – by instructors and faculty, as by administrators and students – and more formally investigated by researchers either within their own disciplines or by those in the field of higher education. As a consequence, a significant research literature on student evaluation of teaching (and courses) has formed in the past 30 years¹⁷ and there is much that it can offer not only to the project team in updating the university’s instrument for student evaluation of teaching and courses, but also, for faculty, administrators, and students as they use the instrument and interpret the data it produces. An overview of key issues from the research literature is presented, below, with focus on some of the most significant of these issues in the sections that follow.

-
- adoption of a system essentially based on two main cycles, undergraduate and graduate;
 - establishment of a system of credits – such as in the ECTS [European Commission Transfer and Accumulation];
 - promotion of mobility by overcoming obstacles to the free movement of students, teachers, researchers and administrative staff;
 - promotion of European co-operation in quality assurance;
 - promotion of the necessary European dimensions in higher education” (2011).

⁹ See the Quality Assurance Agency for Higher Education: <http://www.qaa.ac.uk/>

¹⁰ See the Tertiary Education and Quality Standards Agency: <http://www.teqsa.gov.au/>. Also, as Gapp and Fisher describe, “Since the mid-1980s the university sector in countries such as Australia has seen increasing political interventions, an outcome of which has been the demand to implement course and teaching evaluations. These government initiatives have clearly been made a requirement of various funding models” (2006, p. 156).

¹¹ Further, as Hallinger notes, “Although the use of accountability tools in higher education began as a largely ‘Western’ phenomenon, in recent years their adoption has become global in scale. The impact of quality assurance processes driven by government and accreditation agencies has been further accentuated in recent years by the dissemination of World University Rankings” (2010, p. 253).

¹² In British Columbia, see the BC Education Quality Assurance organization: <http://www.bceqa.ca/>

¹³ See Saroyan 2010, p. 103.

¹⁴ See 2010, p. 28. Here, Algozzine *et al* also cite the following: “Adams, 1997; Algozzine *et al.*, 2004; Berk, 2006; Blunt, 1991; Braskamp, Caulley, & Costin, 1979; Bryson, 1974; Centra, 1977; Cohen, 1981; Darby, 2007; d’Apollonia & Abrami, 1997; Feldman, 1989; Platt, 1993; Remmers, 1927; Rifkin, 1995; Soyjka, Gupta, & Deeter-Schmelz, 2002; Sproule, 2000; Stary, Derry, & Wright, 1973; Theall, Abrami, & Mets, 2001; Theall & Franklin, 1990” (2010, p.28).

¹⁵ See Berk 2006, p.15.

¹⁶ In *Thirteen Strategies to Measure College Teaching*, Berk (2006, p. 14) identifies the following 13 “potential sources of evidence of teaching effectiveness: (1) student ratings, (2) peer ratings, (3) external expert ratings, (4) self-ratings, (5) videos, (6) student interviews, (7) exit and alumni ratings, (8) employer ratings, (9) administrator ratings, (10) teaching scholarship, (11) teaching awards, (12) learning outcome measures, and (13) teaching portfolio” (cited in Algozzine *et al*, 2010, p. 28).

¹⁷ As an example, in the literature review by Al-Issa & Sulieman (2007), they found 2988 articles on student evaluation of teaching (following the widely cited review by Cashin in 1988, which found more than 1300 articles).

2.2.1 THE MULTI-DIMENSIONALITY OF TEACHING AND THE QUESTION OF “GOOD TEACHING”

The complexity and multi-dimensionality of teaching are widely recognized by scholars and practitioners in the field of education, as by faculty in their experiences in university or college classrooms.¹⁸ It is perhaps not surprising, then, that “higher education has yet to establish a universally accepted definition of the characteristics and skills necessary for teaching excellence.”¹⁹ This is in part because what is expected from a “good teacher” can “vary from faculty to faculty since different are the needs, the disciplinary contents and the possible practical activities that show an applicative use.”²⁰ In one study, “components of effective teaching identified from analysis of student ratings include six common dimensions...to as many as twenty-eight dimensions.”²¹ An understanding of teaching excellence clearly extends beyond the questions used on student evaluation of teaching forms and it is worthwhile to note that:

Although researchers use different terms, there is a consensus regarding six dimensions [of “good teaching”]: rapport with students, course value, course organization and design, fairness of grading, difficulty, and workload. Ramsden (1992) lists six key principles of effective teaching in higher education as: (a) Interest and explanation, (b) Concern and respect for students and student learning, (c) Appropriate assessment and feedback, (d) Clear goals and intellectual challenge, (e) Independence, control and active engagement, and (f) Learning from students. Feldman (1976) identified 22 constructs of effective teaching, concluding that the two most highly rated dimensions were stimulation of interest and clarity of presentation.²²

Perhaps more important than common questions on student evaluation of teaching forms is, as Ghedin and Aquario remind us, that how a “good teacher” is defined depends very much on *conceptions about teaching*²³ and this holds for faculty and students, both.

As the establishment of teaching centres within universities has initiated discussion about teaching and broadened faculty approaches to and understandings of teaching,²⁴ studies in the field of higher education have attempted to identify and characterize faculty approaches, beliefs, and theories about teaching. The typology developed by Ramsden, for example, identifies what he refers to as “different theories of teaching represented in lecturers’ attitudes to teaching and their instructional strategies” and these are: teaching as “telling and transmission...teaching as organising student activity...[and] teaching as making learning possible”²⁵ Similarly, Pratt *et al* developed a “Teaching Perspectives Inventory”²⁶ based on a series of 13 studies in which faculty classified their approach to teaching according to learning focus, with five approaches identified as follows:

- (1) The Transmission Perspective: Effective Delivery of Content;
- (2) The Apprenticeship Perspective: Modelling Ways of Being;
- (3) The Developmental Perspective: Cultivating Ways of Thinking;
- (4) The Nurturing Perspective: Facilitating Self-Efficacy; and
- (5) The Social Reform Perspective: Seeking a Better Society.²⁷

¹⁸ See, for example, Ghedin and Aquario 2008.

¹⁹ See Arreola 2007, p. 98.

²⁰ See Ghedin and Aquario 2008, p. 595.

²¹ See Paulsen 2002, p. 8.

²² See Khandelwal 2009, p. 299. Khandelwal also notes that, “Murray, Rushton, and Paunnen (1990) reported that forty to seventy percent of the variance in student teacher ratings could be accounted for by six personality traits, namely leadership, extroversion, liberalism, supportiveness, intellectual curiosity, and changeableness” (p. 300).

²³ See Ghedin and Aquario 2008. The authors also cite studies by Kember and Wong (2000) and Lecouter and Del Fabbro (2001).

²⁴ See Knapper 2010, p. 238.

²⁵ See Ramsden 2003, p. 106-110.

²⁶ See <http://teachingperspectives.com/>

²⁷ See Entwistle 2010, p. 30.

Underlying each of the above characterizations of teaching are assumptions and perspectives not only about the purpose(s) of higher education, but also, about the roles and responsibilities of both teacher and student. Particular attention has been given in recent years to questions of student learning and as a consequence, there has been a notable shift in higher education from a “teacher-centred” to a “learner-centred”²⁸ classroom and curriculum.²⁹ Among the contributing factors in this shift have been changed views about and approaches to knowledge (where, how, and by whom “knowledge” is produced, legitimized, shared, etc.) alongside growing understandings about student learning (in particular, recognizing the increasing cultural and socio-economic diversity of the student population), an appreciation for different learning styles, and implications of the 21st century technological and media environment.

2.2.2 UNDERSTANDINGS ABOUT STUDENT LEARNING AND THE SHIFT TO LEARNER-CENTRED CURRICULUM

While there are serious concerns about student learning in college and university, Harris and Cullen (with their view shared by many others) note, “One of the oddities of the tradition of higher education is that professors are rarely provided any instruction or professional development in the role that represents a major element of their responsibility: teaching.”³⁰ Yet, teaching faculty have a great deal of experience encountering issues of student learning. Weimer, in a somewhat light-hearted but also sincere description, presents the following as the most common concerns that faculty raise in regard to students as learners:

- Students are passive
- Students lack confidence as learners
- Many students lack the basic study skills necessary to succeed in college
- The only thing that motivates students are grades, points, and marks.³¹

²⁸ It should be noted that the term “student-centred” can also be used here. In this document, “learner-centred” is being used for reasons that are summarized well by Weimer: “Being student-centered implies a focus on student needs. It is an orientation that gives rise to the idea of education as a product, with the student as a customer and the role of the faculty as one of serving and satisfying the customer. Faculty resist the student-as-customer metaphor for very good reasons. When the product is education, the customer cannot always be right, there is no money-back guarantee, and tuition dollars do not ‘buy’ the desired grades. Being learner-centered focuses attention squarely on learning: what the student is learning, how the student is learning, the conditions under which the student is learning, whether the student is retaining and applying the learning, and how current learning positions the student for future learning” (2002, p. xvi).

²⁹ Similar to Pratt *et al*, though focusing on a teacher- or student-centred spectrum of teaching, there is also Trigwell’s description of “Five qualitatively different approaches to teaching...Approach A: Teacher-focused strategy with the intention of transmitting information to students...Approach B: Teacher-focused with the intention that students acquire the concepts of the discipline...Approach C: A teacher-student interaction strategy with the intention that students acquire the concepts of the discipline...Approach D: A student-focused strategy aimed at students developing their conceptions...Approach E: A student-focused strategy aimed at students changing their conceptions” (2010, p. 117)

³⁰ See Harris and Cullen 2008, p. 58. Although part of graduate training often involves opportunities for teaching assistant positions and professional development offerings to support faculty in teaching are available on most campuses, doctoral-level programs are widely recognized as placing focus on training for research skills within a particular discipline. This is likely a contributing factor for “faculty on average hav[ing] unrealistically high self-perceptions of their own teaching effectiveness” (Khandelwal 2009, p. 299, citing Marsh and Dunkin 1997, p. 276). Knapper identifies several problematic areas in university teaching: [1] Teaching remains overwhelmingly didactic and reliant on traditional lectures, and assessment methods are often trivial and inauthentic; [2] Curriculum development relies far too much on disciplinary tradition and faculty interests, rather than on student and societal needs; [3] There is still a ‘tyranny of the academic disciplines’ which mitigates against integration of knowledge and insights from different fields; [4] Evaluation of teaching effectively and learning outcomes is often superficial” (2010, p.238).

³¹ The description by Weimer, though certainly not to be taken as a fair description of all students, is quite likely to be familiar to many faculty: “Based on my own experience as an educational developer, [faculty] would identify one or several of these four problems. First, *students are passive*. They want to sit back and have education ‘done unto them,’ and they hope the experience will be pleasant and painless...They want the teacher to write the material on the board, or better yet show it on a PowerPoint presentation that can conveniently be downloaded. They really like it when teachers provide complete sets of notes from class. Next, *students lack confidence as learners*. They do not like it when they have to make decisions related to learning. They want teachers to spell out exactly what they should do. Teachers should specify paper length, appropriate fonts, margin size, and the number of references for the bibliography. Students want to know whether they should write in the first person, or include examples or quotations. They want teachers to ‘go over’ what will be on the test, detailing those sections of the text that should be

Although Weimer recognizes the challenges that faculty confront in regard to student learning, she also calls on faculty to consider how teaching can contribute to these difficulties, and that research on student learning can be drawn on to understand why these kinds of difficulties occur and ways to overcome them. In turning to the research for possible ways forward, Meyer describes, for example, a shift in the late 1970s that resulted in part from asking students about their own experiences of learning, making reference in particular to the widely-cited research by Marton and Säljö, who, “disturbingly demonstrated...that students differed from one another in their learning intentions...reflecting a categorisation not of students themselves but rather of the *variability in what they do and why they do it* in terms of learning process and underlying intention” (italics in original).³² An important finding in this research is the significance of “metalearning” in which students, becoming aware of their learning, gain “a recognition of self in relation to ‘learning’ in...[a particular] context and a consideration of what, in their own minds, they actually do when they are ‘learning’ in that context.”³³ A similar principle is also significant for faculty in their awareness of their teaching.

A learner-centred classroom requires awareness by faculty of how their teaching supports learning.³⁴ Perhaps one of the most promising aspects of a learner-centred approach to teaching is that, “When the focus is on student learning, the assessment of teaching takes into account what students are doing because it is assumed that they are [or can be] active participants who are engaged in learning.”³⁵ Weimer again provides a useful perspective on what this means in the day-to-day realities of the classroom, suggesting five changes that are needed in post-secondary teaching practice in order to overcome the challenges that faculty commonly confront:

- (1) *The balance of power*: Motivation to learn is greatly affected if students feel a sense of empowerment in their learning.

the focus on in-depth study. Sometimes they find it difficult to say for sure whether or not they understand something. Third, faculty might observe that *many students lack the basic study skills necessary to succeed in college*. They struggle to read technical material, even texts that have become simpler and better organized than they used to be. They depend on a limited repertoire of study strategies—flash cards, recopying notes, and underlining material in texts. If those strategies fail, students still use them, thinking that the only option is to use them more. Often they write and calculate poorly. Finally, faculty might complain that *the only thing that motivates students are grades, points, and marks*. Students rarely demonstrate any intellectual curiosity. They will volunteer but only if the learning opportunity involves points, even just a few of them. They engage in intellectual dialogue with passion only when the argument involves the possibility of more points or an increase in partial credit (2010, p. 81-82).

³² See Meyer 2010, p. 192 (citing Marton and Säljö 1976). The categorization the authors refer to is that of *deep learning versus surface learning*. Lindblom-Ylänne describes, for example, how a student “applying a surface approach to a reading assignment concentrates on the text itself. A deep approach, on the other hand, is based on a genuine interest in the subject matter and the aim is interpreting the meaning of the text...Previous research has shown that the deep approach to learning is more likely to be related to higher-quality learning outcomes than a surface approach” (2010, p. 64).

³³ See Meyer 2010, p. 200. Meyer also notes, however, that teaching faculty often “require something to prompt [in students] an empowering initial realisation of self as learner in process terms, the emphasis being on ‘what am I doing?’ rather than ‘who am I?’ The distinction is crucial because empowerment for students begins with the realisation that their learning processes can vary” (2010, p. 200).

³⁴ Kinzie notes, for example, the need to understand what students expect from their experience in order to use instructional approaches that support them in becoming intentional learners able to reach intended outcomes. Kinzie also notes that, “what faculty members emphasize and think is important to learning can influence what students do” (2010, p. 143). In the same vein, Meyer offers the reminder that from “assessment cues (in lectures) and the ‘ticky box’ multiple-choice test at the end, a student forms a temporal ‘cram and dump’ conception of ‘learning’” and she provides the following excerpt from a student participating in one of her studies: “I’m more focused on achieving a high mark than I am in gaining more knowledge. If I were to shift more focus towards placing value on the learning itself, high marks might come as an added benefit. ... I harbour some resentment towards students who consistently memorise information without understanding it, write it down on an exam without really thinking, and then forget it all afterwards, but I am also guilty of the same practice” (2010, p. 195).

³⁵ See Higgerson 2006, p. 36. See also Combs *et al* (2008) and their discussion of the importance – for faculty and students – of clear learning objectives and the usefulness of using a learning objectives tool, both for student evaluation of teaching and for student learning: “Students benefit directly from the use of this learning objective tool in several ways. The use of the tool requires stated learning objectives that clarify what the course is to deliver, which contributes to students’ understanding of what the instructor views as the important components of the course content. The identification of learning objectives also helps show students how the different course elements link to one another” (p. 90).

- (2) *The function of content*: The relationship between course content and the learner is significant to the learning process. Here, constructivist approaches to learning are useful for understanding how “learners’ actively construct...their own knowledge rather than passively receiving information transmitted to them from teachers and textbooks.”³⁶
- (3) *The role of the teacher*: Faculty expertise, rather than being exclusive, encourages and legitimizes different ways for students to interact with and relate to course content.
- (4) *The responsibility for learning*: Many students, by the time they reach college or university, have become dependent learners in teacher-centred learning environments in the sense of depending on the teacher “to identify what needs to be learned, to prescribe the learning methods, and...to assess how well they have learned,” which runs counter to the development of “independent, autonomous learners who assume responsibility for their own learning.”³⁷ These skills must be taught, however, and require an understanding (by teachers *and* students) of the responsibilities of the teacher for teaching and the learner for learning.
- (5) *Evaluation purpose and processes*: Students learn throughout their education to be motivated to focus on learning material on which they will be tested or otherwise assessed. It is important for teachers and students, both, to consider the implications of this approach to evaluation and to build in students an ability to conduct a self-assessment of the extent to which they do or do not understand what is being taught.³⁸

Weimer cautions, however, that some level of student resistance can be expected in shifting from a teacher-centred to learner-centred classroom and recommends an article that she reads before the start of every course.³⁹ That said, it is also worth noting, as Prosser reports from a series of studies, that teachers who adopt teacher-focused approaches to teaching have students tending toward surface approaches to learning; while teachers who adopt student-focused approaches to teaching have students tending toward deeper approaches to learning.⁴⁰

2.2.3 ISSUES OF VALIDITY AND RELIABILITY IN STUDENT EVALUATION OF TEACHING

Although, as Arreola notes, “Decades of research have demonstrated that tools used in faculty evaluation systems, especially student rating forms, can be designed to be valid and reliable,” there remain persistent challenges of validity and reliability in developing and implementing instruments for student

³⁶ See Weimer 2002, p. 13, citing Stage, Muller, Kinzie and Simmons 1998, p. 35.

³⁷ See Weimer 2002, p. 15.

³⁸ See Weimer 2002, p. 8-15.

³⁹ Weimer recommends and quotes from “Navigating the bumpy road to student-centered instruction” by Felder and Brent: “It’s not that student-centered instruction doesn’t work when done correctly—it does, as both the literature and our personal experience...richly attest. The problem is that although the promised benefits are real, they are neither immediate nor automatic. The students, whose teachers have been telling them everything they need to know from the first grade on, don’t necessarily appreciate having this support suddenly withdrawn” (Felder and Brent 1996, p. 43, cited in Weimer 2002, p. 150). Edström also provides useful perspectives on challenges of revisiting student-teacher responsibilities: “The practice to rate teachers, and the blame-the-student thinking displayed by teachers, appear as two sides of the same coin. A student may attribute success or failure to good or bad teachers, much as a teacher attributes the results to good or bad students. These thought patterns allow students and teachers, respectively, to focus on the other party’s contribution in the teaching and learning process and shy away from discussing (or even seeing) their own responsibilities. Changing the course evaluation practices means challenging these comfortable positions and, therefore, resistance may be expected [from teachers and students]” (2008, p. 104).

⁴⁰ See Prosser 2010, p. 130. See also Frick *et al* (2009): “First Principles of Instruction are relevant to complex learning of authentic, real-world, whole tasks. Based on a synthesis of instructional design theories, Merrill (2002) claimed that student learning will be promoted when: (1) instruction is problem- or task-centered, (2) student learning is activated by connecting what they already know or can do with what is to be newly learned, (3) students are exposed to demonstrations of what they are to learn, (4) they have opportunities to try out what they have learned with instructor coaching and feedback, and (5) they integrate what they have learned into their personal lives. If one or more of these First Principles are missing during instruction, Merrill argues that learning will be negatively impacted” (p. 116).

evaluation and teaching.⁴¹ Chen and Hoshower provide a helpful description of the distinction between reliability and validity studies:

Reliability studies (Marlin & Gaynor, 1989; Scherr & Scherr, 1990; Nimmer & Stone, 1991; Wachtel, 1998) generally address the question 'Are student ratings consistent both over time and from rater to rater?'. On the other hand, validity studies (Howard et al., 1985; Byrne, 1992; Tagomori & Bishop, 1995) address the questions 'Do student ratings measure teaching effectiveness?' and 'Are student ratings biased?'.⁴²

Although validity and reliability in student evaluation of teaching and courses is to be discussed in detail in a separate report,⁴³ some attention is given here to these issues. As a starting point, a more specific definition of validity is as follows:

[Validity is] the effectiveness of the test in representing, describing, or predicting the attribute that the user is interested in. *Content validity* refers to the faithfulness with which the test represents or reproduces an area of knowledge. *Construct validity* refers to the accuracy with which the test describes an individual in terms of some psychological trait or construct. *Criterion-related validity*, or *predictive validity*, refers to the accuracy with which the test scores make it possible to predict some criterion variable of educational, job, or life performance.⁴⁴

For student evaluation of teaching and courses, a key question for validity is, whether or not the instrument measures what it is intended to measure. This, then, is a key point of discussion for SFU faculty, students, and administrators, because if the instrument is to measure teaching effectiveness, a decision must be made, for example, about the components of effective teaching to include in questions (and options for responses) in the student evaluation of teaching instrument. Similarly, the findings in the research literature on both the components of effective teaching⁴⁵ and instruments used to measure effective teaching⁴⁶ are important to inform the update of the student evaluation of teaching instrument. As Marsh (2007) notes, "poorly worded or inappropriate items will not provide useful information, while scores averaged across an ill-defined assortment of items offer no basis for knowing what is being measured" (p. 321). In regard to reliability, the test phase of the development of the instrument is important for ensuring reliability in inter-rater results, re-test results, and internal consistency.⁴⁷ Although concerns regarding reliability and validity are important to recognize and address, as Chen and Hoshower note:

Overall, the literature supports the view that properly designed student ratings can be a valuable source of information for evaluating certain aspects of faculty teaching performance (Cohen, 1981; Marsh, 1984; Calderon et al., 1994).⁴⁸

Although these kinds of findings from research on student evaluation of teaching can be reassuring, there are also studies identifying sources of student bias that can affect validity and reliability of student evaluation of teaching and courses.⁴⁹

⁴¹ See Arreola 2007, p. 98.

⁴² See Chen and Hoshower 2003, p. 72.

⁴³ Hui Niu, the project team member leading the development of the instrument, will discuss validity and reliability in detail in her report.

⁴⁴ See Arreola 2007, p. 99 (citing Thorndike and Hagen 1969, pp. 163-177).

⁴⁵ Hence, the overview on pages 4-9 of this report.

⁴⁶ As noted, Hui Niu, the project team member leading the development of the instrument, will discuss findings in the research literature on instruments used for student evaluation of teaching.

⁴⁷ Inter-rater reliability refers to the consistency of results from different raters using the same instrument. Re-test reliability refers to the consistency of results from the same rater with the same instrument but in multiple tests. Internal consistency refers to items within an instrument being consistent in what they measure.

⁴⁸ See Chen and Hoshower 2003, p. 72.

⁴⁹ Khandelwal, for example, in a review of the literature notes that student evaluation of teaching can be affected by "students' prior motivation or desire to take the course (Marsh, 1984), anticipated grades (Howard & Maxwell, 1982), workload (Greenwald & Gillmore, 1997; Marsh and Roche, 2000), course level (positive relationship - Braskamp et al., 1985), class size (negative relationship - Cashin & Slawson, 1977; Smith & Glass, 1980), and grading leniency of the instructor (Greenwald, 1997). Cashin's

Understanding student bias as a variable affecting evaluation of teaching is important for developing the instrument to be used and in interpreting the data it produces. Further, in preparing the questions used in the instrument and in interpreting the data, it is now understood from studies on student evaluation of teaching that there are some aspects of faculty teaching that students are not equipped to evaluate.⁵⁰ That said, in the absence of formal and careful evaluation of teaching, faculty can or “will be evaluated by hearsay evidence, gossip, an other shoddy means.”⁵¹ (p. 7).

It is perhaps useful in SFU’s endeavours to update the form and process for student evaluation of teaching and courses to be reminded that:

[P]ractically no campus is satisfied with the protocols it has in place for evaluating teaching. The quest for perfecting the teaching evaluation process is quite likely to be doomed to be unending as we fuss and fiddle with the terminology, the question wording, and compare strategies. But it is in the fussing and fiddling that we move toward a better understanding of what goes on in the college and university classroom.⁵²

Keeping in mind that although a *perfect* process for the evaluation of teaching and courses may not be an attainable goal, members of the SFU community are encouraged to hold in sight the purpose and promise of the project, which is to support faculty, administrators, and students in creating an environment that is rewarding and conducive to successful teaching and learning.⁵³

3 EVALUATION OF TEACHING AND COURSES

Hardré and Cox (among others) offer the reminder that “faculty performance is critical to the health of institutions of higher education and to the education of citizens.”⁵⁴ It is also important, however, that evaluation of teaching and courses is something “done with” rather than “done to” those involved in teaching.⁵⁵ It is necessary, therefore, for SFU faculty to be provided with multiple opportunities to be involved in updating the process and form used for student evaluation of teaching and courses at the university. Similarly, the SFU community is encouraged to consult this report for an overview of findings in the research literature to support their involvement in the development and responsible use of the revised instrument.

review (1988) concluded that student motivation (willingness to participate actively in the learning process) has the greatest positive influence on student satisfaction than any other instructional factor like grade expectations, sex of teacher/student, age of teacher/student, time of day, etc.” (cited in Khandelwal 2009, p. 300).

⁵⁰ The list that Pallett provides is representative of what is found elsewhere in the research literature: [1] The appropriateness of an instructor’s objectives; [2] The instructor’s knowledge of the subject matter; [3] The degree to which instructional processes and materials are current, balanced, and relevant to objectives; [4] The quality and appropriateness of assessment methods; [5] The appropriateness of grading standards; [6] The instructor’s support for department teaching efforts such as curriculum development and mentoring new faculty; [7] The instructor’s contribution to a department climate that values teaching” (2006, p. 56).

⁵¹ See Seldin 2006, p. 7.

⁵² See Halonen and Ellenberg 2006, p. 151.

⁵³ Although the purpose of this report is to support SFU faculty, administrators, and students, it is worthwhile to note that in the larger context of research on post-secondary education, universities have been criticized as having “forgotten their purpose; namely, creating educated adults who will take responsibility for their society” (Harris & Cullen 2008, p.57). The same authors, citing Bok, echo a widely shared concern about faculty as teachers, “that not enough attention is paid to pedagogy” (2008, p. 58), as faculty training tends to focus heavily on disciplinary knowledge and research. By the same token, in regard to students as learners, Meyer’s view is that part of the challenge for student learning “lay in the fact that most entering university students are not likely to have had an opportunity to talk to their teachers (or anybody else) at school about *how* they typically went about learning *in process terms*. For the most part, students have never really thought about themselves in this way and, if asked, they generally experience difficulty, beyond habitual or preferential activities, in describing what *they know about themselves as learners*” (2010, p. 200, italics in original).

⁵⁴ See Hardré and Cox 2009, p. 383. In discussing the importance of evaluating faculty work, the authors also cite Fairweather (1996) and Gappa, Austin and Trice (2007).

⁵⁵ See MacBeath and McGlynn (2002, p. ix), cited in Burden (2010, p. 111).

One of the key findings that surfaces repeatedly in the research literature is the importance of using *a set* of evaluative processes. As mentioned earlier in this report, there are several approaches that are commonly used in the evaluation of teaching.⁵⁶ Although the focus of this report is on student evaluation of teaching and courses, some attention is given here to two approaches to teaching evaluation that are most often used to complement student evaluation of teaching and courses: (1) peer evaluation of teaching, and (2) teaching portfolios. It should be noted, however, that there is a need for validity of “all indicators of teaching effectiveness, not just SETs, [to] be systematically examined before they are actually used.”⁵⁷

3.1 PEER EVALUATION OF TEACHING

Peer evaluation of teaching is most commonly conducted through classroom observations by faculty in the same department as the faculty member being evaluated. Peer evaluation of teaching can be valuable because there are clearly aspects of teaching that students cannot be expected to be able to assess,⁵⁸ such as “mastery of content, course goals, course organization and materials.”⁵⁹ Evaluation by peers is also understood as fundamental to academic work, as faculty are accustomed to peer review in submitting material to journals and scholarly publishers, for example. As Paulsen notes:

Faculty expect public review of the methods and products of their research. In contrast, methods and products of teaching are rarely discussed or shared with peers. Just as the quality of research improves due to dialogue and debate among disciplinary peers, so would the quality of teaching benefit from similar opportunities.⁶⁰

As with peer review for publication, however, there are challenges to overcome in terms of the ability and willingness of a peer to provide informed, fair, and helpful evaluation. That said, peer evaluation of teaching, which has for some time been common in the UK, is seen by much of the teaching faculty, “as an integral part of their own professional development as teachers and [they] see value from the process, both in engaging as the observer and the one being observed.”⁶¹

Peer observation and evaluation of teaching offers the following:

- Insight into what helps learners to learn and what happens in effective teaching sessions;
- Feedback on individual teaching skills and style;
- Discussion, collaboration, and the exchange of ideas;
- Mutual support between colleagues;
- Earmarked ‘quality time’ to talk about learning and teaching;
- Feedback on piloting a new idea, method, or solution to a problem;
- Triangulation with other evaluative procedures, e.g. student perception questionnaires, module and program evaluations;
- Focused reflection on teaching sessions;
- The opportunity to see exemplary practitioners at work.⁶²

⁵⁶ See footnote 13 of this report for the following list: “(1) student ratings, (2) peer ratings, (3) external expert ratings, (4) self-ratings, (5) videos, (6) student interviews, (7) exit and alumni ratings, (8) employer ratings, (9) administrator ratings, (10) teaching scholarship, (11) teaching awards, (12) learning outcome measures, and (13) teaching portfolio” (Berk 2006, p. 14, cited in Algozzine *et al* 2010, p. 28).

⁵⁷ See Marsh 2007, p. 343.

⁵⁸ See footnote 47.

⁵⁹ See Paulsen 2002, p. 10.

⁶⁰ See Paulsen 2002, p. 10. Paulsen notes that, “Specialists in teaching and its evaluation also agree that the work of an individual faculty member is valued more when it has been subjected to rigorous peer review” and suggests that this is a contributing factor to research being valued more highly than teaching (2002, p. 10).

⁶¹ See Fullerton 1999, p. 220.

⁶² See Fullerton 1999, p. 221-222.

Research on peer evaluation of teaching also provides useful guidelines. According to Harris and Cullen, for example, the “three components that should frame the overall evaluation are: [1] teaching philosophy or pedagogical stance; [2] competency in the discipline; and [3] teaching craft.”⁶³ Although a thorough overview of peer evaluation of teaching is beyond the scope of this report, it is an important component in the evaluation of teaching and courses, and therefore warrants further attention as a complement to student evaluation of teaching and courses.

3.2 TEACHING PORTFOLIOS

Teaching portfolios were first used in Canada in the 1970s and have since become widely used across North America.⁶⁴ Items in a teaching portfolio can include the following:

- Courses, modules and units taught;
- Information about the context of your teaching;
- Self-authored documents and/or study guides;
- Module descriptions including aims and objectives;
- Curriculum development work;
- Problem-based learning cases;
- AV materials;
- Teaching strategies;
- Student handouts or workbooks;
- Documentation to support practical teaching;
- Instruments for assessment of student learning;
- Analysis of examination results;
- Student evaluation instruments, views and results;
- Quality assurance reports;
- External examiners’ reports;
- Examples of student work;
- Video, audio and photographic material;
- Computing and IT material;
- Written commentary from observed teaching sessions;
- Details of courses and workshops attended on teaching and learning;
- Educational publications.⁶⁵

A teaching portfolio serves as a “personal record of achievement and professional development” and contains a careful selection of material collected and developed over the course of a faculty member’s career and is a way of demonstrating “a level of attainment, progression, professional development and/or achievement.”⁶⁶ Teaching portfolios (referred to also as “teaching dossiers”) have also become required in applications for faculty positions and are developed as a component of faculty earning credentials in higher education learning and teaching, which has become common in the UK.⁶⁷ A detailed overview of teaching portfolios is not within the scope of this report but teaching portfolios also warrant further attention.

⁶³ See Harris & Cullen 2008, p. 62.

⁶⁴ Citing Kappa (1995), Fry and Ketteridge identify Queen’s University as a “one of the pioneers of the teaching portfolio” (1999, p. 235).

⁶⁵ See Fry and Ketteridge 1999, p. 236-237.

⁶⁶ See Fry and Ketteridge 1999, p. 235.

⁶⁷ See The Higher Education Academy in the UK, for example (<http://www.heacademy.ac.uk/ukpsf>): “The UK Professional Standards Framework (UKPSF) for teaching and supporting learning is for institutions to apply to their professional development programmes and activities thereby demonstrating that professional standards are being met. We support and guide both institutions and individuals as they engage with the UKPSF...The UK Professional Standards Framework provides a general description of the main dimensions of the roles of teaching and supporting learning within the HE environment. It is written from the perspective of the practitioner and outlines a national Framework for comprehensively recognising and benchmarking teaching and learning support roles within Higher Education.”

3.3 STUDENT EVALUATION OF TEACHING AND COURSES

Although the sections to follow in this document provide details of key concerns and research findings in regard to particular aspects of student evaluation of teaching and courses, the widely-cited work of Centra provides a useful overview of the historical trajectory of student evaluation of teaching and courses, and is therefore worth quoting at length:

In the universities of medieval Europe...students evaluated teachers in two ways. First, a committee of students appointed by the rector made sure they covered each topic by the date specified in advance; any irregularities were reported to the rector, who fined the professor for each day he (in those days, always a "he") had fallen behind. Second, students paid fees directly to each teacher (called the *collecta*). Teachers' salaries, therefore, were determined by the number of students attending their classes (Rashdall, 1936). Today's student questionnaires seem tame by comparison. The modern era of student evaluations can be broken roughly into four periods: the thirty-year period preceding 1960, the 1960s, the 1970s, and the period from the 1980s to the present. Before 1960, most of the research on student evaluations was conducted by Herman Remmers and his colleagues at Purdue University. Although Remmers and other researchers...promoted their use, student evaluations were rarely used in colleges and universities until the 1960s. The student protests that rocked so many [US] campuses in the last half of the decade were in reaction not only to the Vietnam War and related national policies but also to policies in effect on their campuses. An irrelevant curriculum and uninspired teachers were among frequently heard student complaints...At many universities, students administered their own rating systems—often haphazardly—and published the results. More often, institutions began to develop their own student evaluation systems, which faculty members could use if they wished. In the 1960s, faculty members usually volunteered to administer the form in their courses themselves, so that they might make necessary improvements on their own. Administrative use of the results was infrequent because few institutions had centrally administered student rating systems. Interest in pertinent issues—the validity of the ratings, in particular—inspired a new wave of research and the 1970s became the golden age of research on student evaluations. Well designed studies investigated questions of bias, validity (Do student evaluations really measure teaching effectiveness?), and utility (Can they help improve instruction?). The generally favourable findings helped support the use of the evaluations for tenure and promotion decisions as well as for institutional improvement. During the 1970s, more and more institutions adopted them, with a majority reporting their use by the end of the decade (Centra, 1979). Department heads ranked student evaluations among the top three sources of information on teaching effectiveness (colleagues and department [end of pg 50] heads were the other two) and, moreover, believed them to be the most important source (Centra, 1979). The fourth period, from the early 1980s to the present, has included the continuing refinement of the research findings, and a series of reviews and meta-analysis has substantiated the findings on important issues.⁶⁸

With this historical context in mind, the sections that follow in this report take up issues of student evaluation of teaching and courses that are particular to the concerns of faculty, administrators, and students.

4 STUDENT EVALUATION OF TEACHING AND COURSES: ISSUES FOR FACULTY, ADMINISTRATORS, AND STUDENTS

One criticism (shared by faculty, administrators, and students) of student evaluation of teaching and courses is that, if used ineffectively, it serves as little more than "a 'fire alarm' function."⁶⁹ If developed and used well, however, there are also advantages to student evaluation of teaching: "[1] Students know best what was effective for their learning; [2] Students observe the whole class; [3] Student samples are larger [than is the case with peer evaluation]."⁷⁰ It is also worth noting that, "there are generally high correlations between students' and faculty members' evaluations of teaching."⁷¹

⁶⁸ See Centra 1993, p. 49-50.

⁶⁹ See Edström 2008, p. 95.

⁷⁰ See Ackerman *et al* 2009, p. 27.

⁷¹ That said, the authors also note that: "Student and faculty views of one another depend on how much they agree on the characteristics of excellent instructors (Goldstein & Benassi, 2006). Interviews with students reveal the attributes most associated by students with outstanding teaching to be rapport, delivery, fairness, knowledge and credibility, organization, and preparation (Faranda & Clarke, 2004)...In addition, some dimensions valued by faculty members such as facilitating the achievement of key

In order for the instrument used at SFU for student evaluation of teaching and courses to be as effective as possible in supporting faculty teaching and student learning, it is important for the SFU community to be aware of the issues and challenges particular to faculty, administrators, and students in using a student evaluation of teaching instrument and in interpreting the data it produces. This section of the report summarizes key findings in the research literature in order to support effective and responsible student evaluation of teaching and courses at SFU.

4.1 ISSUES FOR FACULTY

In the literature on student evaluation of teaching and courses, Ackerman *et al* identify the two main concerns by faculty as centring around: (1) validity (and the extent to which evaluations are affected by student bias); and (2) the use of data in tenure and promotion, given “a lack of sophistication among personnel committees who use student ratings.”⁷² These two concerns are discussed in this report, but first, it is worth mentioning common perceptions (and misperceptions) among faculty about student evaluation of teaching and courses reported in the research literature, as they provide an indication of what is often the starting point (and first obstacle) in working toward effective and responsible use of student evaluation of teaching.

Student evaluation of teaching and courses is sometimes seen by faculty as a “popularity contest...designed to ‘win votes’...[using a] surveillance approach in relation to students’ expectations [that] leads to outcomes that are not always educationally sound.”⁷³ This view presents valid concerns that have been taken up in the research literature. For example, the “popularity” of a faculty member among students (and also among other faculty)⁷⁴ can affect ratings and faculty members have been known to employ tactics to “win” votes.⁷⁵ In terms of unsound outcomes, one of the most widely cited is unspoken leniency in grading traded for higher student evaluation of teaching and courses. These perceptions of flaws (and abuses) in student evaluation of teaching and courses are valid. Faculty misperceptions, however, rest in dismissing student evaluation of teaching and courses altogether, rather than taking into account that: (1) the instrument can be designed to reduce “noise” (which can be expected in any evaluation method) through careful crafting of the questions and options for responses; and (2) guidelines on interpreting the data can provide considerable support for responsible, effective and instructive use of student evaluation of teaching and courses.

4.1.1 RECOGNIZING POSSIBLE STUDENT BIAS

Before presenting an overview of sources of student bias discussed in the research literature, one key concern noted earlier in this report should be reiterated: a consumerist view of higher education is problematic in that it can result in student bias that reduces the learning and the experience of higher education to a product that is purchased by tuition dollars and for which students (as consumers) can be inclined to take a “customer satisfaction” approach to teaching and learning. This is problematic because

learning objectives and modeling rigorous thinking may be unrecognized by students (Buskist, Sikorski, Buckley, & Saville, 2002; Schaeffer, Epting, Zinn, & Buskist, 2003).” See Ackerman *et al* 2009, p. 22. With this, there is perhaps further support for the “meta-learning” suggested by Meyer (2010), the benefits of reflective teaching practice widely supported by researchers in the field of higher education, and the significance of communication between faculty and students about pedagogical approaches being taken.

⁷² See Ackerman *et al* 2009, p. 21. On these two items, the authors cite the work of Betoret (2007), Blackhart *et al* (2007), Clayson (2004), Clayson & Haley (1990), Heckert *et al* (2006), Marks (2000), Olshavsky and Spreng (1995), Simpson and Siguaw (2000), and Sojka *et al* (2002), among others.

⁷³ See Gapp and Fisher 2006, p. 157.

⁷⁴ This is discussed below in the subsection on student bias.

⁷⁵ See Halonen and Ellenberg 2006, p. 154-156. The authors provide the following list of what they refer to as “faculty follies:” (1) the “forget” to conduct the evaluation in their course(s); (2) they “taint” the evaluation by failing to provide enough time; (3) they “pander” (by handing out “treats” on evaluation day; (4) they “wheedle” (guilt-tripping students by noting that the evaluations affect their salaries); (5) they “script” (by priming students in how to respond on the evaluation), (6) they “barter, linger, over-respond, denigrate, ignore” (based on the nasty or overly harsh evaluations from a minority of students). The authors also identify “student follies” and “administrative follies” (these are presented later in this report).

a consumerist view of higher education tends to give little attention to such things as the level of effort and commitment required of students, the difficulty and the demands of the work to be undertaken by students, or in short, that although there are different views on the purpose(s) of higher education, post-secondary studies are, after all, meant to be challenging (and this does not always bode well for “customer satisfaction”).⁷⁶ This is not to dismiss the importance of ensuring quality in higher education. Indeed, maintaining quality standards is fundamental to the strength of higher education and the learning opportunities that colleges and universities provide. The issue is one, however, that deserves attention and discussion between faculty and students in the interest of supporting faculty teaching and student learning, and the goals of higher education more broadly.

According to the research literature, beyond the effects of a consumerist perspective, other factors that can bias student evaluation of teaching and courses include: “prior subject interest, expected grade and actual grade, reason for taking a course, workload and difficulty, class size, level of course or year in school, instructor’s rank, sex of instructor or student, academic discipline, purpose of ratings, administrative conditions, and students’ personality.”⁷⁷ Views on student bias are not uncontested, however. Marsh (2007), for example, cautions that the “literature on potential biases in SETs is frequently atheoretical, methodologically flawed, and not based on well-articulated operational definitions of bias” (p. 346) and he offers the following definition: “Bias exists when a student, teacher, or course characteristic affects the evaluations made, either positively or negatively, but is unrelated to any criteria of good teaching, such as increased student learning.”⁷⁸ The question is not whether the characteristics of the student, teacher or course have effects on evaluations – they inevitably do – but rather, the difficulty is in determining whether or not the effects of these characteristics (of the student, teacher, or course) on the evaluations are related or unrelated to the criteria of “good teaching” put forward in SET forms. Studies on student bias in evaluating teaching and courses have suggested, for example, that:⁷⁹

- Elective courses tend to receive higher ratings than required courses;⁸⁰
- Upper level courses tend to receive higher ratings than lower level courses;⁸¹
- The time of day that a course is offered can affect student ratings;⁸²
- The relationship between class size and ratings from student evaluation of teaching is debated,⁸³ Though smaller classes are more conducive to building rapport;⁸⁴
- There is a relationship between gender of the student, gender of the instructor, and ranking that the student gives the instructor;⁸⁵
- Ethnic/cultural background and beliefs can also affect student evaluation of teaching;⁸⁶

⁷⁶ Germain and Scandura also note, for example, that “Consumerism results in bias due to information not relevant to teaching competency, but important to students such as textbook cost, attendance policy, and the amount of homework” (2005, p. 58).

⁷⁷ See Buchert *et al* 2008, p. 398, citing Marsh and Roche’s review of the literature (1997, p. 1194).

⁷⁸ See Marsh 2007, p. 350 (citing Centra 2003).

⁷⁹ It should be noted here that the list provided is not comprehensive, but rather, reflects areas given the most focus in the research literature. It should also be noted that, as in other areas of research, findings are not always conclusive and different studies can provide conflicting results. That said, care and attention have been given here, in review of the research literature, to report on key findings that can support effective and responsible development and use of student evaluation of teaching and courses.

⁸⁰ See Scherr and Scherr (1990), Marsh and Dunkin (1992), and Pounder (2007).

⁸¹ See, for example, Braskamp *et al* 1985, March 1987, and Buchert *et al* 2008.

⁸² See Zabaleta, for example, whose study showed that, “instructors who teach between 10 am and noon receive the best evaluations” (2007, p. 61). See also, Husbands and Fosh (1993), as cited in Denson *et al* (2010).

⁸³ See, for example, Greenwald and Gillmore (1997), Fernandez *et al* 1998, and Pan *et al* 2007.

⁸⁴ See Pounder 2007.

⁸⁵ As Hobson and Talbot note, “A heated debate continues about the influence on student evaluations of a teacher’s sex and gender-role orientation (masculine, feminine, androgynous, or undifferentiated) (Basow and Howe 1987; Hobson 1997; Sidanius and Crane 1989). With respect to student characteristics, an interaction effect appears to exist between students and instructor’s sex/gender-role orientation” (2001). For example, Centra and Gaubatz, in their study of 741 classes, investigating male and female students in their evaluation of male and female instructors, found that “female instructors received a lower rating from male students” on several scales, while there was no significant difference in how the students of both genders evaluated male instructors” (2000, p. 26).

- Socioeconomic status of students can affect student evaluation of teaching;⁸⁷
- Generation gap issues need to be taken into account as a possible source of student bias;⁸⁸
- There is debate about instructor rank and student ratings, with evidence suggesting only slightly higher student evaluation of teaching for more senior faculty;⁸⁹
- Pre-existing knowledge about the instructor can affect student evaluation of teaching;⁹⁰
- Although there is ample research on discrimination against gay, lesbian, bi-sexual, and transgender persons, research specific to student evaluation of teaching is limited;⁹¹
- Instructor characteristics, including appearance⁹² and charisma, can affect ratings;⁹³
- Differences across academic disciplines can affect student evaluation of teaching;⁹⁴

⁸⁶ See, for example, Germain and Scandura: "Cultural beliefs strongly influence the value and behavior of the people who grew up in the culture, often without their being aware of it...More specifically, a student's cultural background can influence how they react to their learning environment. The effect of cultural context has long been recognized as having an affect on student learning (Halloway, 1988)" (2005, p. 63-64). Similarly, Al-Issa and Sulieman note that, "students' cultural and linguistic backgrounds affect their responses to SET [student evaluation of teaching]" (2007, p. 302). Smith also reports from a study of 31,768 evaluation forms that for "overall value of course and overall teaching ability...Black male faculty received the lowest mean score; Black female faculty received the second lowest mean scores (2008, p. 622)

⁸⁷ See Germain and Scandura (2005): "Socioeconomic status may also affect how students relate to faculty. From an economic standpoint, wealth or hardship of a student could affect grades obtained in the class and subsequent evaluations. Students who are on grants or scholarships are obligated to maintain A's or B's to maintain their benefits...[and] fearing the loss of financial aid might make students grade professors that give C's on course assignments more harshly" (p. 63).

⁸⁸ See Germain and Scandura (2005): "More mature students may be less judgmental with older faculty members. Similarly, younger students may be more likely to give good evaluations to young faculty members" (p. 63). Similarly, Sprinkle reports that older students are "more likely to take responsibility for their own learning and grade, rather than place the burden upon the professor/instructor (Nunn, 1994)...[and] younger students are likely to believe that younger professors/instructors are easier to relate to and more understanding of/sympathetic to their difficulties" (p. 286).

⁸⁹ See, for example, Marsh and Dunkin (1992) and Kogan *et al* (2010).

⁹⁰ See Germain and Scandura (2005): "[T]he degree to which the instructor is well known on campus or whether a student has had the instructor in a previous course...may affect rapport during the semester. Because the student has pre-existing knowledge about the instructor, faculty evaluations may be biased compared with students who have not had the professor previously" (p. 63). There are similar findings in Griffin (2001), who also cites studies by Kelley (1950) and Widmeyer and Loy (1988): "Students prompted prior to viewing a lecture that an instructor was either 'very warm' or 'rather cold' as a person...were enough to affect student ratings of the instructor in a positive or negative direction" (p. 535). That said, in a study by Buchert *et al* (2009), "students considered first impressions more important than professor reputation as determinants of their end-of-the semester evaluations...it appears that students form lasting impressions of faculty during the first 2 weeks of classes" (p. 397).

⁹¹ See Jennings (2010), for example: "Experimental projects have concluded that student assessments of speaker credibility were negatively affected when students perceived a male speaker as homosexual compared with when the speaker was perceived as heterosexual. Further, students believed that they learned less from homosexual male speakers than from those who were not perceived as homosexual (Russ, Simonds, and Hunt 2002). Ewing, Stukas, and Sheehan (2003) found that under some experimental circumstances students rated gay men and lesbian guest speakers lower than speakers whose sexual orientation was unspecified. However...these experimental studies only exposed students to single lectures by guest speakers. They did not study students' evaluations of instructors who were actually teaching the courses and with whom students had repeated contact, experienced multiple opportunities to observe instructors' full range of teaching behaviours, or came to know the instructor personally. Only one project has directly examined the relationship between instructor self-disclosure and student evaluations in a naturally occurring university class (Liddle 1997). In contrast to the two experimental projects by Russ, Simonds, and Hunt and by Ewing, Stukas, and Sheehan, Liddle's research analysed actual classroom evaluations of a course instructor and concluded that student evaluations of the instructor were not affected by instructor self-disclosure" (p. 327).

⁹² See, for example, Morris *et al*, whose study indicated that, "perceptions of instructor competence were highest...[with] formal professional attire...with casual professional attire a close second...and the lowest ratings for...casual attire" (1996, p.143). There is also a substantial research literature on the effects of appearance on likability, and as Gurung and Vespia note from their study, "By far the strongest indicator of self-reported learning [which is a common question in student evaluation of teaching and courses]...was the likeability of the professor. Likeability, in turn, was predicted by instructor attractiveness, approachability, and formality of dress" (2007, p. 8).

⁹³ The most widely cited example of student response to instructor charisma or "enthusiasm" in evaluating teaching and courses is what has become known as the "Dr. Fox" experiment in which a trained actor was "coached to present his topic and conduct his question and answer period with an excessive use of double talk, neologisms, non sequiturs, and contradictory statements. All this was to be interspersed with parenthetical humor and meaningless references to unrelated topics" (Naftulin, Ware, and Donnelly (1973). In the experiment, Dr. Fox received higher ratings from students than actual instructors.

⁹⁴ See, for example, Erdle and Murray's (1986) study on varying "teaching behaviours" varying between arts, social sciences, and natural sciences. Aside from student bias in response to teaching behaviours (the authors differentiate between "interpersonal orientation" and "task orientation"), there are also considerations here for the items included in the evaluation form and options for customization across faculties and departments.

- Courses with a heavy workload tend to receive lower ratings than courses with a workload that students perceive as reasonable for the course level;⁹⁵
- Whether the student is part-time or full-time can affect ratings;⁹⁶
- Perception about the purpose of the evaluation can affect ratings;⁹⁷ and
- There are issues to consider in regard to international students.⁹⁸

In addition to the above list of possible sources of student bias in evaluation of teaching and courses, one further concern commonly raised in the research literature is the relationship between students' expected grades (and/or actual grades) and their evaluation of teaching and courses.⁹⁹ In particular, the concern is in regard to the high correlation between grades and evaluation of teaching, for which the following three explanations are commonly given:

The validity explanation is that grades and evaluations are correlated because students learn more from superior teachers. In contrast, the leniency explanation is that students reward professors with high evaluations in exchange for high grades. Finally, preexisting student interest in course topics may explain both grades and teaching evaluations. Research on this topic continues, but clear evidence for one explanation has not emerged.¹⁰⁰

Another perspective on the relationship between grades and ratings is "the cognitive dissonance explanation [that] states that students attempting to reduce the dissonance caused by a poor grade are more likely to denigrate the teacher than their own ability, which leads to lower evaluations."¹⁰¹ Aside from concerns about the damage done to the integrity of assessment of teaching by an unspoken trade of high ratings on teaching and course evaluations for high grades, at issue here also is a broader concern about grade inflation in universities.¹⁰²

Awareness of possible bias in student evaluation of teaching and courses and knowledge of the research in this area can support the SFU community and the project team as the instrument for student evaluation of teaching and courses is updated, both in terms of deciding on questions to include and in the interpretation of the data that the instrument produces. Although details of the instrument itself are discussed in another report,¹⁰³ the studies on possible areas for student bias are useful for considering, for example, the inclusion of a question on a student's reason for taking a particular course. Similarly, in interpreting the data, taking into account, for example, variables ranging from course level, to gender, to disciplinary differences, are also important. Concerns about interpreting and using the data produced by student evaluation of teaching and courses tend to focus on issues for administrators or the university administration more broadly. This report therefore turns to key findings in the research on these issues,

⁹⁵ See, for example Marsh (1982), Butdsal and Bardo (1986), Jackson *et al* (1999), and Marsh and Roche (2000).

⁹⁶ See Denson *et al*, for example, who found in their study that, "part-time students rated satisfaction with quality of the course lower than full-time students" (2010, p. 346).

⁹⁷ Gapp and Fisher (2006) describe, for example, that an instructor's view on evaluation of teaching and courses can transfer to the students. Al-Issa and Sulieman (2007) note from their study that, "students believed that their assessments were an effective means of voicing their opinions about teaching, but that they were not fully aware of the implications of their evaluations for university administrators and teachers. This raises the question of whether students were motivated to take the evaluation seriously" (p. 304). Further, Burden notes that the evaluation process can be improved by "students learning *how* to do evaluations, or becoming 'more sophisticated evaluators'" (2010, p. 107, citing McKeachie 1997, p. 1223, italics in original). A study by Oliver *et al* (2008) provides an excellent example of revising (and clarifying) questions on a student evaluation form to support students' understanding of what they are asked to evaluate.

⁹⁸ Other than ensuring that students (international or otherwise) fully comprehend the questions on the evaluation form, Burden (2010) notes also, for example, that, "If students have...a positive attitude toward foreign countries, then it is easy for students to 'tick all the boxes down the right (i.e. choose a rating of 5 [out of 5]) without even reading the questions" (p. 105). Similarly, Al-Issa notes that students from some cultural backgrounds may not be "accustomed to 'passing judgment' on their teachers" (2007, p. 303).

⁹⁹ See Howard and Maxwell (1982), Olshavsky and Spreng (1995), McKeachie (1997), and Eiszler (2002), among others.

¹⁰⁰ See Boysen 2008, p. 218, citing the work of Marsh (1984).

¹⁰¹ See Boysen 2008, p. 218. As noted earlier, however, instructors are not immune to cognitive dissonance.

¹⁰² As Germain and Scandura note, "Grade inflation has become an issue in higher education; students' grades have been steadily increasing since the 1960's (Astin, 1998). In June 2001, a record 91 percent of Harvard seniors graduated with honors, and 48.5 percent of grades were A's and A-minuses (2005, p. 58).

¹⁰³ Hui Niu, the project team member leading the development of the instrument, is also providing a report.

which include administrator interpretation and use of data derived from student evaluation of teaching (including issues of validity and reliability), support for teaching and learning, and broader concerns for the university administration.

4.2 ISSUES FOR ADMINISTRATORS AND THE UNIVERSITY ADMINISTRATION

One of the key concerns that faculty have in the use of student evaluation of teaching is about the possible bias of administrators in interpreting data produced by rating instruments and the actions taken (or in some cases, not taken) based on interpretation of the data. The focus in the research literature on administrator bias primarily concerns the use of student evaluation of teaching data in tenure and promotion decisions. The particulars of these concerns are discussed below, but in summary, the key administrative issues that faculty are concerned about involve the capacity (and sometimes the willingness) of deans, departmental chairs and peers on tenure and promotion committees to provide informed and fair assessment of teaching that is based on student evaluation of teaching. While deans, departmental chairs and peer committee members are knowledgeable in their own field(s) of research, they do not always have depth of knowledge in the area of student evaluation of teaching, or more broadly in the area of pedagogy and student learning in higher education. A second concern involves the capacity and commitment of the university administration to use the data produced by student evaluation to provide training and development that supports faculty teaching. While university mission statements commonly identify teaching and learning as priorities, not all universities commit resources and expertise to this part of their mission, and instead privilege research. As a consequence, the data derived from student evaluation of teaching is not effectively used and there is either little incentive or opportunity for faculty development in their teaching.

If student evaluation of teaching and courses is to prove valuable, as Gapp and Fisher note, “first and most critical...[is] the establishment of a safe and trusting environment for both staff and students,”¹⁰⁴ in which “relevant and useful input from participants...leaves all involved feeling empowered within the process.”¹⁰⁵ It is therefore important that there are multiple opportunities for faculty, administrators, and students to be involved in updating the process and instrument for student evaluation of teaching and courses. Also necessary, however, is a shared understanding of the importance of teaching and learning as fundamentally valued by the university, not only in formal documents such as the university mission statement, but also in how faculty and students are supported and engaged in teaching and learning at the level of the university’s courses, programs, and departments.¹⁰⁶ This can be challenging for the university administration, however, as:

Faculty organisational commitment in higher education has declined over time (Cintrón 1999; Judy and D’Amico 1997) and that diminished commitment influences retention and tenure (Werbel and Gould 1984), motivation and involvement (Mowday, Porter, and Steers 1982; Smith, Organ, and Near 1983), task performance and compliance with organisational policy (Angle and Perry 1981) and adoption of organisational values and priorities (O’Reilly and Chatman 1986). For these reasons, it is important to examine the documents that communicate evaluative processes, task expectations and performance standards, in light of the messages they may contain and how these can influence employees’ thinking about their work (Braskamp and Ory 1994; Fairweather 1999; Schön 1983).¹⁰⁷

¹⁰⁴ See Gapp and Fisher 2006, p. 161.

¹⁰⁵ See Gapp and Fisher 2006, p. 158. The authors also note, however, that “No system or method is perfect and innovation and improvement are part of a normal creative improvement process” (2006, p. 161).

¹⁰⁶ As Gapp and Fisher (2006) note, the approach of the university administration can sometimes be limited, unfortunately, to an emphasis on student retention rather than engagement with teaching and learning. That said, there has for many years in the field of higher education been discussion of a “discrepancy between the attention that the teacher [and the university more broadly] pays to the students professional training and the importance that the student should acquire professional skills” and this has implications not only for student satisfaction, but also for changing views on the purpose(s) of higher education (Ghedini and Aquario 2008, p. 593).

¹⁰⁷ See Hardré and Cox 2009, p. 387.

Although the above concerns may be valid, the same authors also note the need for support from the university administration for design and delivery of courses.¹⁰⁸ Further, support is needed from the university administration to recognize and address what Halonen and Ellenberg refer to as “administrative follies” in student evaluation of teaching:

- *Neglect*: failing to ensure validity, reliability, and training for proper use;
- *Constraint*: placing too much weight on student ratings;
- *Conflation*: using a single item as total assessment of performance;
- *Obfuscation*: using unclear questions in the form used for student evaluation of teaching;
- *Jeopardizing*: failing to take adequate measures to protect confidentiality;
- *Delay*: returning results of student evaluation of teaching to faculty at a date too late for the next offering of the course;
- *Quantifying*: accepting figures from student evaluation of teaching as conclusive;
- *Assumptions*: not questioning perspectives on or reasons for ratings;
- *Succumbing*: giving in to somewhat simplistic notions of student satisfaction;
- *Shirking*: not recognizing that other activities (such as peer observation and teaching portfolios) are needed to accompany the use of student evaluation of teaching;
- *Normalizing*: expecting most faculty to be average and that few will be excellent; and
- *Stopping short*: failing to help faculty make sense of the data produced by the student evaluation of teaching instrument.¹⁰⁹

The above list is not meant to reflect the *modus operandi* of university administrators (whether deans, department chairs, or peer reviewers in their administrative activities) or university administration more broadly. It is meant, however, to caution the university community about common administrative flaws in the use of student evaluation of teaching. Effective and responsible use of student evaluation of teaching, it is clear, requires that staff undertaking administrative roles in the process are familiar with and sensitive to a number of issues.¹¹⁰ Further, as Pallett notes, regrettably, “so much emphasis on the summative component of student ratings—getting the right number—[can have the consequence] that what can be learned to improve teaching is often overlooked.”¹¹¹

Support for faculty career development in teaching activities is one of the chief concerns taken up in the literature in regard to administrative issues. Beyond the responsible and informed collection and interpretation of data produced from student evaluation of teaching (complemented by information from other sources on evaluation of teaching), “for real gains in teaching skill to occur, support and mentoring needs to be provided” to faculty.¹¹² As Khandelwal observes, for example, in the “research on student evaluations of teaching effectiveness...that most studies identify characteristics instead of behaviours reduces their utility. Merely knowing that a particular characteristic of effective teaching exists does not

¹⁰⁸ Also, as Marsh (2007) suggests, “ratings for a given instructor should be averaged across different courses to enhance generalizability. If it is likely that an instructor will teach many different classes during his or her subsequent career, then tenure decisions should be based upon as many courses as possible...[and] that a longitudinal archive of SETs is maintained for personnel decisions” (p. 335).

¹⁰⁹ See Halonen and Ellenberg 2006, p. 156-158. In addition, Pallett identifies the following “abuses or misuses of student evaluation” to be aware of and avoid: Abuse 1: Overreliance on student ratings in the evaluation of teaching...Abuse 2: Making too much of too little...While there is substantial evidence that student ratings are reliable, there is always some “noise” in survey data...Abuse 3: Not enough information to make an accurate judgment...Abuse 4: Questionable administrative procedures...If student ratings are taken seriously by faculty and administrators, it is likely that students will take them seriously as well...Abuse 5: using the instrument (or the data collected) inappropriately...Abuse 6: Insufficient attention to selecting/developing an instrument...Abuse 7: Failure to conduct research to support the validity and reliability of a student ratings tool” (2006, p. 56-60).

¹¹⁰ As Harris and Cullen note, for example, administrators “need to be sensitive to issues of power and control, about attitudes toward learning (both student and teacher), and to the many forms of assessment and evaluation and their uses in various contexts. They will need to know the language for writing learning objectives and learning outcomes. They will need to be familiar with techniques for active lecturing, problem-based learning, concept mapping, effective discussion techniques, and more” (2008, p. 61).

¹¹¹ See Pallett 2006, P. 51.

¹¹² Pallett 2006, p. 52.

tell a teacher how to enact that characteristic.”¹¹³ Universities provide different kinds of support for faculty career development in their teaching and the data derived from student evaluation of teaching can prove useful for individual faculty members in seeking support in a particular area, as for the university administration in identifying where support is most needed for faculty.¹¹⁴

Although a review of the literature on career development in faculty teaching activities is beyond the scope of this report, it is perhaps worth noting what Åkerlind identifies as “five qualitatively different approaches to developing as a teacher:

- (1) Building up one’s content knowledge (improving what to teach).
- (2) Building up practical experience (improving how to teach).
- (3) Building up a repertoire of teaching strategies (becoming more skilful as a teacher).
- (4) Finding out what strategies work for the teacher (becoming more effective as a teacher).
- (5) Increasing one’s understanding of what works for the students (becoming more effective in facilitating student learning).¹¹⁵

In the above list, as in the brief discussion of the shift to learner-centred curriculum earlier in this report, it becomes clear that in thinking about and updating the instrument for student evaluation of teaching, there are also student issues that need to be understood. As the research literature makes evident, it is important to involve students in the discussions and the process of updating the instrument used for student evaluation of teaching. This report therefore includes, in the section to follow, an overview of findings from the research literature that can support students in their involvement in the project to update SFU’s student evaluation of teaching instrument.

4.3 ISSUES FOR STUDENTS

If, as the research literature suggests, “student ratings are the most, if not the only, influential measure of teaching effectiveness, active participation by and meaningful input from students can be critical in the success of such teaching evaluation systems.”¹¹⁶ Gaining meaningful input from students, however, involves an understanding of student perspectives on evaluation of teaching. One of the most significant findings from the research on student perceptions of teaching evaluation is that “improvement in teaching” is the most valuable outcome of their participation in such evaluations.¹¹⁷ Although there is a substantial body of research on student views of teaching effectiveness in higher education, the title of

¹¹³ See Khandelwal 2009, p. 300. Khandelwal elaborates thus: “For example, knowing that rapport is a common characteristic of excellent teaching does not provide any detail of the behaviors that convey or communicate rapport. There is more potential for improvement of teaching if behavioral items such as ‘Uses humor’ or ‘Gives real life examples’ are used, instead of abstract, general statements like ‘good presentation’” (2009, p. 300-301).

¹¹⁴ For the support provided to SFU faculty, see the SFU Teaching and Learning Centre: <http://tlcentre.sfu.ca/>. With the support of the Office of the Vice-President, Academic, SFU also offers teaching and learning grants to “recognize teaching development as a scholarly activity and to stimulate the development, implementation, and investigation of innovative teaching and learning” (<http://www.sfu.ca/teachlearn/tlgrants.html>).

¹¹⁵ See Edström 2007, p. 103, citing Åkerlind 2007.

¹¹⁶ See Chen and Hoshower 2003, p. 71. The authors also note that, “Several studies...have observed a significant link between student attitudes toward the evaluation of teaching effectiveness and the success of a teaching evaluation system (Hofman & Kremer, 1980; Marsh, 1984, 1987; Douglas & Carroll, 1987; Tom et al., 1990)” (2003, p. 72). Campbell and Bozeman (2008) also note from their study that students strongly believe that they should complete teaching evaluations.

¹¹⁷ See, for example, Chen and Hoshower (2003), Al-Issa and Sulieman (2007), and Algozzine *et al* (2010). Algozzine *et al* (2010) also note that the “least important use of teaching evaluations, from the students’ standpoint, is for the professor’s tenure, promotion, and salary decisions” (p. 42). This is in keeping with findings from Boysen (2008) that “revenge appears to be one of the least important factors” for students in evaluating teaching (p. 218). Boysen notes further that, “students who believed that evaluations have little consequence were also more likely to provide low ratings on them” (2008, p. 220).

an article by Remedios and Lieberman – “I liked your course because you taught me well” – captures an important finding in research on student views of teaching and teaching evaluation.¹¹⁸ Following from this, one of the key areas of discussion for students as they are involved in updating the instrument for student evaluation of teaching is about their views and understandings of effective teaching. Here, there are important links to the shift from a teacher-centred to a learner-centred classroom and curriculum, discussed earlier in this report.

The shift to a learner-centred classroom and curriculum presents challenges to students, as it does for teachers, given what often can be the limited understanding that students have of themselves as learners in this new context,¹¹⁹ as well as resistance that some students can have to what they have come to expect as the status quo of a teacher-centred classroom in which they are more passive than active participants.¹²⁰ It is because of such challenges, shared by students and teachers (as by the university administration more broadly), that communication between faculty and students (and between administrators and faculty) about a diversity of teaching methods and defined learning outcomes is particularly important, as understandings about teaching and learning in higher education have changed with both research and the realities in classrooms across the disciplines. Arriving at shared understandings about the responsibilities of both faculty and students in teaching and learning is also important for efforts by the university to support innovation and diversity in teaching methods.¹²¹

Although a shared understanding between faculty and students about their respective roles (and responsibilities) in teaching and learning is invaluable, knowledge of the reasons students’ give for their high and low evaluations of teaching is useful for identifying student views on effective teaching. Most published studies on student evaluation of teaching include a rating scale of some kind and data derived from the scale(s). Two studies in particular are cited below, however, because they are not only consistent with general findings on student evaluation of teaching, but also, they identify reasons that students give for low ratings and positive descriptors used in student written comments on teaching effectiveness.

In his study on why students’ give poor evaluations of teaching, the following reasons (in descending order of frequency) were given:

- Poor teaching style or methods
- Rude, disrespectful, or uncaring
- Unprepared or unorganized
- Unfair in grading
- Unclear or poor communication skills
- Generally a bad professor
- Professor unknowledgeable
- Workload too heavy
- Did not learn enough
- Professor was unavailable

¹¹⁸ See Remedios and Lieberman 2009, p. 91. The full title of the article is, “I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students’ evaluation of teaching” and reports on the authors’ study in the UK. From their study, they conclude that, “Although grades and course difficulty did have a small influence on end-of-semester course ratings, structural modelling revealed that ratings were largely determined by how much students enjoyed or felt stimulated by the course content, which in turn depended on the perceived quality of teaching” (p. 91).

¹¹⁹ See, for example, Meyer (2010).

¹²⁰ See, for example, Felder and Brent (1996) cited in Weimer (2002), and Edström (2008).

¹²¹ In keeping with the discussion by Weimer (and others, as cited above) about student resistance to change in higher education classrooms and curriculum, Struyven *et al*, for example, in their study on “students’ likes and dislikes regarding student-activating and lecture-based educational settings,” found that, “While lecture-taught students’ evaluations of the experienced teaching were generally focused and positive, students’ perceptions of the activating methods varied widely and both extremely positive and negative opinions were present” (2008, p. 295).

- Hard grading or no one in class did well
- Did not teach test material
- Objectives unclear
- Class was pointless
- Professor was not helpful
- Class did not match syllabus
- Professor had a poor attitude ¹²²

The above list is significant for students and faculty (as well as for administrators) in a number of ways. For students and faculty, the list is useful for identifying where students and faculty share views on valid reasons for giving a poor evaluation of teaching and courses – such as the professor being unavailable or disrespectful, perhaps, as examples – while also providing insight on where discussion between faculty and students is needed to arrive at shared understandings of expectations – for example, as with workload and grading, or even explicit (rather than tacit) views on teaching methods and purposes of the course. The above list is also useful in identifying where students base evaluations on instructor characteristics or other aspects of the course (such as the amount a student felt he or she learned) rather than on instructor behaviours/actions and on what happens in a course, which is useful for taking steps in moving students, faculty, and administrators toward the use of evaluation of teaching that is both effective and responsible.

In contrast to the above reasons for giving poor teaching evaluations, the following were the most frequently used descriptors in written comments by students giving favourable teaching evaluations (presented here in descending order of frequency):

- Interesting
- Approachable
- Clarity
- Ability to explain
- Effective teaching
- Knowledgeable
- Willing to help
- Aids understanding
- Friendly
- Patient
- Delivery of concepts
- Humorous
- Stimulates thinking
- Effective use of examples
- Encouraging
- Effective questioning
- Engaging
- Good lecture notes
- Concise
- Real-life applications ¹²³

As with the previous list of reasons for poor teaching evaluations, the above list contains items that refer to instructor attributes as well as activities, which can be useful for students, faculty, and administrators, as with the previous list.

¹²² See Boysen 2008, p. 220.

¹²³ See Pan *et al*/2008, p. 83.

Pan *et al* also report that their study “indicate[s] that students value teaching quality more than teacher characteristics.”¹²⁴ That said, as is the case for faculty and administrators, there are also “student follies”¹²⁵ that can be detrimental to student evaluation of teaching. It is therefore important for students to understand that there is learning involved in becoming “more sophisticated evaluators”¹²⁶ and contributing valuable and responsible input for evaluation of teaching and courses. For students, two other points warrant brief mention. The first is the usefulness of student self-evaluation in improving student achievement.¹²⁷ Second, as Shuell notes, “It is helpful to remember that what the student does is actually more important in determining what is learned than what the teacher does.”¹²⁸ In student evaluation of teaching and courses, in particular in the shift from the teacher-centred to student-centred classroom and curriculum, the roles and responsibilities of the student are as important as those of the teacher for successful teaching and learning.

¹²⁴ See Pan *et al* 2008, p. 73.

¹²⁵ In addition to the “faculty follies” and “administrator follies” identified by Halonen and Ellenberg, the following are “student follies” that damage the usefulness and integrity of student evaluation of teaching: “they fade...they suck up...they go woebegone...they go blue...they go for the jugular...they mob...they calculate their [financial] losses...they lose focus...they use alien and imprecise language...they turn seductive...they delude themselves...{and} they become agents [of protest]” (2006, p. 153-154).

¹²⁶ As cited earlier, see McKeachie 1997, p. 1223 (cited in Burden 2010, p. 107).

¹²⁷ See, for example, Combs *et al* (2008, p. 93): “Meta-analyses of college students’ self-evaluation of learning find that it is positively correlated with student achievement (Cohen, 1986; Falchikov & Boud, 1998). More recent, individual studies from various content areas also find overall high correlations between self-assessment results and ratings based on a variety of external criteria (Mehta & Danielson, 1989; Coombe, 1992; Oscarsson, 1997; Chesebro & McCroskey, 2000; Wortham & Harper, 2002; Fitzgerald *et al.*, 2003).”

¹²⁸ See Edström (2008, p. 102), citing Biggs (2003), citing Shuell (1986).

5 REFERENCE LIST

- Abrami, P. (1989) How should we use student ratings to evaluate teaching? *Research in Higher Education*, 30(2): 221–227.
- Abrami, P., Marilyn, H. & Raiszadeh, F. (2001) Business students' perceptions of faculty evaluations, *The International Journal of Educational Management*, 15(1): 12-22.
- Abrami, P., d'Apollonia, S., & Cohen, P. (1990). The validity of student ratings of instruction: What we know and don't know. *Journal of Educational Psychology*, 82(2): 219-231.
- Ackerman, D., Gross, B. & Vigneron, F. (2009). Peer observation reports and student evaluations of teaching: Who are the experts? *The Alberta Journal of Educational Research*, 55(1): 18-39.
- Adams, J. (1997). Student evaluations: The ratings game. *Inquiry*, 1(2): 10-16.
- Åkerlind, G. (2007). Constraints on academics' potential for developing as a teacher. *Studies in Higher Education*, 32(1): 21-37.
- Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Mohanty, G., & Spooner, F. (2010). Multi-method evaluation of college teaching. *Journal on Excellence in College Teaching*, 21 (1), 27-49.
- Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., Mohanty, G., & Spooner, F. (2004). Student evaluation of college teaching: A practice in search of principles. *College Teaching*, 45(1): 1-7.
- Al-Issa, A. & Sulleman, H. (2007). Student evaluations of teaching: Perceptions and biasing factors. *Quality Assurance in Education*, 15(3): 302-317.
- Arreola, R. (2007). *Developing a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems*. San Francisco, CA: Anker Publishing.
- Basow, S. & Howe, K. (1987). Evaluations of college professors: Effects of professors' sex-type, and sex, and students' sex. *Psychological Reports*, 60(2): 671-678.
- Berk, R. (2006). *Thirteen strategies to measure college teaching*. Sterling, VA: Stylus.
- Betoret, F. (2007). The influence of students' and teachers' thinking styles on student course satisfaction and on their learning process. *Educational Psychology*, 27(2): 219-234.
- Biggs, J. (2003). *Teaching for quality learning at university: What the student does*. Buckingham, UK: SRHE and Open University Press.
- Blackhart, G., Peruche, B., DeWall, C. & Joiner, Jr., T. (2006). Factors influencing teaching evaluations in higher education. *Teaching of Psychology*, 33(1): 37-39.
- Blunt, A. (1991). The effects of anonymity and manipulated grades on student ratings of instructors. *Community College Review* 18(4): 48-53.
- Boysen, G. (2008). Revenge and Student Evaluations of Teaching. *Teaching of Psychology*, 35(3): 218-222.
- Braskamp, L., Brandenburg, D., & Ory, J. (1985). *Evaluating teaching effectiveness: A practical guide*. Beverly Hills, CA: Sage.
- Braskamp, L. A., Caulley, D., & Costin, F. (1979). Student ratings and instructor self-ratings and their relationship to student achievement. *American Educational Research Journal*, 16(3): 295-306.
- Bryson, R. (1974). Teacher evaluations and student learning: A reexamination. *The Journal of Educational Research*, 68(1): 12-14.
- British Columbia Education Quality Assurance Organization. (2012). Retrieved Jan 5, 2012, from: <http://www.bceqa.ca/>
- Buchert, S., Laws, E., Apperson, J., & Bregman, N. (2008). First impressions and professor reputation: Influence on student evaluations of instruction. *Social Psychology of Education*, 11: 397-408.

- Burden, P. (2010). Creating confusion or creative evaluation? The use of student evaluation of teaching surveys in Japanese tertiary education. *Educational Assessment, Evaluation and Accountability*, 22: 97-117.
- Burdsal, C., & Bardo, J. (1986). Measuring students' perceptions of teaching: Dimensions of evaluation. *Educational and Psychological Measurements*, 56: 63-79.
- Buskist, W., Sikorski, J., Buckley T. & Saville, B. (2002). Elements of master teaching. In *The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer* (Stephen Davis & William Buskist, Eds.), Mahwah, NJ: Erlbaum, pp. 27-39.
- Campbell, J. & Bozeman, W. (2008). The value of student ratings: Perceptions of students, teachers, and administrators. *Community College Journal of Research and Practice*, 32(1): 13-24.
- Cashin, W. (1988). Student ratings of teaching: A summary of the research . Idea Paper No. 20. Manhattan, KS: Center for Faculty Evaluation and Development in Higher Education.
- Cashin, W. & Slawson, H. (1977). Description of data base 1976–1977, IDEA Technical Report No. 2. Manhattan, KS: Center for Faculty Evaluation and Development in Higher Education.
- Centra, J., & Gaubatz, N. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, 71(1): 17-33.
- Centra, J. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco, CA: Jossey-Bass.
- Centra, J. (1977). Student ratings of instruction and their relationship to student learning. *American Educational Research Journal*, 14(1): 17-24.
- Chen, Y. and Hoshower, L. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment and Evaluation in Higher Education*, 28:1: 71-88.
- Chesebro, J. & McCroskey, J. (2000). The relationship between students' reports of learning and their actual recall of lecture material: A validity test. *Communication Education*, 49(3): 297-301.
- Clayson, D. (2004). A test of the reciprocity effect in the student evaluation of instructors in marketing classes. *Marketing Education Review*, 14(2): 11-36.
- Clayson, D. & Haley, D. (1990). Student evaluations in marketing: What is actually being measured? *Journal of Marketing Education*, 12(3): 9-17.
- Cohen, P. (1986). An updated and expanded meta-analysis of multisection student rating validity studies, paper presented at the 70th Annual Meeting of the American Educational Research Association, San Francisco, CA.
- 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.
- Coombe, C. (1992). The relationship between self-assessment estimates of functional literacy skills and basic English skills test results in adult refugee ESL learners Columbus: Ohio State University.
- Council of Europe. (2012). What is the Bologna Process? Retrieved January 5, 2012, from: http://www.coe.int/t/dg4/highereducation/ehea2010/bolognapedestrians_en.asp
- Darby, J. (2007). Evaluating course evaluations: The need to establish what is being measured. *Assessment and Evaluation in Higher Education*, 32(4): 441-455.
- Denson, N., Loveday, T., & Dalton, H. (2010). Student evaluation of courses: What predicts satisfaction? *Higher Education Research and Development*, 29(4): 339-356.
- Douglas, P. & Carroll, S. (1987). Faculty evaluations: Are college students influenced by differential purposes? *College Student Journal*, 21(4): 360-365.
- Edström, K. (2008). Doing course evaluation as if learning matters most. *Higher Education Research and Development*, 27(2): 95-106.
- Eiszler, N. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education*, 43(4), 483–501.
- Entwhistle, N. (2010). Taking stock: An overview of key research findings. In *Taking Stock: Research on Teaching and Learning in Higher Education* (J. Christensen Hughes and J. Mighty, Eds.). School of Policy Studies, Queen's University, pp. 15-60.

- Erdle, S., and Murray, H. (1986). Interfaculty differences in classroom teaching behaviours and their relationship to student instructional ratings. *Research in Higher Education*, 24(2): 115-127.
- Ewing, V., Stukas, A. & Sheehan, E. (2003). Student prejudice against gay male and lesbian lecturers. *Journal of Social Psychology*, 143(5): 569-579.
- Fairweather, J. & Rhoads, R. (1995). Teaching and the faculty role: Enhancing the commitment to instruction in American colleges and universities. *Educational Evaluation and Policy Analysis*, 17(2): 179-94.
- Falchikov, N. & Boud, D. (1998). Student self-assessment in higher education: a meta-analysis. *Review of Educational Research*, 59(4): 395-430.
- Faranda, W. & Clarke III, I. (2004). Student observations of outstanding teaching: Implications for marketing educators. *Journal of Marketing Education*, 26(3): 271-281.
- Felder, R. & Brent, R. (1996). Navigating the bumpy road to student-centered instruction. Retrieved December 20, 2011, from <http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/Resist.html>
- Feldman, K. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, 31(4): 583-645.
- Fernandez, J., Mateo, M., & Muniz, J.(1998). Is there a relationship between class size and student ratings of teacher quality? *Educational and Psychological Measurement*, 58(4): 596-604.
- Fitzgerald, L., Ferlie, E. & Hawkins, C. (2003). Innovation in healthcare: how does credible evidence influence professionals? *Health and Social Care in the Community*, 11(3): 219- 228.
- Frick, T. Chadha, R., Watson, C. & Zlatkovska, E. (2010). Improving course evaluations to improve instruction and complex learning in higher education. *Educational Technology Research Development*, 58(2):115–136.
- Fry, H. & Ketteridge, S. (1999). Teaching portfolios In *A handbook for teaching and learning in higher education: enhancing academic practice* (Heather Fry, Steven Ketteridge, Stephanie Marshall, Eds.), London, UK: Kogan Page, pp. 235-246.
- Fullerton, H. (2002). Observation of teaching. In *Evaluating faculty performance* (Carol Colbeck, Ed), San Francisco, CA: Jossey-Bass, pp. 220-234.
- Gapp, R., & Fisher, R. (2006). Achieving excellence through innovative approaches to student involvement in course evaluation within the tertiary education sector. *Quality Assurance in Education*, 14(2): 156-166.
- Gappa, J., Austin, A. & Trice, A. (2007). *Rethinking faculty work: Higher education's strategic imperative*. San Francisco, CA: Jossey-Bass.
- Germain, M., & Scandura, T. (2005). Grade inflation and student individual differences as systematic bias in faculty evaluations. *Journal of Instructional Psychology*, 32(1): 58-67.
- Ghedini, E. & Aquario, D. (2008). Moving towards multidimensional evaluation of teaching in higher education: A study across four faculties. *Higher Education*, 56(5): 583–597.
- Goldstein, G. & Benassi, V. (2006). Students' and instructors' beliefs about excellent lecturers and discussion leaders. *Research in Higher Education*, 47(6): 685-707.
- Greenwald, A. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52(11): 1182-1186.
- Greenwald, A. & Gillmore, G. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, 89(4): 741-743.
- Griffin, B. (2001). Instructor reputation and student ratings of instruction. *Contemporary Educational Psychology*, 26: 534–552.
- Gurung, R., & Vespia, K. (2007). Looking good, teaching well? Linking liking, looks, and learning. *Teaching of Psychology*, 34: 5-10.
- Hallinger, P. (2010). Using faculty evaluation to improve teaching quality: A longitudinal case study of higher education in Southeast Asia. *Educational Assessment and Evaluation Accountability*, 22(4): 253-274.

- Halonen, J. & Ellenberg, G. (2006). Teaching evaluation follies: Misperceptions and misbehavior in student evaluations of teachers. In *Evaluating faculty performance: A practical guide to assessing teaching, research, and service* (Peter Seldin and Associates, Eds.). San Francisco, CA: Anker Publishing, pp. 150-164.
- Hardré, P. & Cox, M. (2009). Evaluating faculty work: expectations and standards of faculty performance in research universities. *Research Papers in Education*, 24(4): 383-419.
- Harris, M. & Cullen, R. (2008). Observing the learner-centered class. *Florida Journal of Educational Administration and Policy*, 1(2): 57-66.
- Heckert, T., Latier, A., Ringwald-Burton, A. & Drazen, C. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to "buy" better evaluations through lenient grading? *College Student Journal*, 40(3): 588-596.
- Higgerson, M.L. (1999). Building a climate conducive to effective teaching evaluation. In P. Seldin (Ed.), *Changing practices in evaluating teaching* (pp. 194-212). Bolton, MA: Anker.
- Hobson, S. (1997). The impact of sex and gender-role orientation on student evaluations of teaching effectiveness in counselor education and counseling psychology. Ed.D. diss. Western Michigan University, Kalamazoo.
- Hobson, S., & Talbot, D. (2001) Understanding student evaluations. *College Teaching*, 49(1): 26-31.
- Hofman, F. & Kremer, L. (1980). Attitudes toward higher education and course evaluation. *Journal of Educational Psychology*, 72(5): 610-617.
- Holloway, S. (1988). Concepts of ability and effort in Japan and the United States. *Review of Educational Research*, 58(3): 327-345.
- Howard, G., & Maxwell, S. (1980). Correlation between student satisfaction and grades: A case of mistaken causation? *Journal of Educational Psychology*, 72: 107-118.
- Husbands, C., & Fosh, P. (1993). Students' evaluation of teaching in higher education: Experiences from four European countries and some implications of the practice. *Assessment and Evaluation in Higher Education*, 18(2): 95-114.
- Jackson, D., Teal, C., Raines, S., Nansel, T., Force, R., & Burdsal, C. (1999). The dimensions of students' perceptions of teaching effectiveness. *Education and Psychological Measurement*, 59(4): 589-590.
- Jennings, T. (2010). Teaching 'out' in the university: an investigation into the effects of lesbian, bisexual, and transgender faculty self-disclosure upon student evaluations of faculty teaching effectiveness in the USA. *International Journal of Inclusive Education*, 14(4): 325-339.
- Kappa, C. (1995). The origins of teaching portfolios, *Journal of Excellence in College Teaching*, 6 (1): 5-56.
- Kelley, H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality*, 18(4): 431-439.
- Kember, D., & Wong, A. (2000). Implications for evaluation from a study of students' perceptions of good and poor teaching. *Higher Education*, 40(1), 69-97.
- Khandelwal, K. (2009). Effective teaching behaviours in the college classroom: A critical incident technique from students' perspective. *International Journal of Teaching and Learning in Higher Education*, 21(3): 299-309.
- Kinzie, J. (2010). Student engagement and learning: Experiences that matter. In *Taking Stock: Research on Teaching and Learning in Higher Education* (J. Christensen Hughes and J. Mighty, Eds.). School of Policy Studies, Queen's University, pp. 139-154.
- Knapper, C. (2010). Changing teaching practice: Barriers and strategies. In *Taking Stock: Research on Teaching and Learning in Higher Education* (J. Christensen Hughes and J. Mighty, Eds.). School of Policy Studies, Queen's University, pp. 229-242.

- Kogan, L., Schoenfeld-Tacher, R., & Hellyer, P. (2010). Student evaluations of teaching: perceptions of faculty based on gender, position, and rank. *Teaching in Higher Education*, 15(6): 623-636.
- Lecouter, A., & DelFabbro, P. H. (2001). Repertoires of teaching and learning: A comparison of university teachers and students using Q methodology. *Higher Education*, 42(2), 205-235.
- Liddle, B. (1997). Coming out in class: Disclosure of sexual orientation and teaching evaluations. *Teaching of Psychology*, 24(1): 32-35.
- MacBeath, J. & McGlynn, A. (2002). Self evaluation. What's in it for schools? London, UK: Routledge Farmer.
- McKeachie, W. (1997). Student ratings: the validity of use. *American Psychologist*, 52(11): 1218-1225.
- Marks, R. (2000). Determinants of student evaluations of global measures of instructor and course value. *Journal of Marketing Education*, 22(2): 108-119.
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning II: Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology*, 46: 115-127.
- Marsh, H. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R.P. Perry and J.C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, 319-383. Dordrecht, The Netherlands: Springer.
- Marsh, H. (1987). Students' evaluation of university teaching: Research findings, methodological issues and directions for future research. *International Journal of Educational Research* 11(3): 253-388.
- Marsh, H. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, 52(1): 11-95.
- Marsh, H., & Dunkin, M. (1992). Students' evaluations of university teaching: A multidimensional perspective. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 8, pp. 143-233). New York: Agathon Press.
- Marsh, H., & Roche, L. (1997). Making students' evaluation teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11): 1187-1197.
- Mehta, S. and S. Danielson. (1989). Self-assessment by students: an effective, valid, and simple tool?, Paper presented at the ASEE National Conference, Charlotte, NC.
- Merrill, M. (2002). First principles of instruction. *Educational Technology Research and Development*, 50(3), 43-59.
- Meyer, J. (2010). Helping our students: Learning, metalearning, and threshold concepts. In *Taking Stock: Research on Teaching and Learning in Higher Education* (J. Christensen Hughes and J. Mighty, Eds.). School of Policy Studies, Queen's University, pp. 191-214.
- Morris, T., Gorham, J., Stanley, H., & Huffman, D. (1996). Fashion in the classroom: Effects of attire on student perceptions of instructors in college classes. *Communication Education*, 45: 135-147.
- Murray, H., Rushton, J., & Paunnen, S. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82(2): 250-261.
- Naftulin, D., Ware, J., & Donnelly, E. (1983). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 75: 138-49.
- Oliver, B., Tucker, B., Gupta, R. & Yeo, S. (2008). eVALUate: an evaluation instrument for measuring students' perceptions of their engagement and learning outcomes. *Assessment and Evaluation in Higher Education*, 33(6): 619-630

- Olshavsky, R., & Spreng, R. (1995). Consumer satisfaction and students: Some pitfalls of being customer driven. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 8(2): 69-77.
- Oscarsson, M. (1997). Self-assessment of foreign and second language proficiency. *Encyclopedia of Language and Education*, 7: 175-87.
- Pallett, W. (2006). Uses and abuses of student ratings. In *Evaluating faculty performance: A practical guide to assessing teaching, research, and service* (Peter Seldin and Associates, Eds.). San Francisco, CA: Anker Publishing, pp. 50-65.
- Pan, D., Tan, G., Ragupathi, K., Booluck, K., Roop, R., & Yuen, K. (2007). Profiling teacher/teaching using descriptors derived from qualitative feedback: Formative and summative applications. *Research in Higher Education*, 50: 73-100.
- Paulsen, M. (2002). Evaluating teaching performance. In *Evaluating faculty performance* (Carol L. Colbeck, Ed.), pp. 5-18. Jossey-Bass, San Francisco, CA.
- Pounder, J. (2007). Is student evaluation of teaching worthwhile? An analytical framework for answering the question. *Quality Assurance in Education*, 15(2): 178-191.
- Platt, M. (1993). What student evaluations teach. *Perspectives in Political Science*, 22(1): 29-40.
- Pratt, D., & Collins, J. (2001). Teaching perspectives inventory. Retrieved January 2, 2012, from <http://teachingperspectives.com/>
- Prosser, M. (2010). Faculty research and teaching approaches: Exploring the relationship. In *Taking Stock: Research on Teaching and Learning in Higher Education* (J. Christensen Hughes and J. Mighty, Eds.). School of Policy Studies, Queen's University, pp. 129-138.
- Quality Assurance Agency for Higher Education. (2012). Home page. Retrived Jan 5, 2012, from: <http://www.qaa.ac.uk/>
- Ramsden, P. (2003). *Learning to Teach in Higher Education*. London, UK: Routledge.
- Remedios, R. & Lieberman, D. (2008). I liked your course because you taught me well: the influence of grades, workload, expectations and goals on students' evaluations of teaching. *British Educational Research Journal*, 34(1): 91-115.
- Remmers, H. (1927). The Purdue Rating Scale for Instructors. *Educational Administration and Supervision*, 6: 399-406.
- Rifkin, T. (1995). The status and scope of faculty evaluation. Retrieved January 5, 2012, from <http://www.ericdigests.org/1996-1/status.htm>
- Russ, T., Simonds, C. & Hunt, S. (2002). Coming out in the classroom...An occupational hazard?: The influence of sexual orientation on teacher credibility and perceived student learning. *Communication Education*, 51(3): 311-24.
- Saroyan, A. (2010). Research on student learning: Converging and diverging messages. In *Taking Stock: Research on Teaching and Learning in Higher Education* (J. Christensen Hughes and J. Mighty, Eds.). School of Policy Studies, Queen's University, pp. 95-112.
- Schaeffer, G., Epting, K., Zinn, T. & Buskist, W. (2003). Student and faculty perceptions of effective teaching: A successful replication. *Teaching of Psychology*, 30(2): 133-136.
- Scherr, F., & Scherr, S. (1990). Bias in student evaluations of teacher effectiveness. *Journal of Education for Business*, 65(8): 356-358.
- Seldin, P. (2006). Building a successful evaluation program. In *Evaluating faculty performance: A practical guide to assessing teaching, research, and service* (Peter Seldin and Associates, Eds.). San Francisco, CA: Anker Publishing, pp. 1-20.
- Seldin, P. (1985). Current Practices in Evaluating Business School Faculty. Pleasantville, NY: Center for Applied Research, Lubin School of Business Administration, Pace University.
- Sidanius, J. & Crane, M. (1989). Job evaluation and gender: The case of university faculty. *Journal of Applied Social Psychology*, 19(2): 174-97.
- Simon Fraser University. (2012). Office of the Vice-President, Academic. Retrieved January 5, 2012, from: <http://www.sfu.ca/teachlearn/tlgrants.html>
- Simon Fraser University. (2012). SFU Teaching and Learning Centre. Retrieved January 5, 2012, from: <http://tlcentre.sfu.ca/>

- Simpson, P. & Siguaw, J. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, 22(3): 199-213.
- Smith, M., & Glass, G. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Education Research Journal*, 17(4): 419-433.
- Soyjka, J., Gupta, A., & Deeter-Schmelz, D. (2002). Student and faculty perceptions of student evaluations of teaching: A study of similarities and differences. *College Teaching*, 50(2): 44-49.
- Sproule, R. (2000). Student evaluation of teaching: A methodological critique of evaluation practices. *Education Policy Analysis Archives*, 8(50).
- Struyven, K., Dochy, F. & Janssens, S. (2008). Students' likes and dislikes regarding student-activating and lecture-based educational settings: Consequences for students' perceptions of the learning environment, student learning and performance. *European Journal of Psychology of Education*, 23(3): 295-317.
- Teaching Perspectives Inventory (2012). Home page. Retrieved January 5, 2012, from: <http://teachingperspectives.com/>
- Tertiary Education and Quality Standards Agency. (2012). Retrieved Jan 5, 2012, from: <http://www.teqsa.gov.au/>.
- The Higher Education Academy. (2012). Home page. Retrieved January 5, 2012, from: <http://www.heacademy.ac.uk/ukpsf>
- Theall, M., Abrami, P., & Mets, L. (Eds.). (2001). *The student ratings debate: Are they valid? How can we best use them?* San Francisco: Jossey-Bass.
- Theall, M., & Franklin, J. (Eds.). (1990). *Student ratings of instruction: Issues for improving practice*. San Francisco: Jossey-Bass.
- Trigwell, K. (2010). Teaching and learning: A relational view. In *Taking Stock: Research on Teaching and Learning in Higher Education* (J. Christensen Hughes and J. Mighty, Eds.), School of Policy Studies, Queen's University, pp. 115-128.
- Tom, G., Swanson, S. & Abbott, C. (1990). The effect of student perception of instructor evaluations on faculty evaluation scores. *College Student Journal*, 24(3): 268-273.
- Weimer, M. (2010). Taking stock of what faculty know about student learning. In *Taking Stock: Research on Teaching and Learning in Higher Education* (J. Christensen Hughes and J. Mighty, Eds.). School of Policy Studies, Queen's University, pp. 81-94.
- Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. San Francisco, CA: Jossey-Bass.
- Wagenaar, T. (1995). Student evaluation of teaching: Some cautions and suggestions. *Teaching Sociology*, 23(1): 64-68.
- Widmeyer, W. & Loy, J. (1988). When you're hot, you're hot! Warm-cold effects in first impressions of persons and teaching effectiveness. *Journal of Educational Psychology*, 80(1): 118-121.
- Zabaleta, F. (2007). The use and misuse of student evaluation of teaching. *Teaching in Higher Education*, 12(1): 55-76.



SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

TEACHING AND COURSE EVALUATION PROJECT FINAL REPORT

APPENDIX III: REPORT ON KEY RESEARCH FINDINGS - INSTRUMENTATION

This research report was prepared by Hui Niu as part of the SFU Teaching and Course Evaluation Project. The report has been reformatted for inclusion in the project's final report.

MAY 29, 2012

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	1
1.1	CONTEXT	1
1.2	KEY ISSUES REGARDING INSTRUMENTATION	1
1.2.1	MULTIDIMENSIONALITY OF TEACHING EFFECTIVENESS	1
1.2.2	PURPOSES OF EVALUATING TEACHING EFFECTIVENESS	1
1.2.3	VALIDITY AND RELIABILITY	1
1.2.4	POTENTIAL BIASES: MYTHS OR REALITY?	2
2	INTRODUCTION	3
2.1	THE PROJECT	3
2.2	THE FIELD	3
3	MEASURING TEACHING EFFECTIVENESS	5
3.1	WHAT TO MEASURE?	5
3.2	WHY MEASURE?	5
4	HOW TO MEASURE: INSTRUMENT DEVELOPMENT	7
4.1	ITEM CONSTRUCTION	7
4.2	QUESTIONNAIRE CONTENT AND FORMAT	7
4.3	PSYCHOMETRIC TESTING	8
4.3.1	VALIDITY	9
4.3.2	RELIABILITY	10
4.4	INSTRUMENT IMPLEMENTATION	11
4.5	USE AND INTERPRETATION OF STUDENT EVALUATION DATA	11
5	POTENTIAL BIASES: MYTHS OR REALITY?	13
5.1	PRIOR SUBJECT INTEREST	13
5.2	EXPECTED GRADE/GIVEN GRADES	13
5.3	WORKLOAD/DIFFICULTY	14
5.4	CLASS SIZE	14
6	REFERENCES	15

1 EXECUTIVE SUMMARY

1.1 CONTEXT

Student evaluations of teaching and courses have been widely used internationally for various purposes such as tenure and promotion and teaching improvement. The Teaching and Course Evaluation Project (the Project) at Simon Fraser University (SFU) is an initiative sponsored by the VPA on the recommendation of the Senate Committee on University Teaching and Learning (SCUTL) and the Task Force on Teaching and Learning (TFTL) to develop new instruments and guidelines for student evaluation of teaching and courses. This report serves as one of the supporting documents for the Project¹ to provide the SFU community with an overview of research findings and current practices on student evaluation of teaching and courses with a focus on instrument development and related methodological issues. The Project strives to develop a valid and reliable instrument that best suits SFU's evaluation needs by closely examining the relevant research literature and through community engagement during the process of instrument development and implementation.

1.2 KEY ISSUES REGARDING INSTRUMENTATION

1.2.1 MULTIDIMENSIONALITY OF TEACHING EFFECTIVENESS

Student evaluations of teaching and courses intend to measure a hypothetical construct of "teaching effectiveness." There is a vast amount of research on student evaluations of teaching, however no consensus has been reached on what constitutes "good teaching." It is generally understood that teaching effectiveness is a multidimensional and complex construct that includes multiple and possibly networked sub-constructs. What constitutes effective teaching may also vary from institution to institution, discipline to discipline, and faculty to faculty. Therefore, special attention will be given to gathering information on local preferences, priorities and standards pertinent to effective teaching at SFU. Such standards and priorities will be developed based on the literature and SFU's unique identity and evaluation needs at institutional, departmental, or unit levels. This rationale has prompted the planning of high level community engagement as an essential part of the Project in seeking advice, consultation, and feedback from all stakeholders before and during the process of developing a sound instrument that can best meet SFU's evaluation needs.

1.2.2 PURPOSES OF EVALUATING TEACHING EFFECTIVENESS

The purpose of evaluating teaching effectiveness is twofold: collecting information for summative and formative reasons. Formative evaluation in the context of teaching quality is intended for instructors to improve their teaching. Summative evaluation is commonly used for personnel decisions on tenure, promotion, or teaching appointments. A larger context for the use of teaching assessment is as a response to the growing interest of the public to hold postsecondary institutions accountable for student learning.²

1.2.3 VALIDITY AND RELIABILITY

Numerous research studies, reviews and meta-analyses have shown that student evaluation of teaching can be valid, reliable and stable. However, faculty opinions of student evaluation of teaching vary vastly from viewing it as reliable, valid and useful to unreliable, invalid and useless.

Validity of student evaluation of teaching is primarily dependent upon the interpretation and use of the results, including inferences about teaching effectiveness. Validity has been the most common concern

¹ For a broader review on general issues regarding student evaluations of teaching and courses, see the report submitted to the Senate Committee on University Teaching and Learning by Johanne Provençal in January 2012.

² See the report submitted to the Senate Committee on University Teaching and Learning by Johanne Provençal in January 2012 for more details about this topic.

about student evaluations of teaching in higher education. Such concerns are connected to assumptions that students are not qualified to evaluate teaching and the existence of potential bias variables. These variables have been researched extensively in the literature, and overall, researchers argue that student ratings are consistent with other data sources such as instructor self-evaluation, trained observer evaluation and former student ratings. Although more research is needed to establish validity of student evaluations, such as using more criteria for teaching effectiveness rather than mainly student achievement, researchers are optimistic that student ratings of teaching effectiveness are at least one measure that reflects teaching quality.

Reliability is an important aspect of validity. Research has shown that a student evaluation instrument can be reliable, especially when the instrument is well designed and tested according to psychometric principles. Psychometrics is the field of study concerned with the theory and technique of psychological measurement. The field is primarily concerned with the construction and validation of measurement instruments and includes the development of theoretical approaches to measurement. Reliability of a student evaluation tool is usually determined by inter-rater agreement (agreement across different students in the same class) and internal consistency of a scale (correlations among items measuring the same element of effective teaching)³. Stability refers to reliability over time. Longitudinal studies comparing immediate ratings with ratings at least a year after have shown high correlations and thus demonstrate good stability.

Generalizability refers to reliability across situations. It assumes that evaluation of an instructor's general teaching effectiveness across courses is more reliable than one snapshot of his/her performance in a particular course at a particular time. Research has indicated that student ratings of teaching effectiveness across different courses or the same course across multiple terms are consistent. It has been recommended that for summative purposes, average scores of several courses taught by the same instructor should be used to enhance generalizability.

1.2.4 POTENTIAL BIASES: MYTHS OR REALITY?

Perceptions of potential biases in student evaluation of teaching can be a source of anxiety for instructors. Important concerns regarding background variables that may threaten validity of student ratings have been discussed extensively in the literature over decades. Some researchers believe these variables introduce biases into the evaluation outcomes, others argue that they are merely sources of validity evidence because they cannot be disconnected from features and dimensions of teaching effectiveness⁴. While caution is required in interpreting effects of background variables on student ratings, there seems to be a trend in the literature indicating that some of these biases are not warranted by research evidence. However, it is important for the Project team to take into account evidence-based potential biases when developing and validating our instruments.

³ For example the scale of an instructor's organization skills consists of four items in SEEQ, a well-researched instrument developed by Herbert Marsh: 1) Instructor's explanations are clear; 2) Course materials are well prepared and carefully explained; 3) Proposed objectives agree with those actually taught so you know where the course is going; 4) Instructor gives lectures that facilitate taking notes.

⁴ For example, the effects of expected grades/grading leniency is one of the most common concerns among faculty. However a student's expected grade from a course is likely to correlate with his/her efforts toward and achievement on the subject.

2 INTRODUCTION

2.1 THE PROJECT

The Teaching and Course Evaluation Project is an initiative sponsored by the Vice-President Academic (VPA) on the recommendation of the Senate Committee on University Teaching and Learning (SCUTL) and the Task Force on Teaching and Learning (TFTL) to develop new instruments and guidelines for student evaluation of teaching and courses⁶.

The purpose of this document is to provide the Simon Fraser University (SFU) community with an overview⁷ of research findings and current practices on student evaluation of teaching and courses with a focus on instrument development and related methodological issues. In order to suit the needs of a wide variety of potential readers (faculty⁸, administrators and students), this report will remain brief and avoid excessive use of technical jargon. For those who are interested in the scientific process applied in the instrumentation exploration and development, a final report documenting the entire instrument development process will be available upon conclusion of the Project.

2.2 THE FIELD

Over the past few decades, much attention has been paid to student evaluation of teaching and courses in higher education. Debates, discussions, and thousands of research articles have been published on the topic. There has been a general understanding in the field of teaching performance in postsecondary institutions that student evaluation is merely one source of evidence to gather on teaching excellence. Other means such as peer and expert observations, teaching philosophy, and teaching portfolios⁹ also bring valuable insights into an understanding of instructor's teaching contributions. Nonetheless, in the past 30-40 years, student evaluation of teaching and courses has been the most predominant form of instructional evaluation in higher education throughout the world¹⁰. Student evaluations are widely used by instructors to improve their teaching (formative) as well as by administrators to make personnel decisions such as merit, tenure and promotions (summative). Recently, more efforts have been made in complementing student ratings with other sources of evidence to better delineate the quality of teaching (such as in Arreola, 2000; Braskamp & Ory, 1994; Knapper & Cranton, 2001).

Although researchers and experts still debate over whether student evaluation is a valid and useful measurement of teaching effectiveness, many empirical studies are generating evidence on valid and

⁶ For a more detailed description of the project, please visit our website at www.sfu.ca/teachingandcourseeval.html or Supporting Document 1.

⁷ Thousands of studies have been conducted on student evaluation of teaching and courses. It is beyond the scope of the Project to comprehensively synthesize the literature. Our goal for this document is to provide a general view on research and practice regarding instrumentation issues. Literature included in this report was obtained through the snowballing technique: the search started with a few articles by key authors in the field, such as Abrami and colleagues, Aleamoni, Cashin, Centra, Cohen, Feldman, Franklin & Theall, Greenwald, Marsh, McKeachie, to name a few. Related research cited in these articles was then obtained through database searches, whose bibliographies again were retained for further capture of related articles. Since a comprehensive survey of the literature is out of our scope, no formal search terms were constructed for database searches. The primary inclusion/exclusion criterion is whether or not a publication is pertinent to instrumentation of student evaluation of teaching. Inclusion/exclusion of articles is also based on a) whether or not it is empirical in nature, b) whether or not it is a literature review, and c) whether or not it is a meta-analysis. Pure theoretical publications on the broad topic of student evaluation or teaching effectiveness are excluded from this report. However, such literature may be included in a separate report prepared by Johanne Provencal. After a few rounds of search, results started to converge to a number of key articles. We are confident that we have located a sufficient amount of research studies, reviews and meta-analyses that congruently delineate the state of the art of the instrumentation of student evaluation of teaching. The project team is aware of the possibility of an omission of related articles due to the vast amount of such items from the literature.

⁸ As previously footnoted in a separate report, the term "faculty" is used to represent tenured and tenure-track faculty members, lab instructors, lecturers and senior lecturers, limited term appointments, and sessional instructors.

⁹ There is a list of 12 sources according to Berk (2005) including student ratings.

¹⁰ See Appendix II: Report of Key Research Findings for a more detailed international context of student evaluation of teaching and courses.

reliable instruments that can be used by students (Sprooner, et al, 2007). In practice, institutions are paying more and more attention to improving and refining their evaluation instruments in compliance with more recent teaching and learning theories as part of their move toward accountability and quality assurance of their institutions¹¹.

¹¹ See Report 1 for a fuller discussion on accountability and quality assurance of postsecondary institutions.

3 MEASURING TEACHING EFFECTIVENESS

3.1 WHAT TO MEASURE?

From a measurement point of view (Wilson, 2005), the central goal of constructing a student evaluation instrument is to provide a reasonable and consistent way to summarize student responses that express their attitudes or personal viewpoints on “teaching effectiveness.” Though no consensus has been reached on what precisely constitutes “good teaching”, there has been a general understanding that it is a multidimensional, complex construct that includes multiple and possibly networked sub-constructs. There have been numerous models available from the literature with different numbers of dimensions that represent teaching effectiveness (for example see Kulik & McKeachiem 1975; Feldman, 1976 and Marsh, 1987). Although there are overlapping factors across these models, no common set of dimensionalities has been agreed upon by researchers in the field. Allin, Clarke, and Jopling (2009) attempted to summarize the literature, including the widely cited studies such as Marsh (1987), and generated a four-dimensional structure of effective teaching with twenty-three sub-dimensions (p364)¹²:

- Supportive Learning Environment, including sub-constructs such as provision of intellectual excitement; high degree of subject knowledge; respect for students; approachability; provision of a motivating environment; recognition of student diversity, etc.
- Academic Expectations, including high level of expected output; explaining to students what they are to learn and why; clarity in standards and assessment criteria; appropriate workload and level of difficulty, etc.
- Scaffolding Learning, including varied ways to teach content; anticipation of misconceptions in students' existing knowledge; appropriate pace for the group being taught; excellent management of student behavior; systematic, well organized and well-structured sessions; students work collaboratively with both their peers and their teachers; effective and timely feedback; effective & sympathetic guidance, etc.
- Clarity, including strong, unambiguous presentation skills; high quality explanation.

For the purpose of this project, such models will be contemplated in order to gain insights for our instrument development process. However, as emphasized by researchers there has been no universal set of characteristics of effective teaching that applies to all teaching situations (see Ory & Ryan, 2001 for example), effective teaching may vary from institution to institution, discipline to discipline (Rando, 2001, Cashin, 1990; Feldman, 1978), and faculty to faculty (Ghedin & Aquario, 2008). Special attention will be given to gathering information on local preferences, priorities and standards pertinent to effective teaching at SFU. Such standards and priorities will be developed based on the literature and SFU's unique identity and evaluation needs at institutional, departmental or unit levels. This rationale has prompted the planning of high level community engagement¹³ as an essential part of the project in seeking advice, consultation, and feedback from all stakeholders before and during the process of developing a sound instrument that can best meet SFU's evaluation needs.

3.2 WHY MEASURE?

Scriven (1991) defines evaluation as “the process, whose duty is the systematic and objective determination of merit, worth, or value. Without such a process, there is no way to distinguish the worthwhile from the worthless” (p. 4). The purpose of evaluating teaching effectiveness is twofold: collecting information for summative and formative purposes. Formative evaluation in the context of teaching quality is intended for instructors to improve their teaching. Research has shown that student

¹² This model is cited here merely as an illustration of the multidimensionality and complexity of teaching effectiveness. The Project is not endorsing superiority of this particular model over any other ones in the literature.

¹³ See another document on community engagement authored by Chris Groeneboer (Manager, Applied Research, Teaching and Learning Centre, SFU).

feedback from evaluation of teaching and courses can help improve quality of instruction (Cohen, 1980; Marsh & Roche, 1993; Menges, 1991; Overall & Marsh, 1979) and consequently student learning (Goldschmid, 1978). It is noteworthy that such benefits do not come automatically without additional efforts, such as open discussions with colleagues and administrators about student feedback (Penny & Coe, 2004; Sporeen, et al, 1997). Faculty may also benefit more from student evaluations if they are provided with training or assistance on data interpretation.

Summative evaluation is used for decisions on “merit, worth, or value”. In this context it means personnel decisions on tenure, promotion, or teaching appointments. Research has shown that student evaluations can provide valuable information on rewarding teaching excellence (Aleamoni, 1981; McKeachie, 1979). However it has been generally understood that student evaluation should not be taken as the only indicator of effective teaching when making personnel decisions. There are many other sources of information to gather on teaching excellence which should be considered in triangulation with student ratings.

A larger context for the use of teaching assessment is as a response to the growing interest of the public to hold postsecondary institutions accountable for student learning. Student ratings have been found to be positively associated with student learning and achievement (Aleamoni & Hexner, 1980; Centra, 1976, 1977, 1987; Cohen, 1981; McKeachie, 1990; Murray, et al, 1990). Although critics argue that some correlations have only been moderate (Derry, 1979; McCallum, 1984) or inconsistent among studies (Dowell & Neal, 1982), there is cause for optimism. For a fuller discussion on such matters, see a separate report.¹⁴

¹⁴ See Report 1 prepared by Johanne Provencal.

4 HOW TO MEASURE: INSTRUMENT DEVELOPMENT

4.1 ITEM CONSTRUCTION

Researchers assert that, item composition and selection is very important in ensuring usefulness when measuring teaching effectiveness (Marsh, 2007). However, in practice, item construction is sometimes not carefully executed when institutions develop their instruments locally. "Poorly worded or inappropriate items will not provide useful information" (Marsh & Roche, 1997, p1187). Researchers advise that evaluation questions should avoid the following pitfalls: double-barreled questions (asking two questions at the same time, such as "Rate the assignments and the exams."), overly complex or ambiguous wording (such as "rate the quality of this course"), or poorly scaled response options (Franklin, 2001). Student evaluations usually include Likert type of scales (e.g., from 1-5), quality rating scales (e.g., fair, good, excellent), and frequency rating scales (e.g., how often does the instructor...?). It is important that students understand the meaning of the scales as intended. For example on a 5-point Likert scale, students may have different interpretations of the middle category 3 and thus affect evaluation outcomes (Ory & Ryan, 2001).

4.2 QUESTIONNAIRE CONTENT AND FORMAT

The content and format of instruments are important factors that determine the evaluation outcomes. Instrument development should be firmly grounded on educational and learning theories, practice and research which have been evolving over time. As emphasized by Theall and Franklin (2000), pedagogical changes in postsecondary education call for updates of evaluation tools that were developed when lecturing was the main form of delivery and students were of homogenous backgrounds. For example, changes in instructional practices include collaborative learning, change in student demographics and online programs, and change in evaluation data collection methods. Items that put too much weight on traditional teaching methods may underrepresent the characteristics of effective teaching in an innovative, active learning environment that employs methods such as collaboration, discussion-oriented instruction as opposed to lectures, and the use of computers for instruction (Ory & Ryan, 2001).

According to the literature, it is desirable to include items on multiple aspects of teaching behaviors due to the multidimensional nature of teaching effectiveness. Multidimensionality implies that an instructor may be proficient with some components of good teaching but not with others (Algozzine et al., 2004; Centra, 1993; Marsh, 1987; Marsh & Roche, 1997), for example, high on clarity but low on enthusiasm; therefore a sound instrument needs to address a range of fundamental elements of successful teaching with representative questions. Although there has been no consensus on what these fundamental elements should be, researchers have preferences based on their epistemologies of teaching and learning, and universities develop their own standards based on their historical contexts. For example, Kansas State University developed Individual Development and Educational Assessment (IDEA) with rating forms for teacher behaviors that include course organization and planning, clarity/communication skills, teacher student interaction/rapport, course difficulty/workload, grading and examinations and student self-rated learning (Braskamp and Ory, 1994). Murray (1987) developed the Teaching Behaviors Inventory, which groups teaching performance into 9 categories: clarity, expression, interaction, organization, pacing, disclosure, speech, rapport, and teaching aids. The well-researched Students' Evaluation of Educational Quality (SEEQ) instrument (Marsh, 1984, 1987; Marsh & Dunkin, 1992) is based on nine factors of effective teaching: learning/value, instructor, enthusiasm, organization/clarity, group interaction, individual rapport, breadth of coverage, examinations/grading, assignments/readings, and workload/difficulty.

In practice, institutions use different formats in structuring and delivering student evaluations. However, there are some common features:

(a) A combination of open-ended and closed-ended questions. Closed-ended questions are designed to collect quantitative responses that can be used for norms, distributions and comparisons. Open-ended

questions are designed to give students the opportunity to further elaborate on their experiences. Rando (2001) offered an example of an open-ended question, entitled 'the one-minute paper': "What's the most important thing you learned and what question are you left with?" Such questions allow more freedom on student responses and can generate feedback on student personal experiences. Such feedback is usually kept confidential to the instructor for formative purposes. Researchers tend to disagree on whether qualitative data should be used for summative purposes. For example, Abrami (2001) asserted that they should not be used for personnel decisions because it is difficult to estimate validity and reliability of qualitative data, while Harper and Kuh (2007) argued otherwise. It is noteworthy that research has shown a correlation between quantitative and qualitative outcomes (Cashin, 1995)

(b) At least one "global" item addressing 'overall' effectiveness (such as "overall I learned a great deal from this instructor") with specific items on multiple perspectives on teaching and course content.

For example, at McGill, student evaluation includes 4 global questions:

- Overall, this is an excellent course.
- Overall, I learned a great deal from this course.
- Overall, this instructor is an excellent teacher.
- Overall, I learned a great deal from this instructor.

In practice, many universities use global results for summative decisions. This approach has been debated in the research literature. Researchers who are skeptical about global items argue that effective teaching is multifaceted; any attempt to compress all information on what effective teaching entails into a single measure is faulty (for example see Frey, 1978). Teachers favor less global items than specific dimensions of teaching effectiveness (Murray, 1987). However, using global items for summative purposes has been supported by empirical evidence from the literature. For example Abrami, d'Apollonia, and Rosenfield (1997) demonstrated with his factor analysis of 225 items (from 17 student evaluation forms) that one dominant, general factor is common across all student ratings. Abrami (2001), Cashin and Downey (1992) assert that items on specific teaching behaviors are useful for instructors in improving their teaching, while for summative purposes, a shorter form contains global items is able to capture enough information needed for personnel decisions. A more compromising view using weighted scores of measured dimensions measured is promoted by Marsh (2007).

(c) Common items with customization. This approach may well be due to the general understanding that different departments or disciplines have different teaching standards and priorities. Most universities allow departments and instructors to customize items in order to get feedback for teaching improvement. Some universities allow more degrees of customization, for example UBC provides a modular framework with 6 common items for all course evaluations, and faculty members and departments can contribute customized evaluation questions. Many universities provide question banks for such customization so that instructors and departments could select from existing items when constructing an evaluation instrument, such as University of Toronto, Queen's and University of Michigan.

4.3 PSYCHOMETRIC TESTING

Experts in the field recommend adhering to psychometric principles and practices when developing instruments for effective measurement of teaching performance (for example see Theall & Franklin, 2001). Professionally developed instruments (such as IDEA, SEEQ) usually go through rigorous validation processes involving large sample sizes to ensure sound psychometric qualities. Student evaluation instruments designed by university committees or instructors seldom go through psychometric testing and the relevant revisions (Marsh, 2007; Marsh & Roche, 1997). Researchers in the field proclaim that psychometric analysis is essential even if the final product is an adapted form of a well evaluated instrument, because the latter may not retain the validity of the former in the settings where the instrument is intended to be used (Ory and Ryan, 2001).

4.3.1 VALIDITY

Validity indicates whether or not an instrument measures what it is designed to measure (Wilson, 2005). Validity is the most common concern of student evaluations of teaching and courses in higher education. There is a lack of consensus in the literature regarding validity of student ratings of instruction. Some research has shown that it is more valid than other evaluation tools, although not perfect (such as Cashin, 1988), while others treat student evaluations as only satisfaction or opinion surveys that do not reflect teaching quality.

Despite the evident disputes, there seems to be a trend emerging from the literature regarding validity of student ratings. As noted by Greenwald (1997) *"the validity of student rating measures of instructional quality was severely questioned in the 1970s. By the early 1980s, however, most expert opinion viewed student ratings as valid and as worthy of widespread use"* (p.1182). Research syntheses continue to show plausible results in demonstrating usefulness of student ratings in determining teaching effectiveness. For example, meta-analyses (d'Apollonia and Abrami, 1996, 1997a, 1997b) showed moderate to high correlations between student ratings of instruction and student achievement (an average of .47). Feldman (1997) also reported correlations from .34-.57 between multiple dimensions of effective teaching and student achievement. And Cohen (1987) reported correlations from .28-.55 across dimensions of effective teaching and achievement, and an average of .43 between achievement and overall rating of the teacher effectiveness (Cohen, 1981) between student achievement and various perspectives of student ratings of teaching effectiveness.¹⁵ Researchers are well aware of the limitations of measures of achievement and the complex and imperfect relations between student ratings and their examination performance (Scriven, 1983; Kulik, 2001). Factors other than effective teaching may affect students' performance in exams, for example as exerted in Kulik (2001), *"Bad teachers sometimes put unreasonable pressure on students, and that unethical behavior may produce maximum exam scores."* (p28).

It is important to understand that student achievement, as an important criterion for teaching effectiveness, should not be used as the only condition to determine good teaching (Marsh, 2007). There has been little research examining relations between student evaluation with other aspects of student outcomes such as motivation, study strategies, and career aspirations (Marsh & Roche, 1997).

Nevertheless, researchers are optimistic that student ratings of teaching effectiveness is at least one measure that reflects teaching quality (Centra, 1993). Furthermore, research has provided evidence that student ratings are consistent with other measures, such as trained classroom observers based on instructor behaviors instead of invalid indicators such as personality (Centra, 1993; Murray, 1980), and former student ratings (Howard, Conway & Maxwell, 1985). Instructor self-evaluations also correlate with student ratings (Feldman, 1989; Marsh, Overall & Kesler, 1979) and there is general agreement between students and instructors on the dimensionality of student ratings (Marsh, 2007).

In practice, faculty opinions on student evaluation vary from taking it as reliable, valid and useful to unreliable, invalid and useless (Aleamoni, 1981). Some even consider student ratings as meaningless quantification that leads to personality or popularity contests (Haskell, 1997; Sproule, 2002; Kulik, 2001; Ory, 2001). As a consequence, the use of student ratings may reduce faculty morale (Ryan, Anderson & Birchler, 1980) or encourage undesirable behaviors such as lowering course requirements and become a "teaching teddy bear" (Sacks, 1986), which in turn may compromise student learning. These concerns suggest a disconnection between research and practice as noted by Theall and Franklin (2000). While some concerns may be valid according to researchers, for example, students may not be qualified to judge whether course content is cutting edge or outdated, students are believed to have the ability to

¹⁵ There has also been research arguing otherwise, for example Rodin & Rodin (1972) reported a negative correlation of .75 between student rating and learning. However such studies are often found to be flawed in methodology (Doyle, 1975; Marsh, 1984).

evaluate aspects such as assignments and interactions between instructor and students in and out of classes (Ory & Ryan, 2001).

The Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, National Council of Measurement in Education) had recently updated the traditional validity framework, which contains content, criterion and construct validity, with a unitary concept that measures the extent to which any evidence supports the intended use and inferences of assessment results (Messick 1989, 1995). According to (Ory & Ryan, 2001, p28), "*construct validity should incorporate any evidence that affects the meaning and interpretation of the assessment scores.*" Within this framework, validity of student evaluation of teaching is primarily about its inferences on teaching effectiveness and the use and interpretation of the results (2001). There are multiple types of validity evidence, such as content, external, and consequential, etc (2001) which will be discussed more in detail in the final report for those who are interested in the scientific processes of instrument validation. There are also background variables or "noise" that many researchers believe would threaten the validity of student ratings. These variables, or "biases", have been studied and debated extensively in the literature and will be discussed in a later section in this report. Researchers are more in agreement that the misuse of evaluation results is the major validity threat (for example using the scores to the third decimal points to rank faculty—Theall & Franklin, 2001).

One notable gap in the research literature is how students interpret the evaluation questionnaires, which is an important aspect of validity. For example, do students compare the instructor being evaluated to other instructors when they respond to the evaluation items or do they have a personal standard in mind? And are they well informed with how the evaluation results are going to be used for summative and formative purposes when they respond to the evaluation items. These factors may affect student ratings, for example, research has shown that student ratings are slightly higher when they are told how the results will be used for personnel decisions (Centra, 1976; Feldman, 1979; Overall and Marsh, 1979). Another aspect of validity of student rating resides in the intended and unintended, positive and negative consequences of the interpretation and use of rating scores. Ory & Ryan, (2001) note that this is relatively unexamined territory in the literature and recommend institutions investigate methods and procedures that help enhance intended and positive consequences (such as improving teaching) and reduce unintended and negative ones (such as weakening the curriculum).

4.3.2 RELIABILITY

Reliability is an important aspect of validity. Researchers generally believe that a student evaluation instrument can be reliable (Marsh, 1987), especially when the instrument is well designed and tested according to psychometric principles (Centra, 1993; Aleamoni, 1987; Marsh, 1984). Reliability of a student evaluation tool is usually determined by inter-rater agreement (agreement across different students in the same class) and internal consistency of a scale (correlations among items measuring the same element of effective teaching, such as an instructor's organization skills, see for example, Abrami, 2001; Theall & Franklin, 2001; Wachtel, 1998; Goldschmid, 1978). Research has shown good reliability of teaching evaluations. For example, Feldman (1977) reported that reliability coefficients are usually .80 or higher (with higher coefficients indicating higher reliability). Student ratings also have good internal consistency, that is, the items correlate well with one another. Evaluation scores are also highly correlated across administrations (Ory & Ryan, 2001). There are other aspects of reliability that are not usually examined in teaching evaluations, such as alternative forms measuring the same teaching behaviors (Abrami, 2001).

There has also been abundant evidence demonstrating good stability (rater agreement over time). Longitudinal studies comparing immediate ratings with ratings at least a year after show substantial correlations, for example, in studies by Marsh and Overall (1979) and Overall and Marsh (1980), the average correlation is .83 with 100 courses. In a 13-year longitudinal study by Marsh and Hocevar (1991), no substantial changes of teaching evaluation were found across 195 instructors. Centra (1974)

also reported a correlation of .75 between ratings by students and alumni who had graduated up to 5 years previous.

Generalizability refers to reliable evaluation of an instructor's general teaching effectiveness rather than one snapshot of his/her performance in a particular course at a particular time. Marsh (1984) investigated this factor by examining 1,364 courses across 4 conditions: a) the same instructor teaching the same courses in two different terms; b) the same instructor teaching two different courses; c) the same course taught by different instructors; and d) two different courses taught by different instructors. Findings indicated that student ratings reflect teaching effectiveness of the instructor rather than the course itself. Researchers recommend that for summative purposes, average scores of several courses taught by the same instructor should be used to enhance generalizability (such as Gilmore, Kane & Naccarato, 1978).

4.4 INSTRUMENT IMPLEMENTATION

Seldin (1993) asserts that poor administration can invalidate a good instrument. Therefore the implementation process is important in ensuring student evaluation will generate useful outcomes. Some common features of current practices in North American universities are:

(a) Student responses are anonymous. Anonymity is an important factor that affects evaluation outcomes. Centra (1979) noted that students who identify themselves tend to give more favorable evaluations, especially if the evaluation is returned to the instructor before final grades. Therefore universities are almost uniformly adopting an anonymous approach and returning feedback to instructors after final grades.

(b) Responses are obtained at the end of the term in the absence of the instructor. Universities usually open student evaluations 2-4 weeks before the final exams to allow adequate time for students to reflect on their evaluations. Research has shown that student ratings are slightly higher when the instructor remains in the room during the evaluation process (Stone, Spool, and Rabinowitz, 1977; Feldman, 1979).

(c) Online vs. paper form. Universities are using different formats of evaluation based on different assumptions of how each type would yield desirable evaluation results. The online formatting provides the ease of response, flexible timing, better reflection and easy data entry. However researchers are concerned that this formatting may not generate the best possible response rate. In the literature, both formats are proven to have their advantages and disadvantages, and no consensus has been reached. Similarly in practice, though most universities are moving their evaluations online (for example in spring 2011, MIT moved to a completely online subject evaluation system), some institutions are doing the opposite. For example, In 2008/09 University of Calgary moved away from an online evaluation process to one that is paper-based and in-class. And only where the paper survey cannot be used (e.g., web-based courses) is an online survey administered.

4.5 USE AND INTERPRETATION OF STUDENT EVALUATION DATA

Student ratings are typically used in the following ways in a postsecondary institution (Ory & Ryan, 2001, p39):

- To help the faculty make teaching and course improvements
- To help the administrators make personnel decisions regarding such things as salary and promotion
- To help the campus select campus teaching award winners
- To help the campus conduct program reviews
- To help the students select courses

The use of the data, especially for personnel decisions is the most contentious aspect of validity of an instrument in the field. Researchers such as McKeachie (1997) suggested that more attention should be directed toward methods of ensuring more valid use. Usually universities norm the global scores as benchmarks (norm-based evaluations), or compare individual instructor's scores against a pre-determined

standard (criterion-based evaluation), and keep verbal responses confidential to department heads and instructors. Student evaluation is recommended to be used and interpreted in triangulation with other source of information such as teaching portfolios and peer evaluation.

Researchers express that despite the concerns and research findings, in practice teaching evaluations are primarily used for summative personnel decisions (for example see Beran, Violato, & Kline, 2007; Wagenaar, 1995). Although theoretically student evaluations provide valuable feedback on teaching behaviors, research found that student evaluation results are seldom used by faculty in improving their teaching quality.

Abrami (2001) recommended strategies to improve use of student rating data, such as statistical and administration strategies, and set criteria/standards in advance. For example whether to use criteria- or norm-referenced approaches, averaging global items and scores for courses combined, weighing, and data presentation format such as graphs.

5 POTENTIAL BIASES: MYTHS OR REALITY?

A plethora of research has been focusing on variables that could potentially threaten validity of student ratings by introducing biases to the evaluation outcome (as described by Abrami, 2001, Marsh, 2007, Marsh & Roche, 1997; Theall & Franklin, 2001). Researchers have been debating on whether or not these variables could in fact bias student ratings or their relation to evaluation scores are simply evidence for construct validity (Marsh, 2007; Centra, 2003)¹⁶. Much research has suggested that most of these variables are not in fact conclusive biases, but myths (Aleamoni, 1999) at most. Marsh (2007) uses a more neutral term "*background characteristics*" before concluding whether or not a variable introduces bias to student evaluations, such as class size, workload/difficulty, and grading leniency. He pointed out that most bias research is flawed in methodology and unclear with defining what constitutes bias¹⁷.

Perceptions of potential biases of student evaluation of teaching have been a source of anxiety for instructors (Franklin & Theall, 1989). Such perceptions are based on assumptions such as students are not qualified to evaluate teaching (Nasser & Fresko, 2002); student evaluation is in fact a popularity contest where individuals who are warm, charismatic and give students higher grades are always winners (Aleamoni, 1999; Baldwin & Blattner, 2003; Williams & Ceci, 1997). Although there has been abundant research refuting such assumptions, in practice anxiety and concerns remain (Aleamoni, 1999; Ory, 2001; Theall & Feldman, 2007).

Marsh (2007, p348-349) provided a list of most common concerns from the literature and their effects on student evaluations. Other than the apparent methodological flaws, most effects found on these variables are either small, inconsistent, or inconclusive¹⁸. More importantly, it is argued that most of these variables do not constitute bias toward the validity of student ratings¹⁹.

5.1 PRIOR SUBJECT INTEREST

This variable has been found to correlate with certain aspects of effective teaching, such as learning/value (correlation sizes approximately .4). Thus it is argued to be a source of validity instead of a bias. And the same effects hold across measures of teaching effectiveness, such as instructor self-evaluation.

5.2 EXPECTED GRADE/GIVEN GRADES

Given the widespread concerns of this specific effect, many approaches have been taken to investigate this potential bias in the literature. Correlations between this variable and student ratings tend to be small (.11-.20, Marsh, 2007) or inconsistent. It is difficult to separate the effects between grading leniency and students received grades, which may reflect student learning. Measures of grading leniency are also under question. Additionally, research that reports positive correlations between grades and ratings, for example Greenwald and Gilmore, (1997), often neglects how grading standards vary across subjects, disciplines, and course levels, and consequently deems the bias claim non-generalizable (Kulik,

¹⁶ This will be discussed in more details in the Final Report after data collection and analysis for the Project.

¹⁷ In fact some concerns have been proven to be based on poorly designed research or irrelevant to student rating, discussions and debates on such topics are now often deemed passé (Gravestock & Gregor-Greenleaf, 2008), such as the "Dr. Fox effect": In study by Naftulin, Ware, and Donnelly (1973), a professional actor was hired to deliver a lecture on mathematical game theory to medical educators posing as an expert in the field. His lecture was rated favorably by the audience even though it involves mainly incorrect information. This effect caused by showmanship is termed as the "Dr. Fox Effect".

¹⁸ An in-depth discussion of potential biases is beyond the scope of this report. A few variables are briefly discussed here based on their relation to student ratings as examples. Please see Marsh (2007) and Centra, (2003) for closer examinations of the variables. These variables will be taken into consideration when developing and validating our instruments. More thorough discussions will be provided in the Final Report for those who are interested in the validation process of our instrument.

¹⁹ Marsh (2007) argues that: "If a potential biasing factor actually does have a valid influence on teaching effectiveness and this influence is evident in different indicators of teaching effectiveness... then it may be possible that the influence reflects support for the validity of SETs(student evaluation of teaching), rather than a bias (p349)."

2001). All the research suggests that this variable may generate only weak, if any, biasing effects toward student ratings.

5.3 WORKLOAD/DIFFICULTY

In contrast to common concerns that instructors who assign more work and give difficult assignments tend to receive less favorable ratings, some research found that there is a positive correlation between workload/difficulty and student ratings. Instructors' self-evaluation shows consistent effects with this variable. Although further analyses show that the relationship of this variable and student ratings can be more complicated and non-linear, Marsh (2007) argued that it is not a biasing factor toward student ratings.

5.4 CLASS SIZE

This variable is often found to have a small negative correlation with student ratings and moderate effects on dimensions such as group interactions and individual rapport. Logically a smaller class allows the instructor to allocate more time and attention to each student. The relationship between class size and student ratings tends to be non-linear, with very large classes receiving similar student ratings as small classes. Caution is needed when interpreting such effects. For example favorable student ratings in very large classes may reflect appropriate teaching strategies in such settings. Additionally instructors' self-evaluation tends to be consistent with student ratings on specific aspects of effective teaching, suggesting that this variable is a source of evidence for validity rather than a bias.

In order to alleviate possible anxieties from faculty it is important to acknowledge to the SFU community that the project team is aware of various sources of possible biases and factors that may threaten validity when developing our evaluation instruments. These can be controlled, for example, through data triangulation and appropriately controlling for differences in student ratings associated with these background characteristics when interpreting evaluation results.

Bear in mind that even if all the above concerns were addressed by research results, experts caution that *"...data can be and are misused on a regular basis. Even if ratings results were perfectly reliable and valid (and no educational, psychological, or sociological instrument provides data that are perfect), misuse would still be a major problem."* (Theall & Franklin, 2001, p46). It is important to inform the SFU community that our project is not striving for perfection, but is developing a valid and reliable instrument that best suits SFU's evaluation needs by closely examining the relevant research literature, as well as through a thorough community engagement during the process of instrument development and implementation.

6 REFERENCES

- Abrami, P. C. (2001). Improving judgments about teaching effectiveness using teacher rating forms. In M. Theall, P. C. Abrami, & L. A. Mets (Eds.), *The student ratings debate: Are they valid? How can we best use them? New Directions for Institutional Research*, No. 109 (pp. 59-87). San Francisco: Jossey-Bass.
- Abrami, P.C., d'Apollonia, S., & Rosenfield, S. (1997). The dimensionality of student ratings of instruction: What we know and what we do not. In R.P. Perry & J.C. Smart (Eds.). *Effective teaching in higher education: Research and practice* (pp. 321- 367). Bronx, N.Y.: Agathon Press.
- Aleamoni, L. M. & Hexner, P. Z. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. *Instructional Science*, 9(1), 67-84.
- Aleamoni, L. M. (1981). Student ratings of instruction. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 110-145). Beverly Hills, CA: Sage Publications.
- Aleamoni, L. M. (1987). Typical faculty concerns about student evaluation of teaching. *New Directions for Teaching and Learning*, 31, 25-31.
- Aleamoni, L.M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, 13(2), 153-166.
- Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., Mohanty, G., & Spooner, F. (2004). Student evaluation of college teaching: A practice in search of principles. *College Teaching*, 52(4), 134-141.
- Allan J, Clarke K & Jopling, M. (2009). Notions of effective teaching: An exploration of students' and lecturers' perceptions of first year education undergraduate. *International Journal of Teaching and Learning in Higher Education*, 21, 362-372.
- Arreola, R.A. (2000). *Developing a comprehensive faculty evaluation system* (2nd ed.). Bolton, MA: Anker.
- Baldwin, T. & Blattner, N. (2003). Guarding against potential bias in student evaluations: What every faculty member needs to know. *College Teaching*, 51(1), 27-32.
- Benton, S. & Cashin, W.E. (2010). *Student Ratings of Teaching. A Summary of Research*. (IDEA paper No. 50). Kansas State University.
- Beran, T., Violato, C., & Kline, D. (2007). What's the 'use' of student ratings of instruction for administrators? One university's experience. *Canadian Journal of Higher Education*, 17(1), 27-43.
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17, 48-62.
- Braskamp, L.A. & Ory, J.C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco: Jossey-Bass.
- Cashin, W. E. (1990). Students do rate different academic fields differently. In M. Theall, & J. Franklin (Eds.), *Student ratings of instruction: Issues for improving practice: New Directions for Teaching and Learning*, No. 43 (pp. 113-121). San Francisco: Jossey-Bass.
- Cashin, W. E. (1995). *Student ratings of teaching: The research revisited*. IDEA Paper No. 32. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Cashin, W. E., & Downey, R. G. (1992). Using global student ratings for summative evaluation. *Journal of Educational Psychology*, 84, 563-572. Cashin, 1988
- Centra, J. A. (1976). The influence of different directions on student ratings of instruction. *Journal of Educational Measurement*, 13, 277-282.
- Centra, J. A. (1977). Student ratings of instruction and their relationship to student learning. *American Educational Research Journal*. 14, 17-24.(b)
- Centra, J. A. (1978). Using student assessments to improve performance and validity. In W. R. Kirschling (ed.), *Evaluating Faculty Performance and Validity*. New Directions for Institutional research, no. 20. San Francisco: Jossey-Bass.
- Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.

- Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44(5), 495-518.
- Centra, J. A. The how and why of evaluating teaching. *New Directions for Higher Education*, 1977, 7, 67-78.(a)
- Centra, J.A. (1979). *Determining faculty effectiveness: Assessing teaching, research, and service for personnel decisions and improvement*. San Francisco: Jossey-Bass. Centra, J. A. (1976). The influence of different directions on student ratings of instruction. *Journal of Educational Measurement*, 13, 277-282.
- Cizek, G.J., Rosenberg, S.L., & Koons, H.H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement*, 68, 397-412.
- Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education*, 13, 321-341.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309.
- Cohen, P. A. (1987, April). A critical analysis and reanalysis of the multisection validity meta-analysis. Paper presented at the 1987 annual meeting of the American Educational Research Association. Washington, DC (ERIC Document Reproduction Service No. ED 283 876).
- d'Apollonia, S., & Abrami, P.C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198-1208.
- d'Apollonia, S., and Abrami, P. C. (1996). Variables moderating the validity of student ratings of instruction: A meta-analysis. Paper presented at the 77th Annual Meeting of the American Educational Research Association, New York.
- Derry, J. O. (1979). Can student's ratings of instruction serve rival purposes? *Journal of Higher Education*, 50, pp. 79-88.
- Dowell, D.A. & Neal, J.A. (1982). A selective review of the validity of student ratings of teaching. *Journal of Higher Education*, 53, 51-62.
- Doyle, K. (1975). Student evaluation of instruction. *D.C. Heath*, Lexington, MA.
- Feldman, K. (1979). The significance of circumstances for college students' ratings of their teachers and courses: A review and analysis. *Research in Higher Education*, 10, 149-172.
- Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education*, 9, 199-242.
- Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education*, 10, 149-172.
- Feldman, K. A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 368-395). New York: Agathon Press.
- Franklin, J. (2001). Interpreting the numbers: Using a narrative to help others read student evaluations of your teaching accurately. In K.G. Lewis (Ed.), *Techniques and strategies for interpreting student evaluations* [Special issue]. *New Directions for Teaching and Learning*, 87, 85-100.
- Franklin, J., & Theall, M. (1989). *Who reads ratings: Knowledge, attitude and practice of users of student ratings of instruction*. Paper presented at the Annual meeting of the American Educational Research Association, San Francisco.
- Frey, P. (1978). A two-dimensional analysis of student ratings of instruction. *Research in Higher Education*, 9, 69-91.
- Ghedini, E., & Aquario, D. (2008). Moving towards multidimensional evaluation of teaching in higher education: A study across four faculties. *Higher Education*, 56, 583-597.
- Goldschmid, M.L. (1978). The evaluation and improvement of teaching in higher education. *Higher Education*, 7(2), 221-245.
- Gravestock, P. & Gregor-Greenleaf, E. (2008). *Student Course Evaluations: Research, Models and Trends*. Toronto: Higher Education Quality Council of Ontario.
- Greenwald, A. G. & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, 89, 743-751.

- Harper, S.R., & Kuh, G. (2007). Myths and misconceptions about using qualitative methods in assessment. In S.R. Harper & S.D. Museus (Eds.), *Using qualitative methods in institutional assessment* [Special issue]. *New Directions for Institutional Research*, 136, 5-14.
- Haskell, R.E. (1997). Academic freedom, tenure, and the student evaluation of faculty: Galloping polls in the 21st century. *Education Policy Analysis*, 5(6).
- Howard, G. S., Conway, C., & Maxwell, S. (1985). Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology*, 77, 187-196.
- Knapper, C., & Cranton, P. (Eds.). (2001). *Fresh approaches to the evaluation of teaching* (New Directions for Teaching and Learning, No. 88). San Francisco: Jossey-Bass.
- Kulik, J. A. (2001). Student ratings: Validity, utility, and controversy. In M. Theall, P. C. Abrami, & L. A. Mets (Eds.), *The student ratings debate: Are they valid? How can we best use them?* *New Directions for Institutional Research*, No. 109 (pp. 9-25). San Francisco: Jossey-Bass.
- Marsh, H. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.
- Marsh, H. W., & Dunkin, M. J. (1997). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 241-320). New York: Agathon Press.
- Marsh, H. W., & Hocevar, D. (1991). Students' evaluations of teaching effectiveness: The stability of mean ratings of the same teachers over a 13-year period. *Teaching & Teacher Education*, 7, 303-314.
- Marsh, H. W., & Overall, J. U. (1979). Long-term stability of students' evaluations: A note on Feldman's consistency and variability among college students in rating their teachers and courses. *Research in Higher Education*, 10, 139-147.
- Marsh, H. W., & 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.
- Marsh, H., Overall, J. & Kesler, S. (1979). Validity of student evaluations of instructional effectiveness: A comparison of faculty self-evaluations and evaluations by their students. *Journal of Educational Psychology*, 71, 149-160.
- Marsh, H.W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 253-388.
- Marsh, H.W., & Roche, L. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias and utility. *American Psychologist*, 52(11), 1187-1197.
- Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The Scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-383). Dordrecht, The Netherlands: Springer.
- McCallum, L. W. (1984) A meta-analysis of course evaluation data and its use in the tenure decision, *Research in Higher Education*, 21, pp. 150-158.
- McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe*, 65, 384-397.
- McKeachie, W. J. (1990) Research on college teaching: The historical background, *Journal of Educational Psychology*, 82, pp. 189-200.
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52, 1218-1225.
- Menges, R. J. (1991) The real world of teaching improvement: A faculty perspective, in: M. Theall & J. Franklin (Eds). *Effective Practices for Improving Teaching*, *New Directions for Teaching and Learning*, Vol. 48, pp. 21-37 (San Francisco, Jossey-Bass).
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). Old Tappan, NJ: Macmillan, 1989.

- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82, 250–261.
- Murray, H.G. (1987). Acquiring student feedback that improves instruction. In Weimer, M.G. (Ed.), Teaching large classes well [Special issue]. *New Directions for Teaching and Learning*, 32, 85–96.
- Naftulin, D. H., Ware, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 48(7), 630–635.
- Nasser, F. & Fresko, B. (2002). Faculty view of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, 27(2), 187–198.
- Ory, J. C., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? In T. D. Johnson & D. L. Sorenson (Eds.), *Online student ratings of instruction: New Directions for Teaching and Learning*, No. 5 (pp. 27–44). San Francisco: Jossey-Bass.
- Ory, J.C. (2001). Faculty thoughts and concerns about student ratings. In K.G. Lewis (Ed.), Techniques and strategies for interpreting student evaluations [Special issue]. *New Directions for Teaching and Learning*, 87, 3–15.
- Overall, J. U., & Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology*, 72, 321–325.
- Overall, J., & Marsh, H. (1979). Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. *Journal of Educational Psychology*, 71, 856–865.
- Penny, A. R., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: Meta-analysis. *Review of Educational Research*, 74, 215–253.
- Rando, W. L. (2001). Writing teaching assessment questions for precision and reflection. *New Directions for Teaching and Learning*, 2001: 77–83. doi: 10.1002/tl.30.
- Ryan, J.J., Anderson, J.A., & Birchler, A.B. (1980). Student evaluation: The faculty responds. *Research in Higher Education*, 12(4), 317–333.
- Sacks, P. (1986). *Generation X goes to college: An eye-opening account of teaching in postmodern America*. LaSalle, IL: Open Court.
- Scriven, M. (1991). *Evaluation Thesaurus*. (4th ed.) Thousand Oaks, Calif.: Sage.
- Seldin, P. (1993). *Successful use of teaching portfolios*. Bolton, MA: Anker.
- Spooren, P., Mortelmans, D., & Denekens, J. (2007). Student evaluation of teaching quality in higher education: Development of an instrument based on 10 likert-scales. *Assessment & Evaluation in Higher Education*, 32(6), 667–679.
- Sproule, R. (2000). Student evaluation of teaching: A methodological critique of evaluation practices. *Education Policy Analysis*, 8(50).
- Sproule, R. (2002). The under-determination of instructor performance by data from the student evaluation of teaching. *Economics of Education Review*, 21(3), 287–295.
- Stone, F., Spool, M. & Rabinowitz, S. (1977). Effects of anonymity and retaliatory potential on student evaluations of faculty performance. *Research in Higher Education*, 6, 313–325.
- Theall, M., & Feldman, K. A. (2007). Commentary and update on Feldman's (1997) "Identifying exemplary teachers and teaching: Evidence from student ratings." In R. P. Perry & J. C. Smart (Eds.), *The teaching and learning in higher education: An evidence-based perspective* (pp. 130–143). Dordrecht, The Netherlands: Springer.
- Theall, M., & Franklin, J. (2000). Creating responsive student ratings systems to improve evaluation practice. In K.E. Ryan (Ed.), Evaluating teaching in higher education: A vision for the future [Special issue]. *New Directions for Teaching and Learning*, 83, 95–107.
- Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? In M. Theall, P.C Abrami, & L.A. Mets (Eds.), The student ratings debate: Are they valid? How can we best use them? [Special issue]. *New Directions for Institutional Research*, 109, 45–56.

- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23, 191-211.
- Wagenaar, T.C. (1995). Student evaluation of teaching: Some cautions and suggestions. *Teaching Sociology*, 23(1), 64-68.
- Williams, W. M., & Ceci, S. J. (1997). "How'm I doing?" problems with student ratings of instructors and courses. *Change*, 29(5), 12-23.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.



SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING & COURSES

TEACHING AND COURSE EVALUATION PROJECT
FINAL REPORT

APPENDIX IV: INSTITUTION CONSULTATION KEY
FINDINGS SUMMARY

NOVEMBER 26, 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

DOCUMENT CONTROL

DOCUMENT INFORMATION

	INFORMATION
Document Owner	<i>Corinne Pitre-Hayes</i>
Project /Organization Role	<i>Project Manager, Teaching & Course Evaluation Project</i>

DOCUMENT HISTORY

VERSION	ISSUE DATE	CHANGES
<i>0.1</i>	<i>February 11, 2013</i>	<i>Initial draft</i>
<i>0.2</i>	<i>February 20, 2013</i>	<i>Revisions per input</i>
<i>0.3</i>	<i>February 28, 2013</i>	<i>Add institutions</i>
<i>0.4</i>	<i>March 4, 2013</i>	<i>Revisions per input</i>
<i>0.5</i>	<i>November 4, 2013</i>	<i>Additional interview findings, further revisions</i>
<i>0.6</i>	<i>November 12, 2013</i>	<i>Revisions from final team review</i>
<i>0.7</i>	<i>November 26, 2013</i>	<i>Revisions from SCUTL feedback</i>

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	1
2	INTRODUCTION	1
3	CONSULTATION OVERVIEW	1
3.1	LEADING INSTITUTIONS.....	1
3.2	COMPETITOR INSTITUTIONS	2
3.3	COMPARABLE INSTITUTIONS	2
3.4	OTHER INSTITUTIONS.....	2
3.5	INSTITUTIONS INTERVIEWED AND CATEGORIES REPRESENTED	2
4	ONLINE VS PAPER BASED INSTRUMENTS	4
5	USE OF DATA	5
5.1	SHARING EVALUATION DATA.....	5
5.2	STUDENT ACCESS TO EVALUATION RESULTS	6
5.3	FORMATIVE VERSUS SUMMATIVE USE OF THE DATA	6
6	RESPONSE RATES	8
7	BEST PRACTICES	8
7.1	DEFINE GOVERNING PRINCIPLES / POLICIES	8
7.2	ADOPT A FLEXIBLE FRAMEWORK.....	8
7.3	NEED FOR STRUCTURED SUPPORT	9
7.4	TAKE A CAREFUL, THOUGHTFUL APPROACH.....	9
8	FURTHER INVESTIGATION AS PART OF THE PROOF OF CONCEPT	9

1 EXECUTIVE SUMMARY

This report provides an analysis and summary of phone interviews conducted with a number of key contacts involved in the student evaluation of teaching and courses at other institutions. Interviews were conducted with 23 different institutions that were categorized as leaders, competitors, comparables and others.

The emphasis was on learning from the experiences of these institutions as SFU embarks on a process to replace its form(s) for student evaluation of teaching and courses. Topics related to evaluation frameworks, processes used, types of instruments, use of evaluation data and best practices were explored.

Key best practices that emerged from discussions with those interviewed included:

- Defining governing principles and implementing policies that align with these principles
- Adopting a flexible framework capable of meeting the needs of stakeholders
- Providing structured support from a designated department in the institution
- Generally, taking a careful, thoughtful approach.

Consultation with other institutions raised questions that the team explored further in the Proof of Concept phase of the project.

2 INTRODUCTION

A key reason for reviewing and updating student evaluation of teaching and courses at SFU is the desire to take advantage of more recent knowledge and understanding in this area – both nationally and internationally. To that end, a key component of our research has included consultation with other institutions to learn about their experiences in this regard.

Our goal was to use this information to help inform our recommendations to Senate for instrument selection and related processes. When combined with the rest of our research data, it contributes greatly to a well-developed outline of recommended practices. This report provides a summary of key findings from this stage of our research.

3 CONSULTATION OVERVIEW

The first step in our consultation with other institutions involved obtaining and documenting as much information as possible from the public domain about each institution and their respective details in regard to student evaluation of teaching and courses. Then interviews were conducted with 25 individuals at 23 different institutions. The focus of these interviews was to learn as much as possible from their *experiences* in this area.

A standard set of questions was developed and feedback on these questions requested from each interviewee. The topics explored included:

- The principles / policies that the evaluation framework is built on
- The type of instrument used
- How the evaluation process works
- Use of evaluation data
- Communication with students regarding evaluations
- Requirements to maintain / run the evaluation process.

Institutions interviewed were categorized into four groups: leaders, competitors, comparables and others. Selecting institutions based on their placement in one of these key groups provided the appropriate perspective for each interview – enabling us to tailor the focus and dialogue for each discussion. This approach helped elicit a cross-section of institutional input and representation. The definitions of these four groups are summarized in the sections below.

3.1 LEADING INSTITUTIONS

Institutions in this category are well-regarded leaders in the use of teaching and course evaluations. A review of their documentation and literature indicates that they have based their processes on sound learning, research, validation and testing. Their instrument has effectively proven to meet institutional needs and is also well-received by students, faculty members and administrative personnel.

3.2 COMPETITOR INSTITUTIONS

This group represents institutions within B.C. that, in practical terms, SFU competes with for students.

3.3 COMPARABLE INSTITUTIONS

A subset of institutions based on random selection from the [2013 Maclean's Comprehensive University Rankings](#). Maclean's defines this type of institution as having "a significant degree of research activity and a wide range of undergraduate and graduate programs, including professional degrees."





3.4 OTHER INSTITUTIONS








Institutions outside of the categories mentioned above that were referred to the team as worthwhile contacts.

3.5 INSTITUTIONS INTERVIEWED AND CATEGORIES REPRESENTED

The institutions interviewed and the categories represented are summarized in Table 1.

Table 1: Institutions interviewed and categories represented

Legend			
	Leaders		
	Competitors		
	Comparables		
	Others		

Type	Name	Students	Program and Faculty Information
	Brock University	18,190 students	7 Academic Faculties 582 faculty members
	Griffith University	Over 43,000 students	More than 300 degrees including undergraduate, postgraduate and research degrees
	Harvard University	About 21,000 students	11 principal Academic Units About 2,100 faculty members
	Kwantlen Polytechnic University	More than 17,500 students	Over 120 programs More than 1,600 faculty and staff
	McGill University	37,835 students	1,636 tenured and tenure-stream faculty members 11 Academic Faculties and 11 schools
	McMaster University	21,025 full-time undergraduate students 3,436 full-time graduate students	1,377 full-time faculty
	Memorial University of Newfoundland	18,000 full and part-time students	950 faculty members 850 session instructors 13 Academic Faculties and schools across four campuses

Type	Name	Students	Program and Faculty Information
★	Ryerson University	38, 950 students	More than 100 undergraduate and graduate programs Nearly 2,700 faculty and staff
◆	Saint Mary's University	7,272 students	438 faculty members
○	Stanford University	6,999 undergraduate students 8,871 graduate students	Of the seven schools at Stanford, three award undergraduate degrees: Humanities and Sciences, Earth Sciences and Engineering.
◆	University of Alberta	Approximately 38,800 students	Close to 400 undergraduate, graduate and professional programs 18 Academic Faculties
□	University of British Columbia	43,500 undergraduate and 10,450 graduate students	33 Academic Faculties and schools across two major campuses
○	University of Cambridge	18,994 full time students across undergraduate and post-graduate	31 colleges and over 150 departments, Academic Faculties, schools and other institutions
◆	University of Louisville	15,893 undergraduate students 6,400 graduate students 126 post-doctoral students	12 colleges and schools across 3 campuses 6,901 faculty and staff
○	University of Michigan	59,933 students	8,791 instructional staff 28 schools and colleges across 3 campuses
◆	University of Pennsylvania	24,725 students	4318 faculty members 16 academic and professional schools
◆	University of Saskatchewan	About 20,000 students	13 interdisciplinary colleges, 3 graduate schools and 6 affiliated / federated colleges
○	University of Toronto	65,612 undergraduate students 15,287 graduate students	11,581 faculty members Approximately 709 undergraduate, 215 graduate and 63 professional programs
□★	University of Victoria	20,199 undergraduate and graduate students	12 Academic Faculties and divisions 890 faculty members 547 sessional instructors 892 specialist/instructional staff
◆	University of Washington	More than 92,000 students annually	16 colleges and schools offering 1,800 undergraduate courses each quarter
○	University of Western Ontario	27,525 full-time enrolments across undergraduate and graduate students	13 Academic Faculties 1,408 full-time faculty members
★	University of Windsor	14,058 full and part time undergraduate students 1,637 full and part time graduate students	794 distinct undergraduate programs 524 faculty members At least 9 schools and Academic Faculties
★	York University	Approximately 55,000 students	7,000 faculty members and staff 11 Academic Faculties and 28 research centres

4 ONLINE VS PAPER BASED INSTRUMENTS

Institutions interviewed are using online or paper based instruments or a combination of both. Of the institutions using online instruments, 61% have been developed in-house. Two vendor products being used are: CourseEval (ConnectEDU) and Blue (eXplorance). Table 2 summarizes the types of instruments used by the institutions interviewed.

Table 2: Types of instruments used by institutions

Type	Name	Online	Paper based	Both
★	Brock University			Online: in-house product Paper based: bubble forms
◆	Griffith University	Developed in-house		
○	Harvard University	Developed in-house		
◆	Kwantlen Polytechnic University			Online: in-house product Paper based: bubble forms
○	McGill University	Developed in-house		
◆	McMaster University		Bubble forms	
★	Memorial University of Newfoundland			Online: in-house product Paper based: bubble forms
★	Ryerson University			Online: in-house product Paper based: bubble forms
◆	Saint Mary's University		Bubble forms	
○	Stanford University	Developed in-house		
◆	University of Alberta			Online: CourseEval by ConnectEDU Paper based: bubble forms
□	University of British Columbia	CourseEval by ConnectEDU		
○	University of Cambridge			Online: in-house product Paper based: bubble forms
◆	University of Louisville			Online: Blue by eXplorance Paper based: bubble forms
○	University of Michigan	Developed in-house		
◆	University of Pennsylvania	Blue by eXplorance		
◆	University of Saskatchewan		Bubble forms	
○	University of Toronto	Blue by eXplorance		
□★	University of Victoria			Online: in-house product Paper based: bubble forms

Type	Name	Online	Paper based	Both
◆	University of Washington			Online: in-house product Paper based: bubble forms
○	University of Western Ontario		Bubble forms	
★	University of Windsor		Bubble forms	
★	York University	Developed in-house		
23 institutions		39%	22%	39%

5 USE OF DATA

5.1 SHARING EVALUATION DATA














There were essentially three methods reported with respect to how institutions interviewed share evaluation data:

1. Data is available to the entire institution community: this usually means that evaluation results are made available to students and the rest of the institution community (i.e. those with an ID and password) – in each case, not ALL data is made available
2. Data is delivered to instructors, department heads and administrators only
3. Data is delivered to instructors only (allowing them to decide who to share the results with).

Table 3 provides a summary of how the institutions interviewed share evaluation data with their respective communities.

Table 3: How institutions share evaluation data with the institution community

Type	Name	Entire institution community	Instructors, Department Heads and Administrators Only	Instructors Only
★	Brock University			✓
◆	Griffith University		✓	
○	Harvard University		✓	
◆	Kwantlen Polytechnic University		✓	
○	McGill University	✓		
◆	McMaster University		✓	
★	Memorial University of Newfoundland	✓		
★	Ryerson University	✓		
◆	Saint Mary's University			✓
○	Stanford University	✓		
◆	University of Alberta	✓		

Type	Name	Entire institution community	Instructors, Department Heads and Administrators Only	Instructors Only
	University of British Columbia	✓		
	University of Cambridge	✓		
	University of Louisville		✓	
	University of Michigan	✓		
	University of Pennsylvania	✓		
	University of Saskatchewan		✓	
	University of Toronto	✓		
 	University of Victoria		✓	
	University of Washington		✓	
	University of Western Ontario	✓		
	University of Windsor	✓		
	York University	✓		
23 institutions		57%	35%	8%

5.2 STUDENT ACCESS TO EVALUATION RESULTS

All institutions interviewed indicated they feel it is important for students to have access to evaluation results – even if only fragments of aggregated data. Institutions that are not currently sharing results with students are actively pursuing ways that they can by trying to overcome policies, legislation and union bylaws.

The strongest, most consistent reason stated for wanting to share data with students is because it is viewed as effective in demonstrating that students concerns have been heard – an indication that student feedback does make a difference. This form of engagement with students was generally seen as critical, regardless of whether an institution felt they are doing a good job of it or not.

5.3 FORMATIVE VERSUS SUMMATIVE USE OF THE DATA

Formative and summative use of data at institutions interviewed is summarized in Table 4. About two thirds of the institutions interviewed reported using evaluation data for both summative and formative purposes, whereas about a third use it for summative purposes only.

Of those institutions using the evaluation results for summative purposes only, many stated that they would like to see a more formative approach taken. In some cases, using the data in a formative way is left up to departments and individual instructors to interpret. Many institutions suggested that the evaluation data was available too late for instructors to make any changes (e.g. at the end of the year or semester).

In order to use the evaluation data for both formative and summative purposes, institutions indicated asking the right questions, assisting with data interpretation and offering support to instructors were key factors to realizing the best outcome.

Suggestions for types of support included:

- Starting conversations with faculty members that appear to be struggling
- Offering teaching support services
- Creating professional development opportunities
- Interpreting data for use in teaching dossiers and assignment of teaching awards.

TABLE 4: Formative and summative use of data at institutions interviewed

Type	Name	Both Formative & Summative	Summative Only
★	Brock University		✓
◆	Griffith University		✓
○	Harvard University	✓	
◆	Kwantlen Polytechnic University	✓	
○	McGill University	✓	
◆	McMaster University		✓
★	Memorial University of Newfoundland	✓	
★	Ryerson University	✓	
◆	Saint Mary's University	✓	
○	Stanford University	✓	
◆	University of Alberta		✓
□	University of British Columbia		✓
○	University of Cambridge	✓	
◆	University of Louisville	✓	
○	University of Michigan	✓	
◆	University of Pennsylvania	✓	
◆	University of Saskatchewan		✓
○	University of Toronto	✓	
□★	University of Victoria	✓	
◆	University of Washington	✓	
○	University of Western Ontario		✓

Type	Name	Both Formative & Summative	Summative Only
★	University of Windsor		✓
★	York University	✓	
23 institutions		65%	35%

6 RESPONSE RATES

According to the institutions interviewed, offering students incentives such as prizes or cash to complete evaluations did not seem to have much impact on the response rate. One incentive that does appear to work for a couple of schools is granting students access to their grades earlier if they complete an evaluation.

Switching from using paper forms to an online instrument seems to have had an initial negative impact on the student response rate at some institutions. Leaving it up to each student to voluntarily choose to participate has proven that many choose not to follow through and complete the evaluation form. Paper based evaluation systems often include a process whereby students complete the form while in class. This generally appears to help encourage a higher response rate.

Some institutions have worked for some time to improve their online response rate and report these rates are now within the same range as they had previously or higher. One institution offered the advice: *“It’s less important to focus on the response rate...focus more on whether or not you’re getting data that’s reliable.”*

A few tips were offered regarding how to encourage students to participate:

- Good communication with students (and internal stakeholders) about why the evaluation is done and why it is important
- Buy-in and participation from faculty: instructors can play a key role in encouraging students to complete the evaluation, including providing class time to do so
- Use of an instrument that is easy to access for students (e.g. on a mobile device or laptop)
- Prompting students to complete the evaluation form when they go online to check their grades
- Using questions that are easy to understand
- Sharing the evaluation results with students.

7 BEST PRACTICES

Several institutions interviewed highlighted the positive impact that student evaluation of teaching and courses can have. One institution offered the comment, *“It’s a culture shift, for sure.”* To maximize the benefits to SFU, any future changes to the evaluation framework should, in part, be based on techniques that have been proven through use and refinement at other institutions.

The following is a brief summary of key best practices based on findings, observations and advice captured during the consultation with colleagues at other institutions.

7.1 DEFINE GOVERNING PRINCIPLES / POLICIES

A clear best practice for effective student evaluation of teaching and courses is to define a set of principles that outline the desired student learning experience at the institution. Such a definition and the policies put in place to align with these principles should form the foundation for the evaluation system. As an interviewee at one institution put it, *“We’ve found that people get really hung up on the questions, but really it’s more about what they want to know.”*

7.2 ADOPT A FLEXIBLE FRAMEWORK

Having a defined framework with guidelines and standard procedures helps promote a consistent approach. It is important to adopt a framework that is flexible enough to meet the needs of stakeholders at various levels across the institution. Such flexibility can enable the institution to focus on priorities that are specific to the entire institution as well as the priorities of each individual school, faculty, department and instructor.

Best practices with respect to flexibility include:

- Allow each key stakeholder group to include their own questions in the evaluation
- Provide opportunities for stakeholders to create/tailor questions specific to their priorities
- Create a bank of questions that can be used by the institution, faculties, schools, departments and instructors.

7.3 NEED FOR STRUCTURED SUPPORT

It was strongly suggested by the institutions interviewed that structured support from a designated department in the institution is required for the development of evaluation questions and interpretation of evaluation data.

Such a department can help instructors select or create evaluation questions that will elicit the response data they need to support their personalized plan for achieving specific teaching goals and standards. Based on an instructor's adaptations, further analysis of response data can indicate how the changes impacted the learning experience. The department's team of professionals can also take the lead in providing analysis and helping interpret evaluation data for faculty members and other stakeholders.

Interviewees suggested that structuring support in this way would help ensure that the necessary assistance is available to those who need it. Suggestions for types of support offered include providing workshops, customized reports, data interpretation tools, teaching specialists, classroom observation and assistance creating faculty dossiers for tenure or promotion.

One institution illustrated it this way: *"It's a lot like using a bathroom scale. The scale doesn't prescribe what you need to do, it just tells you that you need to take action...like go to the doctor, go to the gym, take tests and eliminate stress. From there you need to make a plan."*

7.4 TAKE A CAREFUL, THOUGHTFUL APPROACH

A general best practice embedded in much of the advice offered by interviewees can be summarized as taking a careful, thoughtful approach. Advice offered included:

- Do your research and get institution-wide buy-in
- Take time to develop a comprehensive evaluation framework that has been validated by experiential evidence
- Build everything else on key foundational elements – evaluation processes, questions, technological support, etc
- Discover what the teaching and learning priorities are across the institution by focusing on each stakeholder's vision of an ideal student learning experience – use this to identify ways to enhance the evaluation framework.

The institutions interviewed reported that taking this approach can prove to be very rewarding. As one interviewee put it, *"The most interesting and enriching part of the process has been the discussions at each division about their priorities. The conversations about their teaching has reached levels that we just didn't expect."*

8 FURTHER INVESTIGATION AS PART OF THE PROOF OF CONCEPT

The interviews with other institutions provided the project team with an abundance of good information and experience to draw upon. However, while these interviews were enormously helpful, it did raise questions regarding the most appropriate practices to recommend for adoption at SFU.

Questions were raised in the following key areas:

- Have we identified a suitably flexible framework that can be adapted for our institution?
- What sort of support will be required to make the new system work?
- How should evaluation reports be managed?
- What are the implications for online versus paper based evaluations?
- How should the new evaluation process be managed?

The team determined that the Proof of Concept phase of the project could provide valuable practical insights into these questions. Accordingly, the questions were incorporated into that phase of the project. Please see Appendix X and XI for more information on the Proof of Concept phase of the project.



SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

TEACHING AND COURSE EVALUATION PROJECT FINAL REPORT

APPENDIX V: COMMUNICATIONS PLAN

SPRING 2012 - 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

TABLE OF CONTENTS

1	COMMUNICATIONS STRATEGY	1
1.1	PROJECT DRIVERS.....	1
1.2	CRITICAL COMMUNICATION ISSUES	1
1.2.1	FOR THOSE WHO ARE EVALUATED (FACULTY MEMBERS AND INSTRUCTORS)	1
1.2.2	FOR THOSE WHO EVALUATE (STUDENTS)	1
1.2.3	FOR THOSE WHO USE THE EVALUATION DATA (ADMINISTRATORS)	1
1.2.4	FOR THOSE WHO COORDINATE THE EVALUATION PROCESS (DEPARTMENTAL MANAGERS)	1
1.3	DESIRED OUTCOMES	1
1.3.1	SHORT-TERM (CONSULTATION PROCESS).....	1
1.4	LONG-TERM (DURING AND AFTER IMPLEMENTATION OF NEW EVALUATION INSTRUMENT, PROCESS, AND GUIDELINES)	2
1.5	COMMUNICATIONS STRATEGY FOR CONSULTATION PROCESS	2
1.5.1	STAGE 1 – CONSULTATIONS ABOUT PROCESS (APRIL–SEPTEMBER 2012).....	2
1.5.2	STAGE 2 – CONSULTATIONS ABOUT CONTENT (SEPTEMBER 2012 AND BEYOND)..	2
1.6	KEY MESSAGES.....	2
1.7	TOOLS	2
2	COMMUNICATION AND INFORMATION RESOURCES FOR SUMMER 2013 PROOF OF CONCEPT.....	3
2.1	ABOUT THIS PROJECT (1–2 PAGES)	3
2.2	QUESTION SELECTION AND CUSTOMIZATION FORM (1-PAGE INTRO PLUS QUESTIONS)	3
2.3	HOW TO USE THE RESULTS	3
2.4	SCRIPTS	3
2.5	STUDENT Q&A (1–2 PAGES)	3
2.6	DELIVERABLES FOR WEBSITE (BY JULY 2).....	3

1 COMMUNICATIONS STRATEGY

1.1 PROJECT DRIVERS

- Recommendations of SCUTL and TFTL
- Desire to take advantage of advances in TCE research, methods, and instruments
- Desire for more consistent and appropriate collection and use of data
- Desire for more efficient and appropriate methods for gathering, compiling, interpreting, and using data

1.2 CRITICAL COMMUNICATION ISSUES

1.2.1 FOR THOSE WHO ARE EVALUATED (FACULTY MEMBERS AND INSTRUCTORS)

- Concerns about fairness, consistency, and relevance of test formats
- Concerns about inappropriate interpretation and use of data
- Concerns about privacy, academic freedom, and related issues

1.2.2 FOR THOSE WHO EVALUATE (STUDENTS)

- Perception that data will be disregarded
- Desire for positive change based on evaluations
- Desire for publication of data results

1.2.3 FOR THOSE WHO USE THE EVALUATION DATA (ADMINISTRATORS)

- Desire for consistent, flexible data collection
- Desire for credible, relevant, meaningful, and easily usable data
- Desire for best practices in collection and use of data

1.2.4 FOR THOSE WHO COORDINATE THE EVALUATION PROCESS (DEPARTMENTAL MANAGERS)

- Desire for more consistent and efficient methods for gathering, compiling, storing, and accessing data

1.3 DESIRED OUTCOMES

1.3.1 SHORT-TERM (CONSULTATION PROCESS)

- **Faculty members and instructors:** Establish a high level of awareness of the TCE project; establish an understanding that the project is motivated by a desire to provide them with positive benefits; establish a perception that their needs and concerns will be addressed; achieve a high rate of involvement in consultation process
- **Students:** Establish a high level of awareness of the TCE project, linked with a perception that student voices will be heard during the consultation process; achieve a high rate of involvement in consultation process
- **Administrators:** Establish a high level of awareness of the TCE project; establish a high level of awareness of the potential benefits of the project; establish a high level of official support for the project
- **Departmental managers:** Establish a high level of awareness of the TCE project; establish a high level of awareness of the potential benefits of the project; establish a high level of participation in the consultation process

1.4 LONG-TERM (DURING AND AFTER IMPLEMENTATION OF NEW EVALUATION INSTRUMENT, PROCESS, AND GUIDELINES)

- **Faculty members and instructors:** Establish a positive attitude toward the new instrument, process, and guidelines; establish a high level of awareness of the benefits of the new instrument, process, and guidelines
- **Students:** Establish a high level of awareness of the uses of evaluation data; establish a high level of confidence that evaluation data will prompt meaningful, positive change
- **Administrators:** Establish a high level of belief in the benefits of the new instrument, process, and guidelines; establish a high level of adoption for the new instrument, process, and guidelines
- **Departmental managers:** Establish a high level of support for the new instrument, process, and guidelines

1.5 COMMUNICATIONS STRATEGY FOR CONSULTATION PROCESS

1.5.1 STAGE 1 – CONSULTATIONS ABOUT PROCESS (APRIL–SEPTEMBER 2012)

- i. Build understanding of and support for the TCE project among administrators and departmental managers through digital media, individual meetings, and (in the case of departmental managers) the creation of an advisory council to address the TCE process (excluding questions of content)
- ii. Build awareness of and support for the TCE project among faculty, instructor, and student groups through individual meetings and digital media
- iii. Build awareness of the TCE project and alleviate concerns and fears among faculty members and students through digital media

1.5.2 STAGE 2 – CONSULTATIONS ABOUT CONTENT (SEPTEMBER 2012 AND BEYOND)

- i. Build understanding of and support for the TCE project among faculty members and students through digital media, surveys, open forums, and individual meetings
- ii. Build an awareness of the benefits and improvements offered to all stakeholders by highlighting capabilities of new options through digital media and pilot projects

1.6 KEY MESSAGES

- A vast body of research and literature on TCE exists, and many advanced TCE models exist at other institutions; the SFU project will be built upon this solid foundation
- The new evaluation instrument, process, and guidelines will be consistent, flexible, and customizable by course format, discipline, etc., to allow for relevant questions and meaningful comparisons
- The new data collection process will be efficient and designed to encourage high rates of student response and to facilitate more extensive and more meaningful use of results
- The new process will include best-practice guidelines for collection and use of data
- Instructors' concerns about privacy, academic freedom, and related issues will be explored
- Students' concerns about the process and accessibility of the results will be explored

1.7 TOOLS

- Website (completed)
- News releases to SFU News, The Peak, and other SFU-related media
- Face-to-face meetings with key individuals in stakeholder groups
- Advisory council meetings with departmental managers
- Project e-newsletter
- Feedback tools in consultations and meetings (paper forms, online polls, etc.)
- Advertorial features in The Peak

2 COMMUNICATION AND INFORMATION RESOURCES FOR SUMMER 2013 PROOF OF CONCEPT

2.1 ABOUT THIS PROJECT (1–2 PAGES)

- Intro – Context, purpose, objective of TCE project and summer PoC
- Timeline (key dates and activities for instructors, students, project team)
- Q&A – How PoC results will be used, how instructor feedback will be collected, etc.
- Contact info for further inquiries
- Pointers to background documentation (website)

2.2 QUESTION SELECTION AND CUSTOMIZATION FORM (1-PAGE INTRO PLUS QUESTIONS)

- Explanation of four-tier structure and purpose of questions at each level
- Description of how results will be used; address confidentiality
- Description of factors to consider in choosing questions
- Table of questions at all levels
- We will be using all 183 questions from the U of T question bank

2.3 HOW TO USE THE RESULTS

- Explanation of how to read report; include ideas for using results etc.

2.4 SCRIPTS

- Intro – Instructor role in boosting participation
- Overview of survey process with script dates indicated
- Sample scripts for use in encouraging student participation

2.5 STUDENT Q&A (1–2 PAGES)

- Project background
- Purpose of proof of concept
- How test will be conducted
- How results will be used
- How feedback will be solicited
- How they will be informed of outcomes (both question results and project results)
- Contact info for further inquiries (website, Sarah)

2.6 DELIVERABLES FOR WEBSITE (BY JULY 2)

- Post U of T background report
- Repackage proof-of-concept information package for website
 - Intro
 - Timeline
 - Instructor Q&A
 - Student Q&A
 - Pointers to U of T, eXplorance, etc.



SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING & COURSES

TEACHING AND COURSE EVALUATION PROJECT FINAL REPORT

APPENDIX VI: FACULTY ADMIN SURVEY KEY FINDINGS SUMMARY

NOVEMBER 7, 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

DOCUMENT CONTROL

DOCUMENT INFORMATION

	INFORMATION
Document Owner	<i>Corinne Pitre-Hayes</i>
Project /Organization Role	<i>Project Manager, Teaching & Course Evaluation Project</i>

DOCUMENT HISTORY

VERSION	ISSUE DATE	CHANGES
<i>0.1</i>	<i>April 10, 2013</i>	<i>Initial draft</i>
<i>0.2</i>	<i>October 2, 2013</i>	<i>Additions to draft</i>
<i>0.3</i>	<i>October 11, 2013</i>	<i>Revisions per C. Pitre-Hayes</i>
<i>0.4</i>	<i>October 17, 2013</i>	<i>Further revisions</i>
<i>0.5</i>	<i>November 2, 2013</i>	<i>Revisions for inclusion in appendix of final project report</i>
<i>0.6</i>	<i>November 7, 2013</i>	<i>Revisions from final team review</i>

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	1
2	INTRODUCTION	1
3	SURVEY OVERVIEW	1
4	TYPES OF INSTRUMENTS USED	3
5	KEY THEMES	5
	5.1 THE PROCESS.....	5
	5.2 FACULTY SUPPORT	5
	5.3 STUDENT ENGAGEMENT.....	6
	5.4 USE AND INTERPRETATION OF EVALUATION DATA.....	6
6	CONCLUSION.....	6

1 EXECUTIVE SUMMARY

This report provides an analysis and summary of phone interviews conducted with a variety of staff members involved in the administration of student evaluation of teaching and courses. Interviews were conducted with managers, advisors, secretaries, assistants and directors.

Key themes that emerged from discussions with 33 individuals across the institution involved in teaching and course evaluation processes include:

- Significant challenges with the existing largely paper-based processes
- Opportunities for better faculty and student engagement
- Concern regarding use and interpretation of evaluation results.

Almost everyone interviewed agreed that an online approach would help create efficiencies. In general, the top priorities for a new system are reducing time and cost while increasing interactivity with students and faculty. Overwhelmingly, all academic units interviewed would like to see a redesigned evaluation process that helps create a consistent approach to improving teaching and learning through effective course delivery.

2 INTRODUCTION

In 2008, the Senate Committee on University Teaching and Learning (SCUTL) conducted a survey of how academic units use evaluations. The results of this survey can be found in the document "Evaluating How We Evaluate: Examining SFU's course and instructor evolution system" (Summer 2008). Building on this earlier work, our engagement with the SFU community included interviews with staff involved with the administration of student evaluation of teaching and courses. The objectives of these interviews were to:

- Update our information on the types of instruments currently in use at SFU
- Gain a deeper understanding of the current student evaluation business practices across the institution
- Obtain insight into the administrative readiness of the SFU community to embrace a new approach to student evaluation of teaching and courses.

The goal was to use this information to help inform our recommendations to Senate for instrument selection and related processes. Understanding what our faculties, schools and departments are doing today provides meaningful input for instrument selection and support. This report provides a summary of key findings from this stage of engagement with the SFU community.

3 SURVEY OVERVIEW

The team conducted interviews with 33 individuals in the following areas:

- Beedie School of Business (1 person)
- Faculty of Applied Sciences (4 people)
- Faculty of Arts and Social Sciences (10 people)
- Faculty of Communication, Art and Technology (6 people)
- Faculty of Education (1 person)
- Faculty of Environment (2 people)
- Faculty of Health Sciences (2 people)
- Faculty of Science (6 people)
- Centre for Online and Distance Education (1 person).

Table 1 below provides a summary of the faculties represented, the individuals interviewed and the individuals' respective roles in the organization.

Table 1: Faculties represented and the roles of the individuals interviewed

Faculty	Role
Beedie School of Business	Academic Director, Undergraduate Program
Faculty of Applied Sciences	Manager, Administrative and Academic Services Undergraduate Program Assistant Assistant Manager, Academic and Administrative Services Manager, Academic and Administrative Services
Faculty of Arts and Social Sciences	Department Assistant Manager, Academic and Administrative Services Manager, Academic and Administrative Services Chair's Secretary Manager, Academic and Administrative Services Undergrad Secretary Manager, Academic and Administrative Services & Undergraduate Advisor Manager, Academic and Administrative Services Manager, Academic and Administrative Services Manager, Academic and Administrative Services
Faculty of Communication, Art and Technology	Undergraduate Advisor Manager, Academic and Administrative Services Director of Administration Administrative Assistant Undergraduate Advisor School Secretary
Faculty of Education	Director, Special Projects
Faculty of Environment	Manager, Academic and Administrative Services Office Assistant
Faculty of Health Sciences	Director Director, Administration and Operations
Faculty of Science	Manager, Administrative Services Manager, Academic and Administrative Services Manager, Administrative Services Manager, Academic and Administrative Services Manager, Administrative Services & Student Academic Advisor Department Manager
Centre for Online and Distance Education	Program Director

A set of standard interview questions were organized into the following topic areas to help structure the conversations:

- Mechanics of the evaluation process
- Staff involvement and decision-making
- How data is used or stored
- Challenges of the existing process
- Recommendations or suggestions for improvement

4 TYPES OF INSTRUMENTS USED

As shown in Table 2, 84% (of 25 Faculties/schools/departments) interviewed reported that they currently use paper-based instruments. Paper-based included SFU's standard bubble form, in-house forms developed by the academic unit, or a combination of both. Those reporting using only the standard SFU bubble form represented about 48% of the total. Approximately 8% reported using online forms and another 8% a combined approach by pairing the use of an online instrument with paper forms. Of the few using online evaluations one academic unit mentioned their solution was developed in-house, and one indicated using an online tool.

Table 2: Instruments Reported Used

Faculty School / Department	Online	Paper-based Bubble In-House Combo	Both Online & Paper-Based
Beedie School of Business		Bubble	
Centre for Online and Distance Education	In-house		
Faculty of Applied Sciences			
School of Computing Science		Bubble	
School of Engineering Science		Bubble	
Faculty of Arts and Social Sciences			
Department of Economics		Bubble	
Department of English		Bubble	
Department of French		Bubble	
Department of History		Bubble	
Department of Humanities, Asia-Canada Program, Language Training Institute		Bubble	
Department of Philosophy		In-house	
Department of Psychology		Bubble	
School of Criminology		In-house	

Faculty School / Department	Online	Paper-based		Both Online & Paper-Based
		Bubble	In-House	
Faculty of Communication, Art and Technology				
School for Contemporary Arts				Combo
School of Communication				Combo
School of Interactive Arts and Technology				Combo
Faculty of Education				
Faculty of Education			In-house	
Faculty of Environment				
Department of Geography		Bubble		
Department of Archaeology		Bubble		
Faculty of Health Sciences				
Faculty of Health Sciences	Online tool			
Faculty of Science				
Department of Earth Sciences		Bubble		
Department of Biological Sciences		Bubble		
Department of Biomedical Physiology and Kinesiology		Bubble		
Department of Chemistry				Web & Bubble
Department of Mathematics				Bubble & Online
Department of Statistics and Actuarial Science				Combo
25 Faculties / Schools / Departments		8%	84%	8%

The SCUTL survey conducted in 2008 (approximately 50 interviewed or surveyed participants) reported that:

- 66% of the academic units used the standard SFU bubble form
- 23% used a combination of the standard forms and in-house forms
- 11% used in-house forms only

The updated information indicates (from 33 individuals representing 25 Faculties/schools/departments):

- 48% are currently using the standard SFU bubble form
- 16% are using a combination of bubble form and in-house forms
- 20% are using in-house only

While it is important to note that the data is not directly comparable, there is a suggestion that academic units may be increasingly "doing their own thing".

5 KEY THEMES

At the outset of conducting these administrative interviews, the goal of the project team was to update and expand on the information collected during the SCUTL survey completed in 2008. This further consultation identified an overwhelming number of similarities across the institution, specifically in regards to how the evaluation process works and the challenges associated with it. It also identified a small number of key differences.

The similarities were quite marked and included:

- The process is viewed as antiquated and time-consuming
- Questions are deemed inflexible and many suggest there is little support for effective analysis of results
- There was frequent mention of a lack of support offered to instructors, especially related to improving teaching and learning
- Many report a profound lack of engagement of students with the evaluation process
- Results are consistently reported to be used for tenure, promotion and instructor evaluation

Key differences included:

- A wide range of policies and processes reported for managing, interpreting and storing evaluation data
- Different academic units had different staff / roles involved in the evaluation process
- There were inconsistencies reported in how results were being used by instructors / departments to improve course delivery

The following sections provide additional detail for some of the most significant themes that emerged in our interviews with administrative community members across the institution.

5.1 THE PROCESS

The most common theme across all interviews was that the current process is outdated, inflexible and very manual – “an administrative burden.” Typically, most areas reported using either the standard SFU bubble form or one that has been developed in-house. The time and cost involved in ordering and distributing the forms in addition to compiling and summarizing the results is immense.

A few specific administrative items mentioned that could use some work included:

- Many report a discrepancy in how the written comments are handled – in some cases a staff member or temp is assigned to type the notes, while in other departments instructors are allowed to review the student’s handwritten comments directly
- Graduate evaluations are described as challenging due to small class sizes and the issue of anonymity
- Room for improvement was noted in the area of helping the junior staff who administer the evaluations. It was suggested that more effective communication about the importance of the course evaluation and its data could help encourage junior staff to feel more involved in the process.

Faculties that are already using online forms generally agreed that technical adjustments and improvements to questions would enhance their evaluation experience.

5.2 FACULTY SUPPORT

Currently, evaluation results are generally reported to be used by departments for tenure, promotion and faculty evaluation activities. All interviewees reported that instructors have the option to see their results. However, in many cases, a great deal of effort is required to make this happen. Many report that the logistics and timing of the current evaluation process makes it difficult to give meaningful feedback to instructors before the next iteration of a course.

It was noted that greater efficiency could be achieved if instructors could have evaluation results made available to them in a more direct and consistent way. The importance of receiving the results in a format

that's easy to understand was also emphasized. It was suggested that this could help gain faculty member buy-in to the importance of the evaluation process as it relates to teaching and course delivery.

Supporting instructors in developing their skills and offering multiple evaluation methods (i.e. peer-to-peer and observational evaluations) were also mentioned as suggestions for the redesign of the existing evaluation process.

5.3 STUDENT ENGAGEMENT

We did not identify any unit that currently offers students evaluation feedback or the option to view evaluation results. It was noted that students don't appear to think the evaluations are taken seriously – “it doesn't seem to make any difference” is apparently a common answer to survey questions.

It was also suggested that the current evaluation process doesn't allow for much interaction with students. Many indicated that it would be beneficial to provide a more consistent level of service to students. It was noted that it would be better to engage students with an outline of the process and its importance followed by sharing results, rather than expecting students to complete the evaluations "because they're supposed to".

Several were convinced that more student responses would lead to better, more reliable data. It was suggested that it would be ideal if every student submitted an evaluation – even the ones who don't make it to class. The question asked by many was, “how do we get more students to participate?”

5.4 USE AND INTERPRETATION OF EVALUATION DATA

A wide range of approaches were reported in the use and interpretation of evaluation data. One example given was the development of internal algorithms to extract additional comparative data and to make the data more meaningful (e.g. taking the difficulty level of the course into account when comparing across the unit).

There was also a wide range of data retention policies in use across the institution. For example, some schools/departments reported storing all their data indefinitely, while others indicated that they destroy it after five semesters.

In addition, some reported gaps in the evaluation data (e.g. some reported that laboratory experiences are currently not evaluated). There was a widespread perspective that the inability to access detailed comparative results greatly hampered the ability to effect change and offer support to instructors.

Some recommendations for the future included finding a system that is more centralized and electronic – online perhaps. It was suggested that an online system would enable instructors to have the opportunity to more quickly implement changes to their course delivery. Similarly, students would have the option to respond to the survey either in-class or at their leisure. While the benefits of an electronic system were noted, the challenges were also highlighted. The challenges of using an online approach that were mentioned included maintaining the use of open-ended questions, protecting student privacy and the prospect of low response rates. Concerns were expressed that these challenges could overshadow the benefits of an online system.

6 CONCLUSION

Despite reporting that the current processes are cumbersome and ineffective, administrative staff generally expressed an appreciation for the importance of student evaluation of teaching and courses – especially for faculty. They expressed a willingness and readiness to adapt to whatever system is decided upon. There were widespread wishes for a new system that is more streamlined and straightforward than the existing one.



SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

TEACHING AND COURSE EVALUATION PROJECT FINAL REPORT

APPENDIX VII: RESULTS FROM SURVEY OF STUDENTS: TEACHING AND COURSE EVALUATION PROCESS

Prepared by:

Janice Wiebe, Analyst; Jessica Tilley, Analyst; Kiran Bisra, Analyst; Evan Thompson, Co-op Analyst
Institutional Research and Planning, Simon Fraser University

SPRING 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

TABLE OF CONTENTS

1	INTRODUCTION	1
2	METHODOLOGY.....	1
3	RESULTS	1
	3.1 FEEDBACK ON THE CURRENT TEACHING AND COURSE EVALUATION PROCESS	1
	3.2 IMPROVEMENTS TO THE CURRENT PROCESS.....	2
4	SURVEY INSTRUMENT	4
5	DISTRIBUTION OF STUDENT RESPONSES	4

1 INTRODUCTION

At Simon Fraser University (SFU), at the end of each term, students are asked to fill out a questionnaire about each class they take, asking them to evaluate both the course and the instructor(s). In May 2011 the SFU Senate approved a plan to revise the current teaching and course evaluation process. The development of recommendations for a new teaching and course evaluation process has involved community consultation, of which this survey forms a part.

2 METHODOLOGY

Two survey questions about teaching and course evaluations were included in the Fall 2012 Undergraduate Student Survey (UGSS). The UGSS was administered online by SFU's department of Institutional Research and Planning, during October 2012. All 24,521 registered undergraduate students¹ with known e-mail addresses were invited by e-mail to participate in the survey. In order to accommodate more questions, the Fall 2012 survey was split into two surveys, with the population of students randomly split between the two. 12,260 students were invited to participate in the survey that contained the two questions about teaching and course evaluations. A total of 2,785 students responded. This represents a response rate of 23%. Assuming the sample is representative, proportions calculated on all respondents are accurate within $\pm 1.7\%$, 19 times out of 20.

3 RESULTS

The survey instrument and distribution of responses to the survey questions can be found in sections 4 and 5 respectively.

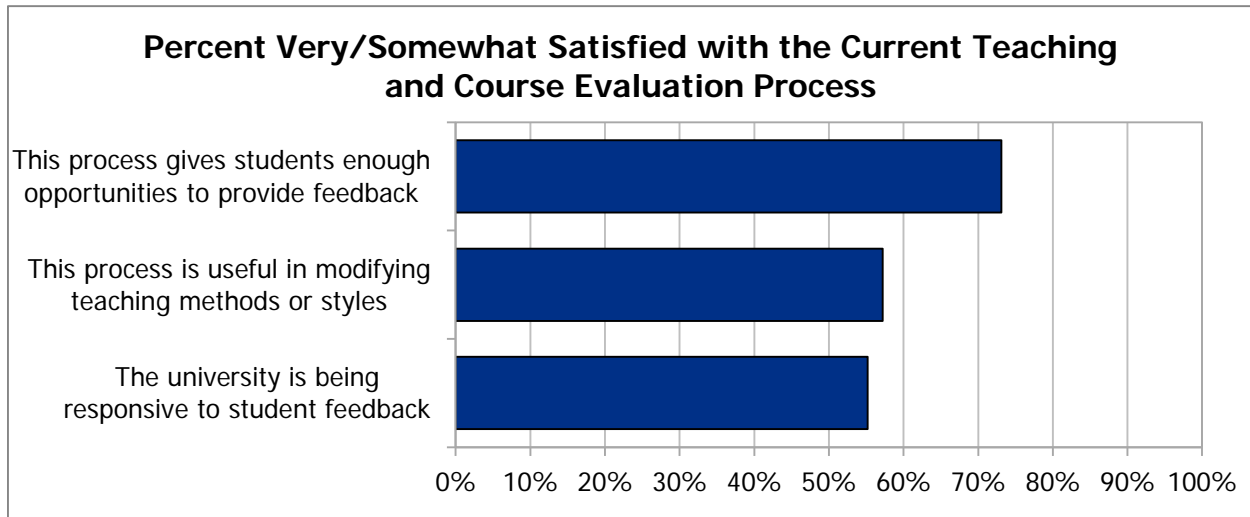
3.1 FEEDBACK ON THE CURRENT TEACHING AND COURSE EVALUATION PROCESS

Students were asked about their satisfaction with the current teaching and course evaluation process (Figure 1). Among respondents:

- 73% are very or somewhat satisfied that the process gives students enough opportunities to provide feedback,
- 57% are very or somewhat satisfied that the process is useful for modifying teaching methods/styles, and
- 55% are very or somewhat satisfied that the university is being responsive to student feedback.

¹ Students who were on co-op terms in Fall 2012 and were not taking any non-co-op courses were not invited to participate in the survey.

Figure 1



3.2 IMPROVEMENTS TO THE CURRENT PROCESS

Students were next asked to rate the level of importance of various potential changes/improvements to the teaching and course evaluation process (Figure 2). Among respondents:

- Over 90% said it is very or somewhat important to:
 - have individual instructors to use the findings to modify their teaching or the course,
 - show that departments are taking the findings seriously, and
 - clearly inform students about what's being done with the findings.
- 85-90% said it's very or somewhat important to:
 - provide training for instructors who score below a certain level, and
 - have students provide input mid-way through the term, so that the instructor has time to make adjustments.
- 79% said it would be very or somewhat important to add a question that asks students whether the course covered what they expected based on the course outline.

Figure 2

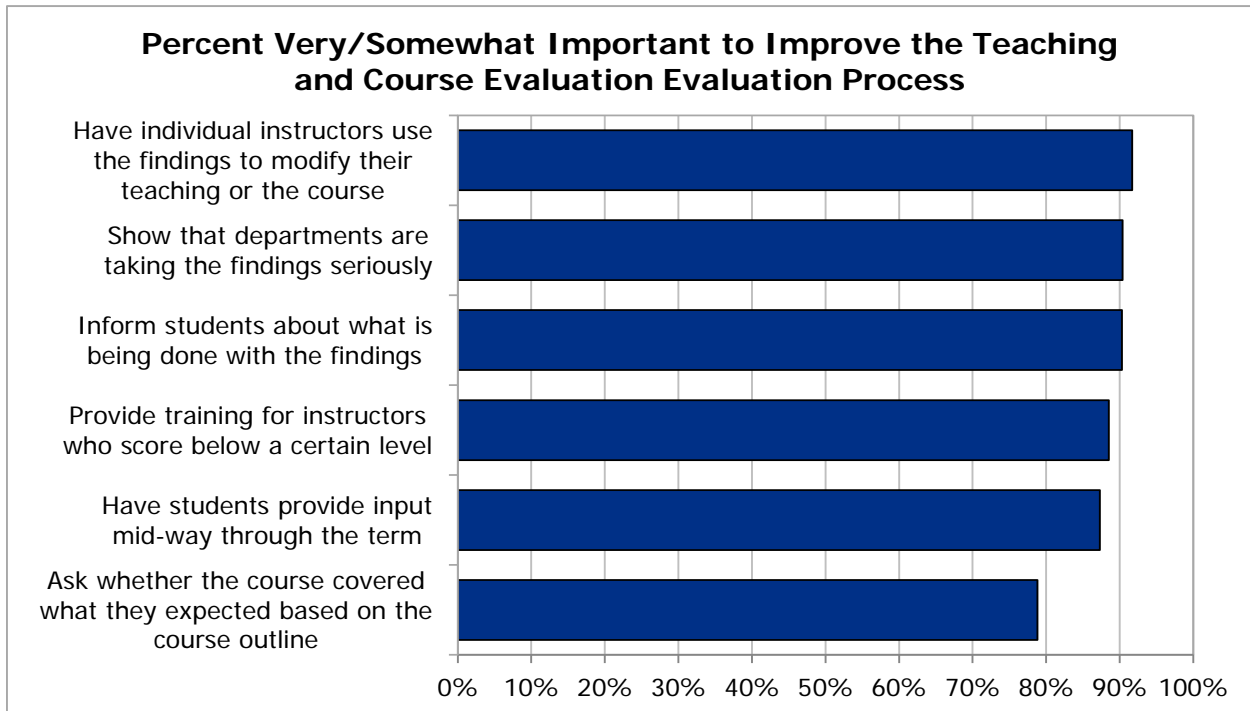
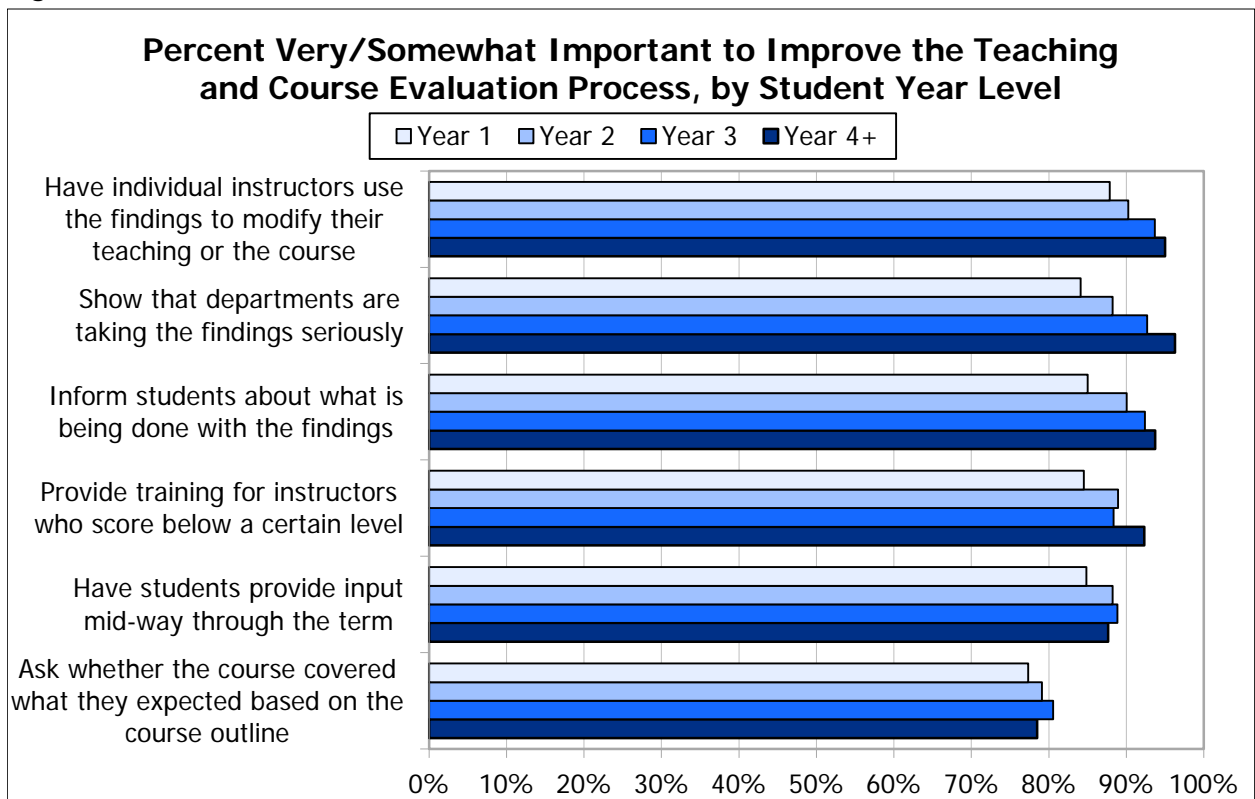


Figure 3 displays the results by student year level. In general, respondents in higher year levels rate these changes/improvements as more important.

Figure 3



4 SURVEY INSTRUMENT

The "Teaching and Course Evaluation Process" refers to the current process whereby at the end of each term, students are asked to fill out a questionnaire about each class they take, asking them to evaluate both the course and the instructor(s).

- 1) The Teaching and Course Evaluation process: How satisfied are you that...
 - a. this process is useful in modifying teaching methods or styles?
 - b. this process give students enough opportunities to provide feedback?
 - c. the university is being responsive to student feedback?

- 2) If it were up to you, what would you do to improve the Teaching and Course Evaluation Process, and to better use the results? Please rate the importance of each of these options.
(Scale: Very Important, Somewhat Important, Not Very Important, Not at all Important, Undecided)
 - a. Have individual instructors use the findings to modify their teaching or the course.
 - b. Show that departments are taking the findings seriously.
 - c. Provide training for instructors who score below a certain level
 - d. Clearly inform students about what's being done with the findings
 - e. Have students provide input mid-way through the term, so the instructor has time to make adjustments.
 - f. Include a question that asks students whether the course covered what they expected based on the course outline.

5 DISTRIBUTION OF STUDENT RESPONSES

This section displays the distribution of student responses to the survey questions. The tables that follow show the number and percentage of students selecting each response to the questions. Please note that percentages in these tables are rounded to one decimal place. Summaries that group categories (e.g. where "Very Satisfied" and "Somewhat Satisfied" are grouped) may result in percentages that do not exactly match the sum of the percentages displayed in the tables. This is an artifact of rounding.

19) The Teaching and Course Evaluation process: How satisfied are you that...

this process is useful in modifying teaching methods or styles?		
Very Satisfied	314	14.4%
Somewhat Satisfied	931	42.8%
Undecided	178	8.2%
Not Very Satisfied	508	23.3%
Not at all Satisfied	245	11.3%
Total Responses	2,176	100.0%
Not Applicable/No Experience	252	
Missing Cases	357	
this process gives students enough opportunities to provide feedback?		
Very Satisfied	545	24.3%
Somewhat Satisfied	1,094	48.8%
Undecided	65	2.9%
Not Very Satisfied	386	17.2%
Not at all Satisfied	151	6.7%
Total Responses	2,241	100.0%
Not Applicable/No Experience	188	
Missing Cases	356	

the university is being responsive to student feedback?		
Very Satisfied	310	14.7%
Somewhat Satisfied	857	40.5%
Undecided	209	9.9%
Not Very Satisfied	488	23.1%
Not at all Satisfied	252	11.9%
Total Responses	2,116	100.0%
Not Applicable/No Experience	310	
Missing Cases	359	

Summary		% Very / Somewhat Satisfied
this process gives students enough opportunities to provide feedback?		73.1%
this process is useful in modifying teaching methods or styles?		57.2%
the university is being responsive to student feedback?		55.2%

20) **If it were up to you, what would you do to improve the Teaching and Course Evaluation Process, and to better use the results? Please rate the importance of each of these options.**

Have individual instructors use the findings to modify their teaching or the course.		
Very Important	1,415	58.8%
Somewhat Important	793	32.9%
Undecided	74	3.1%
Not Very Important	116	4.8%
Not at all Important	9	0.4%
Total Responses	2,407	100.0%
Missing Cases	378	

Show that departments are taking the findings seriously.		
Very Important	1,446	60.0%
Somewhat Important	732	30.4%
Undecided	67	2.8%
Not Very Important	152	6.3%
Not at all Important	13	0.5%
Total Responses	2,410	100.0%
Missing Cases	375	

Provide training for instructors who score below a certain level.		
Very Important	1,403	58.1%
Somewhat Important	734	30.4%
Undecided	89	3.7%
Not Very Important	161	6.7%
Not at all Important	28	1.2%
Total Responses	2,415	100.0%
Missing Cases	370	

Clearly inform students about what's being done with the findings.		
Very Important	1,405	58.3%
Somewhat Important	770	32.0%
Undecided	50	2.1%
Not Very Important	163	6.8%
Not at all Important	21	0.9%
Total Responses	2,409	100.0%
Missing Cases	376	

Have students provide input mid-way through the term, so that instructor has time to make adjustments.		
Very Important	1,324	54.8%
Somewhat Important	786	32.5%
Undecided	75	3.1%
Not Very Important	193	8.0%
Not at all Important	38	1.6%
Total Responses	2,416	100.0%
Missing Cases	369	

Include a question that asks students whether the course covered what they expected based on the course outline.		
Very Important	916	38.1%
Somewhat Important	981	40.8%
Undecided	77	3.2%
Not Very Important	366	15.2%
Not at all Important	66	2.7%
Total Responses	2,406	100.0%
Missing Cases	379	

Summary		% Very / Somewhat Important
Have individual instructors use the findings to modify their teaching or the course.		91.7%
Show that departments are taking the findings seriously.		90.4%
Clearly inform students about what's being done with the findings.		90.3%
Provide training for instructors who score below a certain level.		88.5%
Have students provide input mid-way through the term, so that instructor has time to make adjustments.		87.3%
Include a question that asks students whether the course covered what they expected based on the course outline.		78.8%



SFU

SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

TEACHING AND COURSE EVALUATION PROJECT FINAL REPORT

APPENDIX VIII: RESULTS FROM SURVEY OF FACULTY AND INSTRUCTORS: TEACHING AND COURSE EVALUATION PROCESS

Prepared by:

Janice Wiebe, Analyst; Joanne Quan, Analyst; Jessica Tilley, Analyst
Institutional Research and Planning, Simon Fraser University

MAY 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

TABLE OF CONTENTS

1	INTRODUCTION	1
2	METHODOLOGY.....	1
3	RESULTS	1
	3.1 FEEDBACK ON THE CURRENT TEACHING AND COURSE EVALUATION PROCESS	1
	3.2 FEEDBACK ON HOW THESE EVALUATION RESULTS SHOULD BE USED	2
	3.3 IMPROVEMENTS TO THE CURRENT PROCESS.....	3
4	SURVEY INSTRUMENT	8
5	DISTRIBUTION OF FACULTY AND INSTRUCTOR RESPONSES	9

1 INTRODUCTION

At Simon Fraser University (SFU), at the end of each term, students are asked to fill out a questionnaire about each class they take, asking them to evaluate both the course and the instructor(s). In May 2011 the SFU Senate approved a plan to revise the current teaching and course evaluation process. The development of recommendations for a new teaching and course evaluation process has involved community consultation, of which this survey forms a part.

2 METHODOLOGY

The survey, which was administered by SFU's department of Institutional Research and Planning during February and March, 2013, was sent to all active continuing faculty and all instructors who taught at least one course at SFU in 2012. The purpose of the survey was to seek faculty and instructor feedback on the current evaluation process, as well as their opinions on how the process could be improved, and how the data from teaching and course evaluations should be used.

Of the 1,429 faculty and instructors invited to participate in this survey, 519 responded. This represents a response rate of 36%. Assuming that the sample is representative, proportions calculated on all respondents are accurate within $\pm 3.5\%$, 19 times out of 20.

3 RESULTS

The survey instrument and distribution of responses to the survey questions can be found in sections 4 and 5 respectively.

3.1 FEEDBACK ON THE CURRENT TEACHING AND COURSE EVALUATION PROCESS

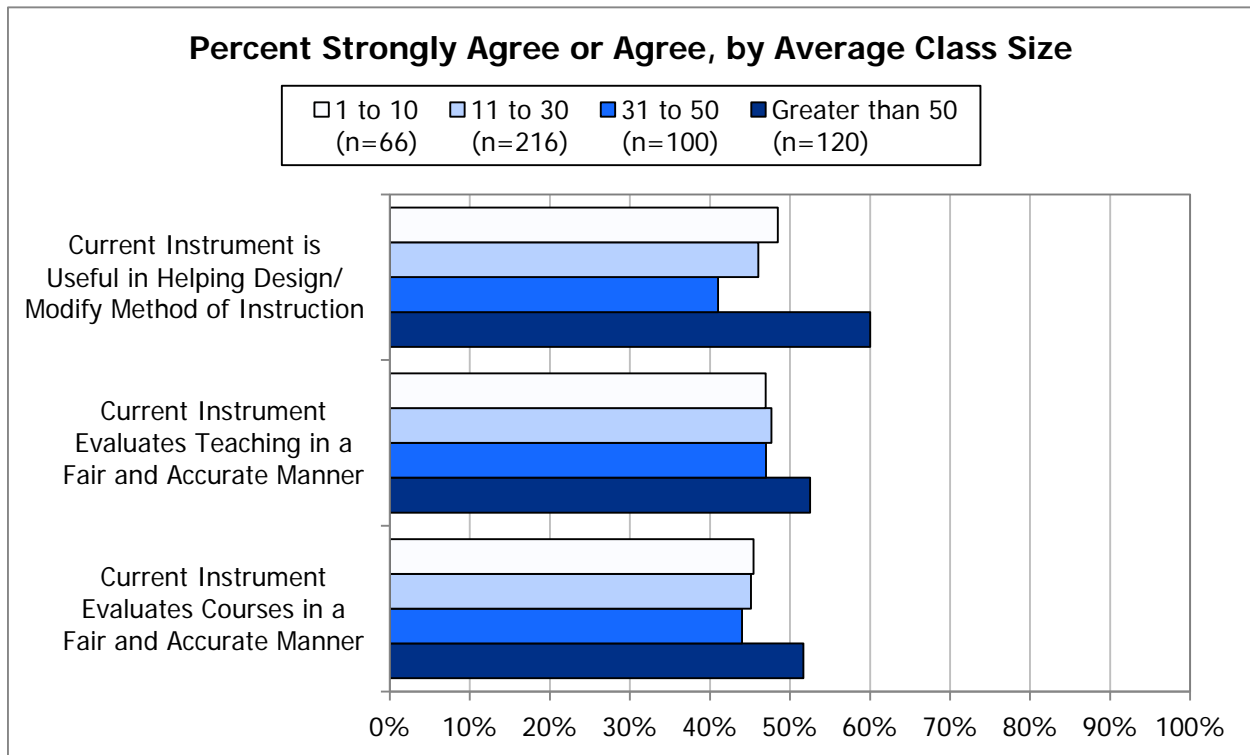
Faculty members and instructors were first asked about their agreement with three statements about the current teaching and course evaluation process. Among all respondents:

- 47-49% agree that the instrument asks questions that evaluate teaching and courses in a fair and accurate manner and that the evaluation results are useful for designing/modifying their method of instruction,
- 43-45% disagree with those statements, and
- 8-9% are undecided.

Figure 1 displays the percentage of respondents who indicated that they strongly agree or agree with the statements, broken down by the average class size the instructor taught in recent terms. From the figure:

- Those teaching large classes (with an average of over 50 students) were most likely to agree with all the statements (52-60% agree).
- Respondents teaching courses with an average of 31-50 students were least likely to agree that the results help them modify their method of instruction (41% agree).

Figure 1



Notes: These averages are based on the four most recent courses taught by each faculty member or instructor within the Spring, Summer, and Fall terms of 2012. Some individuals have fewer courses included in the average if they taught fewer than four courses in the year, while others have more than four courses: if they taught, for example, two courses in Fall 2012, and three in Summer 2012, then all five courses would be included in their average. Open Labs have been excluded from the average class size as they do not have enrollment numbers. The average class sizes have been grouped into categories that are not evenly spaced, in the hope of reflecting the different class types at SFU.

3.2 FEEDBACK ON HOW THESE EVALUATION RESULTS SHOULD BE USED

Faculty were next asked to indicate to what extent teaching and course evaluation results should impact various decision-making processes.

- 77-78% of respondents indicated that results should be used to a great or moderate extent in:
 - making changes to teaching methods,
 - identifying opportunities for teacher development,
 - review of candidates for teaching roles/positions, and
 - making changes to course designs.
- 68% indicated that tenure and promotion should be impacted by the results to a great or moderate extent.
- Respondents were least likely to indicate that results should greatly or moderately impact merit increases (62%).

Table 1 displays the results broken down by faculty/instructor rank.

From the table:

- Sessionals were most likely to indicate that the results should be used to a great/moderate extent in:
 - making changes to course designs (83%),
 - making changes to teaching methods (82%),
 - identifying opportunities for teacher development (81%), and
 - review of candidates for teaching roles/positions (78%).
- For all four of these statements, non-tenure track lecturers were least likely to say that decisions should be greatly/moderately impacted by evaluation results (68-76%).
- Sessionals were least likely (59%) and non-tenure track lecturers were most likely (68%) to say that evaluation results should impact “merit increases” to a great or moderate extent.
- Tenure track respondents were in the middle for all statements.

Table 1: Percent Saying Evaluation Results Should Greatly/Moderately Affect Decisions, by Rank

Decision	Overall (n=492)	Tenure Track (n=281)	Non-Tenure Track (n=62)	Sessionals (n=149)
Making changes to teaching methods	78%	77%	76%	82%
Identifying opportunities for teacher development	78%	77%	69%	81%
Review of candidates for teaching roles/positions	77%	77%	74%	78%
Making changes to course designs	77%	75%	68%	83%
Tenure and promotion	68%	69%	69%	66%
Merit increases	62%	62%	68%	59%

Notes: The ranks used are the most recent ranks of faculty members and instructors.

Those with more than one rank in the most recent semester have been categorized under the higher rank.

Tenure track respondents include Professors, Associate Professors, and Assistant Professors; non-tenure track respondents include Lecturers and Senior Lecturers; sessionals include Sessional Instructors and Sessional Lecturers.

3.3 IMPROVEMENTS TO THE CURRENT PROCESS

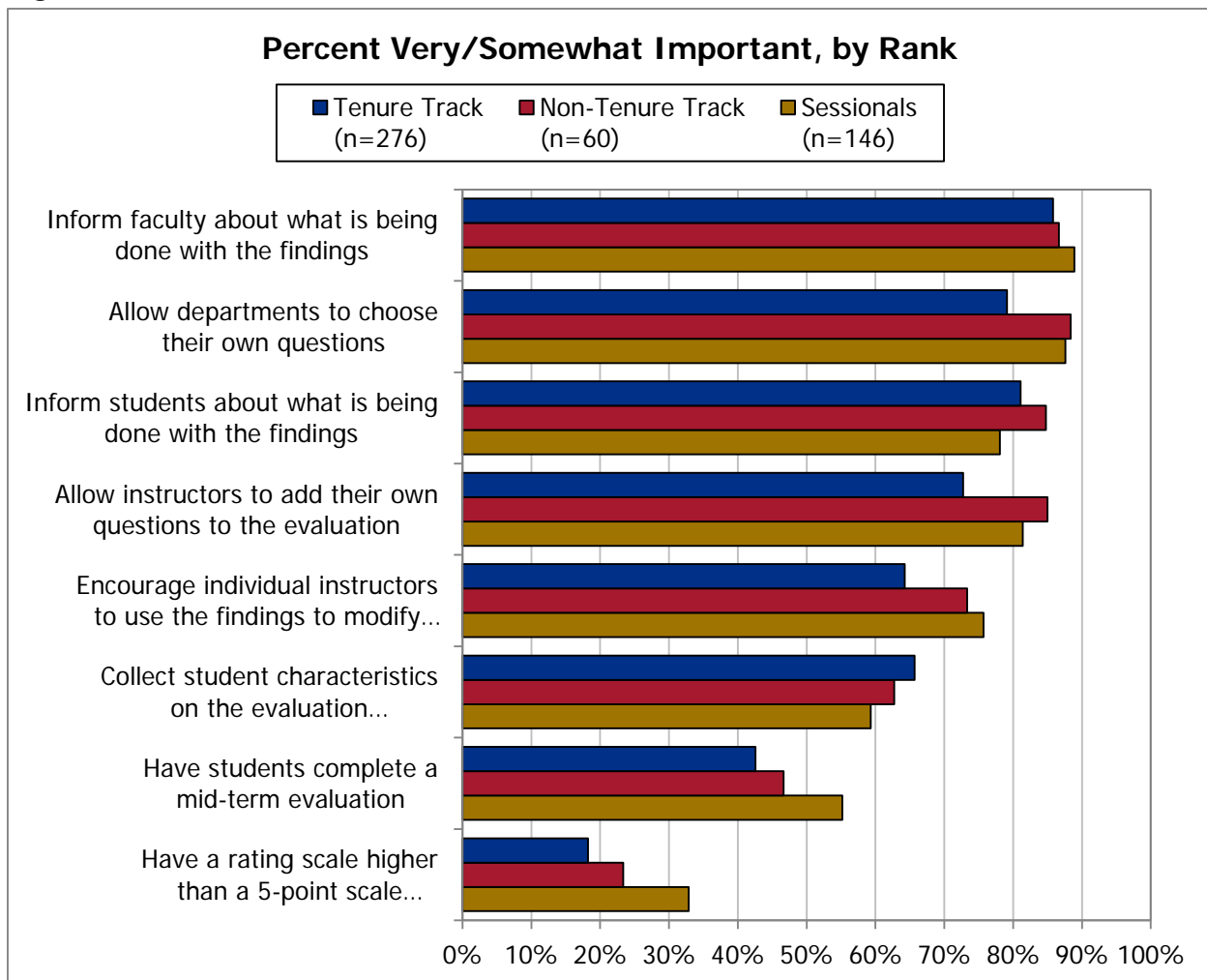
The third survey question asked faculty members what they would do to improve the teaching and course evaluation process and to better use the results. Respondents were given a list of options and asked to rate their importance. Overall:

- Respondents were most likely to report informing faculty members about what is being done with the findings as very or somewhat important (87%), followed by allowing customizable sections where departments may choose their own questions (83%), and clearly informing students about what is being done with the findings (81%).
- Few respondents indicated that it was important to have the students complete a mid-term evaluation (47%) and for the rating scale to be changed (23%).

Tenure track, non-tenure track, and sessional respondents tended to rate the importance of the different options similarly (see Figure 2), with the following main differences:

- a higher percent of sessionals rated “have a rating scale higher than a 5-point scale” as very or somewhat important compared to tenure track and non-tenure track (33% vs. 18-23%),
- those in tenure track were less likely to rate “encourage individual instructors to use the findings to modify their teaching or the course” as important (64% vs. 73-76%), and
- those in tenure track were less likely to rate “allow instructors to add their own questions to the evaluation, with the responses viewable only by the instructor” as important (73% vs. 81-85%).

Figure 2



Notes: The ranks used are the most recent ranks of faculty members and instructors. Those with more than one rank in the most recent semester have been categorized under the higher rank. Tenure track respondents include Professors, Associate Professors, and Assistant Professors; non-tenure track respondents include Lecturers and Senior Lecturers; sessionals include Sessional Instructors and Sessional Lecturers.

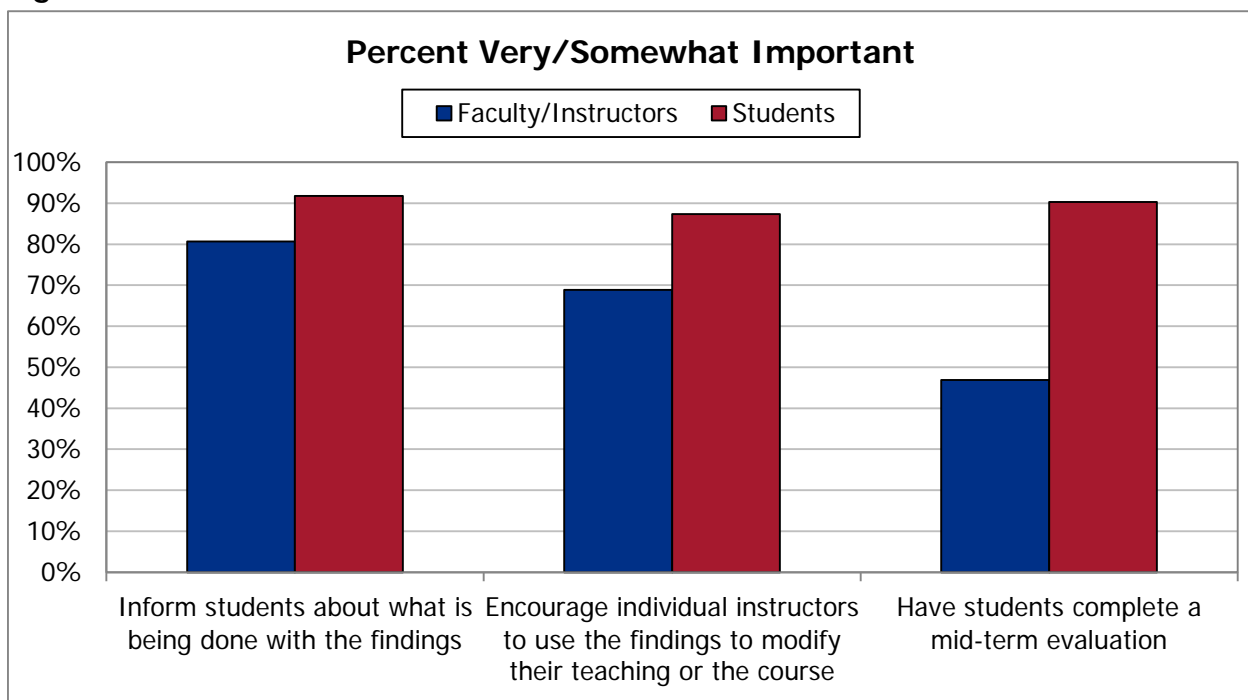
In the Fall 2012 Undergraduate Student Survey, students were asked a similar question about the teaching and course evaluation process. Students were given a set of possible changes to the process, and asked to rate the importance of each. Three of the changes are similar to

those asked of faculty and instructors: “clearly inform students about what is being done with the findings”, “have students complete a mid-term evaluation in addition to the end-of-term evaluation”, and “encourage individual instructors to use the findings to modify their teaching or the course.”

Figure 3 displays the results separately for faculty/instructors and students. From the figure:

- 92% of student respondents said that informing students about what is being done with the findings is very or somewhat important (vs. 81% of faculty/instructor respondents),
- 90% of student respondents said completing a mid-term evaluation is important (vs. 47% of faculty/instructor respondents), and
- 87% of student respondents indicated that having individual instructors use the findings to modify their teaching or the course is important (vs. 69% of faculty/instructor respondents).

Figure 3



Finally, the last question asked faculty and instructors which aspects of teaching they thought needed to be included on an evaluation in order for it to be a representative assessment of teaching. The top choices were:

- organization of teaching and course material (82%),
- clarity of presentation (81%),
- appropriate assessment and feedback (78%), and
- stimulation of critical thinking (78%).

Table 2: Top Aspects to Include on an Evaluation of Teaching, by Faculty

Faculty Respondents	of Top 3 Aspects to Include on an Evaluation of Teaching
Applied Sciences (n=38)	<ul style="list-style-type: none"> • Clarity of presentation (92%) • Organization of teaching and course material (89%) • Two-way tie: <ul style="list-style-type: none"> ○ Concern for student learning (74%) ○ Appropriate assessment and feedback (74%)
Arts and Social Sciences (n=147)	<ul style="list-style-type: none"> • Two-way tie: <ul style="list-style-type: none"> ○ Organization of teaching and course material (84%) ○ Appropriate assessment and feedback (84%) • Clarity of presentation (80%)
Beedie School of Business (n=34)	<ul style="list-style-type: none"> • Stimulation of critical thinking (88%) • Clarity of presentation (82%) • Concern for student learning (79%)
Communication, Art and Technology (n=50)	<ul style="list-style-type: none"> • Communication of clear goals and expectations (82%) • Appropriate assessment and feedback (80%) • Three-way tie: <ul style="list-style-type: none"> ○ Stimulation of critical thinking (76%) ○ Clarity of presentation (76%) ○ Organization of teaching and course material (76%)
Education (n=52)	<ul style="list-style-type: none"> • Appropriate assessment and feedback (83%) • Two-way tie: <ul style="list-style-type: none"> ○ Stimulation of critical thinking (81%) ○ Organization of teaching and course material (81%)
Environment (n=23)	<ul style="list-style-type: none"> • Two-way tie: <ul style="list-style-type: none"> ○ Clarity of presentation (83%) ○ Respect for students (83%) • Two-way tie: <ul style="list-style-type: none"> ○ Stimulation of critical thinking (78%) ○ Instructor enthusiasm (78%)
Health Sciences (n=19)	<ul style="list-style-type: none"> • Expertise in subject matter (95%) • Stimulation of critical thinking (84%) • Two-way tie: <ul style="list-style-type: none"> ○ Clarity of presentation (68%) ○ Organization of teaching and course material (68%)
Science (n=102)	<ul style="list-style-type: none"> • Organization of teaching and course material (88%) • Clarity of presentation (84%) • Ability to stimulate student interest (81%)

Notes: Faculty members and instructors have been categorized into their primary Faculty. Those without a primary Faculty have been categorized into the Faculty of their most recently taught course. Data from some Faculties are based on small sample sizes, so care should be taken when interpreting the results.

The top choices by Faculty are displayed in Table 2. From the table:

- “Clarity of presentation” was among the top three selected by respondents in all Faculties except Education.
- “Organization of teaching and course material” was among the top choices in all Faculties except the Beedie School of Business and the Faculty of Environment.

- Some aspects were part of the top choices for only one Faculty:
 - Only respondents from Communication, Art and Technology selected “communication of clear goals and expectations” among their top choices,
 - those from the Faculty of Environment had “respect for students” and “instructor enthusiasm” in their top choices,
 - Health Sciences was the only Faculty to have “expertise in subject matter” among their top choices, and
 - only respondents from Science had “ability to stimulate student interest” among the top three aspects of teaching.
- The least chosen aspects by respondents in all Faculties were “facilitation of group interaction”, “adherence to the course outline”, “rapport with students”, and “workload assigned to students.”

Table 3 displays the results by the number of years faculty/instructors have been in their current position at SFU. From the table:

- Respondents in all groups agree on the top 2 aspects to include on an evaluation of teaching: “organization of teaching and course material” and “clarity of presentation.”
- Faculty and instructors who have been in their current position for 3-5 years had “expertise in subject matter” as their third choice, whereas those who have been in their current position for less than 3 years or more than 5 years have “stimulation of critical thinking” as their third choice.

Table 3: Top Aspects to Include on an Evaluation of Teaching, by Years in Current Position at SFU

Years in Current Position	Top 3 Aspects to Include on an Evaluation of Teaching
0-2 Years (n=230)	<ul style="list-style-type: none"> • Organization of teaching and course material (83%) • Clarity of presentation (80%) • Stimulation of critical thinking (80%)
3-5 Years (n=114)	<ul style="list-style-type: none"> • Organization of teaching and course material (84%) • Clarity of presentation (83%) • Expertise in subject matter (78%)
More than 5 Years (n=121)	<ul style="list-style-type: none"> • Clarity of presentation (79%) • Organization of teaching and course material (78%) • Stimulation of critical thinking (76%)

Notes: The start date used to calculate years in current position is the date on the maximum effective dated record. For those with contracts, the effective date of their most recent contract is used.
The end date used to calculate years in current position is January 1, 2013.

4 SURVEY INSTRUMENT

The “Teaching and Course Evaluation Process” refers to the current process whereby at the end of the term, students are asked to fill out a questionnaire about each class they take, asking them to evaluate both the course and the instructor(s).

- 1) Given the current Teaching and Course Evaluation Process, please state your level of agreement with the following:
(Scale: Strongly Agree, Agree, Disagree, Strongly Disagree, Undecided)
 - a. The instrument asks questions that evaluate teaching in a fair and accurate manner.
 - b. The instrument asks questions that evaluate courses in a fair and accurate manner.
 - c. The evaluation results are useful in helping me to design/modify my method of instruction.

- 2) In your opinion, to what extent should teaching and course evaluation results have an impact on the decision-making process for each of the following? Please select “Not At All” if you feel that these evaluation results should not be used.
(Scale: Not At All, Very Little, To a Moderate Extent, To a Great Extent, No Opinion)
 - a. Tenure and promotion
 - b. Review of candidates for teaching roles/positions
 - c. Merit increases
 - d. Identifying opportunities for teacher development
 - e. Making changes to course designs
 - f. Making changes to teaching methods

- 3) If it were up to you, what would you do to improve the Teaching and Course Evaluation Process, and to better use the results? Please rate the importance of each of these options.
(Scale: Very Important, Somewhat Important, Not Very Important, Not at all Important, Undecided)
 - a. Collect student characteristics on the evaluation (e.g. gender, GPA, major).
 - b. Clearly inform faculty about what is being done with the findings.
 - c. Allow instructors to add their own questions to the evaluation, with the responses viewable only by the instructor.
 - d. Clearly inform students about what is being done with the findings.
 - e. Allow customizable sections where departments may choose their own questions.
 - f. Have students complete a mid-term evaluation in addition to the end-of-term evaluation.
 - g. Have a rating scale higher than a 5-point scale (e.g. 1-7, 1-10).
 - h. Encourage individual instructors to use the findings to modify their teaching or the course.

- 4) What aspects of teaching do you think need to be included on an evaluation in order for it to be a representative assessment of teaching? Please select all that apply.
 - a. Rapport with students
 - b. Organization of teaching and course material
 - c. Instructor enthusiasm

- d. Fairness of grading
 - e. Ability to stimulate student interest
 - f. Clarity of presentation
 - g. Respect for students
 - h. Concern for student learning
 - i. Appropriate assessment and feedback
 - j. Facilitation of group interaction
 - k. Expertise in subject matter
 - l. Stimulation of critical thinking
 - m. Communication of clear goals and expectations
 - n. Overall teaching behavior
 - o. Workload assigned to students
 - p. Adherence to the course outline
 - q. Other, please specify..._____
- 5) Please write any concerns or comments you have about the Teaching and Course Evaluation Process.

5 DISTRIBUTION OF FACULTY AND INSTRUCTOR RESPONSES

This section displays the distribution of faculty and instructor responses to the survey questions. The tables that follow show the number and percentage of faculty and instructors selecting each response to the questions. In cases where faculty and instructors could select more than one response, the percentage of all respondents (percentage of all those who answered the question) are provided.

Please note that percentages in these tables are rounded to one decimal place. Summaries that group categories (e.g. where "Very Important" and "Somewhat Important" are grouped) may result in percentages that do not exactly match the sum of the percentages displayed in the tables. This is an artifact of rounding.

1) **Given the current Teaching and Course Evaluation Process, please state your level of agreement with the following:**

The instrument asks questions that evaluate teaching in a fair and accurate manner.		
Strongly Agree	10	2.0%
Agree	234	46.6%
Undecided	39	7.8%
Disagree	145	28.9%
Strongly Disagree	74	14.7%
Total Responses	502	100.0%
Missing Cases	17	

The instrument asks questions that evaluate courses in a fair and accurate manner.		
Strongly Agree	10	2.0%
Agree	223	44.5%
Undecided	45	9.0%
Disagree	164	32.7%
Strongly Disagree	59	11.8%
Total Responses	501	100.0%
Missing Cases	18	

The evaluation results are useful in helping me to design/modify my method of instruction.		
Strongly Agree	26	5.2%
Agree	218	43.5%
Undecided	44	8.8%
Disagree	132	26.3%
Strongly Disagree	81	16.2%
Total Responses	501	100.0%
Missing Cases	18	

Summary		% Strongly Agree / Agree
The evaluation results are useful in helping me to design/modify my method of instruction.		48.7%
The instrument asks questions that evaluate teaching in a fair and accurate manner.		48.6%
The instrument asks questions that evaluate courses in a fair and accurate manner.		46.5%

2) **In your opinion, to what extent should teaching and course evaluation results have an impact on the decision-making process for each of the following? Please select "Not At All" if you feel that these evaluation results should not be used.**

Tenure and promotion		
To a Great Extent	86	17.5%
To a Moderate Extent	249	50.7%
Very Little	109	22.2%
Not At All	34	6.9%
No Opinion	13	2.6%
Total Responses	491	100.0%
Missing Cases	28	

Review of candidates for teaching roles/positions		
To a Great Extent	135	27.6%
To a Moderate Extent	242	49.5%
Very Little	85	17.4%
Not At All	20	4.1%
No Opinion	7	1.4%
Total Responses	489	100.0%
Missing Cases	30	

Merit increases		
To a Great Extent	76	15.5%
To a Moderate Extent	227	46.3%
Very Little	122	24.9%
Not At All	47	9.6%
No Opinion	18	3.7%
Total Responses	490	100.0%
Missing Cases	29	

Identifying opportunities for teacher development		
To a Great Extent	135	27.6%
To a Moderate Extent	244	49.9%
Very Little	81	16.6%
Not At All	20	4.1%
No Opinion	9	1.8%
Total Responses	489	100.0%
Missing Cases	30	

Making changes to course designs		
To a Great Extent	92	18.7%
To a Moderate Extent	285	58.0%
Very Little	91	18.5%
Not At All	20	4.1%
No Opinion	3	0.6%
Total Responses	491	100.0%
Missing Cases	28	

Making changes to teaching methods		
To a Great Extent	100	20.4%
To a Moderate Extent	284	57.8%
Very Little	85	17.3%
Not At All	17	3.5%
No Opinion	5	1.0%
Total Responses	491	100.0%
Missing Cases	28	

Summary		
	% To a Great / Moderate Extent	
Making changes to teaching methods	78.2%	
Identifying opportunities for teacher development	77.5%	
Review of candidates for teaching roles/positions	77.1%	
Making changes to course designs	76.7%	
Tenure and promotion	68.2%	
Merit increases	61.8%	

3) **If it were up to you, what would you do to improve the Teaching and Course Evaluation Process, and to better use the results? Please rate the importance of each of these options.**

Collect student characteristics on the evaluation (e.g. gender, GPA, major).

Very Important	134	28.0%
Somewhat Important	169	35.4%
Undecided	13	2.7%
Not Very Important	99	20.7%
Not at all Important	63	13.2%
Total Responses	478	100.0%
Missing Cases	41	

Clearly inform faculty about what is being done with the findings.

Very Important	285	59.5%
Somewhat Important	131	27.3%
Undecided	9	1.9%
Not Very Important	32	6.7%
Not at all Important	22	4.6%
Total Responses	479	100.0%
Missing Cases	40	

Allow instructors to add their own questions to the evaluation, with the responses viewable only by the instructor.

Very Important	180	37.5%
Somewhat Important	189	39.4%
Undecided	21	4.4%
Not Very Important	65	13.5%
Not at all Important	25	5.2%
Total Responses	480	100.0%
Missing Cases	39	

Clearly inform students about what is being done with the findings.

Very Important	220	45.8%
Somewhat Important	167	34.8%
Undecided	16	3.3%
Not Very Important	44	9.2%
Not at all Important	33	6.9%
Total Responses	480	100.0%
Missing Cases	39	

Allow customizable sections where departments may choose their own questions.

Very Important	194	40.6%
Somewhat Important	202	42.3%
Undecided	14	2.9%
Not Very Important	43	9.0%
Not at all Important	25	5.2%
Total Responses	478	100.0%
Missing Cases	41	

Have students complete a mid-term evaluation in addition to the end-of-term evaluation.		
Very Important	86	17.9%
Somewhat Important	139	29.0%
Undecided	32	6.7%
Not Very Important	114	23.8%
Not at all Important	109	22.7%
Total Responses	480	100.0%
Missing Cases	39	

Have a rating scale higher than a 5-point scale (e.g. 1-7, 1-10).		
Very Important	29	6.1%
Somewhat Important	82	17.2%
Undecided	29	6.1%
Not Very Important	141	29.6%
Not at all Important	196	41.1%
Total Responses	477	100.0%
Missing Cases	42	

Encourage individual instructors to use the findings to modify their teaching or the course.		
Very Important	103	21.5%
Somewhat Important	226	47.3%
Undecided	24	5.0%
Not Very Important	87	18.2%
Not at all Important	38	7.9%
Total Responses	478	100.0%
Missing Cases	41	

Summary		% Very / Somewhat Important
Clearly inform faculty about what is being done with the findings.		86.8%
Allow customizable sections where departments may choose their own questions.		82.9%
Clearly inform students about what is being done with the findings.		80.6%
Allow instructors to add their own questions to the evaluation, with the responses viewable only by the instructor.		76.9%
Encourage individual instructors to use the findings to modify their teaching or the course.		68.8%
Collect student characteristics on the evaluation (e.g. gender, GPA, major).		63.4%
Have students complete a mid-term evaluation in addition to the end-of-term evaluation.		46.9%
Have a rating scale higher than a 5-point scale (e.g. 1-7, 1-10).		23.3%

4) **What aspects of teaching do you think need to be included on an evaluation in order for it to be a representative assessment of teaching? Please select all that apply.**

Organization of teaching and course material	381	81.9%
Clarity of presentation	376	80.9%
Appropriate assessment and feedback	363	78.1%
Stimulation of critical thinking	363	78.1%
Communication of clear goals and expectations	345	74.2%
Expertise in subject matter	339	72.9%
Ability to stimulate student interest	332	71.4%
Respect for students	328	70.5%
Concern for student learning	324	69.7%
Instructor enthusiasm	305	65.6%
Fairness of grading	281	60.4%
Overall teaching behaviour	279	60.0%
Workload assigned to students	227	48.8%
Rapport with students	203	43.7%
Facilitation of group interaction	176	37.8%
Adherence to the course outline	153	32.9%
Other	76	16.3%
Total Respondents	465	

5) **Please write any concerns or comments you have about the Teaching and Course Evaluation Process.**

Comments are not provided here, but were provided to the Teaching and Course Evaluation Project Committee to inform recommendations.



SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

TEACHING & COURSE EVALUATION PROJECT FINAL REPORT

APPENDIX IX: FOCUS GROUP KEY FINDINGS SUMMARY

NOVEMBER 18, 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

DOCUMENT CONTROL

DOCUMENT INFORMATION

	INFORMATION
Document Owner	<i>Corinne Pitre-Hayes</i>
Project /Organization Role	<i>Project Manager, Teaching & Course Evaluation Project</i>

DOCUMENT HISTORY

VERSION	ISSUE DATE	CHANGES
<i>0.1</i>	<i>October 15, 2013</i>	<i>Initial draft</i>
<i>0.2</i>	<i>October 17, 2013</i>	<i>Revisions to Education</i>
<i>0.3</i>	<i>November 15, 2013</i>	<i>Revisions and addition of focus groups</i>
<i>0.4</i>	<i>November 17, 2013</i>	<i>Minor revisions and formatting</i>
<i>0.5</i>	<i>November 18, 2013</i>	<i>Final edits before distribution to SCUTL</i>

TABLE OF CONTENTS

1	INTRODUCTION	1
2	OVERVIEW OF FOCUS GROUPS	1
3	APPROACH	4
3.1	OBJECTIVES.....	4
3.2	QUESTIONS FOR FACULTY	4
3.3	QUESTIONS FOR STUDENTS	4
4	HIGH-LEVEL OVERVIEW	5
5	DETAILED SUMMARIES	10
5.1	BEEDIE SCHOOL OF BUSINESS	10
5.2	FACULTY OF APPLIED SCIENCES	12
5.3	FACULTY OF APPLIED SCIENCES (SCHOOL OF COMPUTING SCIENCE)	14
5.4	FACULTY OF ARTS AND SOCIAL SCIENCES	16
5.5	FACULTY OF EDUCATION	19
5.6	FACULTY OF HEALTH SCIENCES	21
5.7	FACULTY OF SCIENCE	23
5.8	SFU FACULTY ASSOCIATION	25
5.9	TPC CHAIRS	27
5.10	STUDENTS	29

1 INTRODUCTION

The literature research and consultation with other institutions provided the Teaching and Course Evaluation (TCE) project team with context for engagement with the SFU community. An essential component of community engagement was a series of focus groups with stakeholder groups. The focus groups were aimed at exploring questions in more depth with participants. Stakeholder groups were strongly encouraged to include participants with a wide range of views on the topic. It was repeatedly stressed that all points of view were welcome.

The goal was to use this information to help inform our recommendations to Senate for instrument selection and related processes. Understanding what our faculties and departments are doing today provides meaningful input for instrument selection and support. This report provides a summary of key findings from this stage of engagement with the SFU community.

2 OVERVIEW OF FOCUS GROUPS

The TCE project team conducted 11 focus groups across 6 faculties and 3 important stakeholder groups within the SFU community. Table 1 summarizes the focus groups and the dates they were held. Table 2 provides a summary of the roles of participants in each focus group.

Table 1: Focus Groups and the Dates they were Conducted

Focus Group	Date
Faculty of Education	May 31, 2012
Students	Oct 30, 2012
Faculty of Arts and Social Sciences	December 6, 2012 December 13, 2012
Faculty of Applied Sciences (School of Engineering) (School of Computing Science)	February 13, 2013 March 14, 2013
Beedie School of Business	February 22, 2013
Faculty of Science	February 25, 2013
Faculty of Health Sciences	April 15, 2013
TPC Chairs	April 19, 2013
SFU Faculty Association	May 30, 2013

Table 2: Overview of the Roles of Focus Group Participants

Roles of Participants
Beedie School of Business
Senior Lecturer
Lecturer
Lecturer
Lecturer
Admin Staff
Admin Staff
Lecturer
Associate Professor

Faculty of Applied Sciences

Associate Professor
Senior Lecturer
Limited-term Lecturer
Professor
Senior Lecturer
Assistant Professor
Senior Lecturer
Lecturer
Admin Staff
Lecturer
Lecturer

Faculty of Arts and Social Sciences

Senior Lecturer
Associate Professor
Senior Lecturer
Assistant Professor
Education Consultant
Assistant Professor
Associate Professor
Professor
Assistant Professor
Lecturer

Faculty of Education

Professor
Professor
Assistant Professor
Professor

Faculty of Health Sciences

Faculty (Lecturer)
Admin Staff
Faculty (Lecturer)
Faculty
Student Rep (Undergraduate)
Faculty (Associate Professor)
Faculty (Assistant Professor)

Faculty of Science

Professor

Senior Lecturer

Associate Professor

Professor

Professor

Professor

Professor

SFU Faculty Association

Admin staff

Faculty of Science

Beedie School of Business

Department of Statistics

SFU Library

Department of Linguistics

TPC Chairs

Former TPC Chair (Molecular Biology and Biochemistry)

Former TPC Chair (English)

Former TPC Chair (Education)

Students

18 Learning Commons Peer Educators

3 APPROACH

Our goal was to gather the needs of the SFU community and understand its concerns. To facilitate this, each session began with a brief overview of the project that included highlights from the literature review and consultation with other institutions. Then dialogue was structured around key questions that helped to explore how concerns might be addressed through relevant instrument design principles.

3.1 OBJECTIVES

The objectives of the focus groups were to:

- Gather faculty/department/school/program needs with respect to student evaluation of teaching and courses
- Understand attitudes, feelings and concerns with respect to the student evaluation of teaching and courses instrumentation and use of data acquired with the instrument
- Ensure the community feels that we are actively listening to their attitudes, feelings and concerns
- Address some of the concerns that arise with instrument design principles, review of the research, and practice at other institutions.

3.2 QUESTIONS FOR FACULTY

The four questions considered at each focus group session with faculty were:

- 1) How are student evaluations of teaching and courses used in your faculty? (Tenure and promotion? Improvement of teaching? Improvement of courses?)
- 2) How do you feel about the current instrument?
- 3) How do you feel about how the evaluation data are used?
- 4) What are your main concerns with respect to student evaluation of teaching? Within your own faculty? In the institution as a whole?
- 5) How can we make improvements that will address these concerns?

3.3 QUESTIONS FOR STUDENTS

Questions considered at the focus group with students included:

- 1) What is your perception of how student evaluations of teaching and courses are used at Simon Fraser University? (Instructor evaluation? Improvement of teaching? Improvement of courses?)
- 2) What do you feel are the most important reasons for giving a low evaluation of teaching and courses?
- 3) What do you feel are the most important reasons for giving a high evaluation of teaching and courses?
- 4) What are your main concerns with respect to student evaluation of teaching and courses? Within your program? In the institution as a whole? What should be done with the results versus what is currently done?
- 5) How can we make improvements that will address your concerns?

4 HIGH-LEVEL OVERVIEW

Table 3 below provides a high-level overview of comments and discussion from each of the stakeholder groups, with the exception of the student focus group. The different perspectives of evaluation topics that were discussed at each session, including the student focus group are provided in the detailed summaries that follow.

Table 3: High Level Overview of Comments and Discussion from each Stakeholder Group (except Students)

CURRENT INSTRUMENT	
Beedie School of Business	<p>Current system asks students about teaching</p> <p>No context to different types of courses/classes is provided</p> <p>Some like the existing instrument – others think it needs to be revised</p> <p>Provides data for tenure and promotion</p>
Faculty of Arts and Social Sciences	<p>Evaluations are imposed</p> <p>Many instructors conducting their own formative evaluations</p> <p>Don't trust that the instrument produces accurate results</p> <p>Instrument doesn't help students see the difference between learning outcomes and teaching performance</p>
Faculty of Applied Sciences (School of Engineering)	<p>Paper based forms that include both bubble and open-answer formats</p> <p>Intro paragraph that explains what evaluation results are used for</p> <p>Find the forms "useless"</p> <p>Some instructors have created their own surveys</p>
Faculty of Applied Sciences (School of Computing Science)	<p>Paper based instrument that gathers student feedback on teaching</p> <p>Used for tenure and promotion</p> <p>Instrument is too narrow – doesn't ask the right questions</p> <p>Meets SFU policy</p> <p>Instructors conducting their own formative evaluations</p>
Faculty of Education	<p>Instrument is not used effectively</p> <p>Unreliable / unvalidated instrument</p>
Faculty of Health Sciences	<p>Online instrument is used</p> <p>Data is used by instructors and committees responsible for promotion, tenure and hiring</p> <p>Dislike the low response rate (approximately 50%)</p> <p>Instructors encourage students to complete the survey, incentives are offered</p> <p>Instrument is too narrow – asks the wrong questions</p>
Faculty of Science	<p>Used to gather student feedback on teaching</p> <p>Data is used for making improvements to teaching and making decisions about tenure and promotion</p> <p>Instrument is too narrow – it doesn't ask the right questions</p> <p>The instrument is used too late in the semester to be effective</p>
SFU Faculty Association	<p>Instruments place more emphasis on tenure and promotion than on enhancing teaching and learning</p> <p>Paper based instruments work well</p> <p>In some cases, there's a lack of clarity about what the instrument is evaluating</p>
TPC Chairs	<p>TPC uses results to review salaries or make decisions about instructor merit, tenure or promotion</p> <p>Compare results for similar courses</p> <p>Review student comments</p>

USE OF DATA	
Beedie School of Business	<p>Used for tenure and promotion</p> <p>Used for making hiring decisions about new lecturers</p> <p>Used by the Associate Dean and department chairs to assess instructors</p> <p>Some instructors use data and written comments in a formative way</p> <p>Some instructors conduct their own mid-term evaluations to receive formative data</p>
Faculty of Arts and Social Sciences	<p>Results are used for promotion and tenure</p> <p>Departments all use data in different ways – use of data is selective</p> <p>Evaluation may or may not impact teaching methods</p> <p>Some data is not useful for instructors at all</p>
Faculty of Applied Sciences	<p>Used for making tenure and promotion decisions</p> <p>Lack of context provided for data</p> <p>Used to improve teaching and curriculum</p> <p>Instructors interested in written comments</p>
Faculty of Applied Sciences (School of Computing Science)	<p>Data is almost exclusively used for teaching and promotion decision-making</p> <p>Data is usually considered out of context</p> <p>Instructors use written comments</p>
Faculty of Education	<p>Issues with formatting and inflexibility of questions</p> <p>No best practices used</p> <p>Not sure the data is accurate</p> <p>Not enough time to interpret the data</p> <p>No comparative analysis</p> <p>Not consistently used for teaching improvement or tenure/promotion</p>
Faculty of Health Sciences	<p>Summative data is used for administrative purposes</p> <p>Instructors use written comments in a formative way</p>
Faculty of Science	<p>Very limited use, mostly for tenure and promotion</p> <p>Numbers are not a good indicator of teaching quality</p> <p>Context of the data is not considered</p> <p>Data is not flexible and accessible</p>
SFU Faculty Association	<p>Used for tenure and promotion</p> <p>Does not do a good job of measuring lab performance</p> <p>In some cases, the data is considered useless</p> <p>The timing of evaluations is important</p>
TPC Chairs	<p>The way results are used depends on the department rather than the faculty</p> <p>Inconsistencies could be related to the lack of training for TPC chairs</p>

KEY CONCERNS	
Beedie School of Business	<ul style="list-style-type: none"> Perception of how data used The current instrument just assesses teaching, not courses Access to data is an issue Students don't know how to complete evaluations Timing of the evaluation
Faculty of Arts and Social Sciences	<ul style="list-style-type: none"> Student responses are not accurate or helpful Ineffective instrument Pressured to share results with students Many distrust the evaluation process because of inconsistencies Instrument is challenging to use which could contribute to low response rates Data is inaccurate Some courses are destined to receive negative feedback from students Student resistance to course learning outcomes and learner-centered instruction
Faculty of Applied Sciences	<ul style="list-style-type: none"> Data used for tenure and promotion is out of context Results are negatively impacted because of external factors and not being able to get all students to complete the evaluation
Faculty of Applied Sciences (School of Computing Science)	<ul style="list-style-type: none"> Students don't care about giving feedback Low response rate Lack of accountability for students Added workload for faculty and staff Questions are too generic Too many characteristic-oriented questions Questions are not worded so that all students can interpret them
Faculty of Education	<ul style="list-style-type: none"> Communication with students about the process is necessary Students misinterpret questions Results not being used consistently for tenure / promotion or for making teaching improvement
Faculty of Health Sciences	<ul style="list-style-type: none"> Students lack of interest and negative comments External factors and instructors' job satisfaction can negatively impact evaluations Teachers are being hired despite bad reviews Inaccurate data
Faculty of Science	<ul style="list-style-type: none"> The ethical and appropriate use of data Types of questions used Added workload for faculty and staff Students and faculty don't care about evaluations
SFU Faculty Association	<ul style="list-style-type: none"> Most concerns relate to how data is used Viewpoint that TPCs "always extract the negative information" from evaluations Several faculty members don't care about the evaluation process Issues moving to an online system Students colluding or getting together to fill out evaluations Concern that the existing system is not obsolete
TPC Chairs	<ul style="list-style-type: none"> Lower response rates and instructor ratings using online instrument Instructors want to return to paper based system Faculty do not respect the evaluations Giving the survey out at the end of a class devalues the evaluations Validity and reliability of the instrument and data is poor Peers without expertise shouldn't evaluate peers Cannot force new forms of evaluation on instructors Students are not qualified to comment on some aspects No standards for extracting and analyzing data / student comments The context of questions on the survey impact the data Evaluation data is often used "sloppily"

SUGGESTED IMPROVEMENTS

Beedie School of Business	<ul style="list-style-type: none">Clearly identify the problems with the existing evaluation that need to be fixedProvide meaningful dataConsistency with tenure and promotionProvide communication and instruction about the evaluation process to both instructors and studentsInclude peer evaluationsDesign and structure questions that would provide a better experience for students and better data for departments/instructorsConsider the timing and instrument used for the evaluation
Faculty of Arts and Social Sciences	<ul style="list-style-type: none">Better communication with students – define expectations for evaluation processKeep students involved – instructors should respond to feedbackOffer an online optionProvide better recognition for teaching excellenceStructure the survey better – ask more detailed questionsConsider influences that could impact student responses when analyzing dataEvaluations from peers/education consultantsCollect formative and summative data
Faculty of Applied Sciences	<ul style="list-style-type: none">Use an online instrumentCollect feedback on programsAsk questions after each classUse a question bankUse core questions combined with customizable optionsAsk open-ended questionsHave a tool that can compile results for sharing and provide contextIndicate how many students haven't completed the evaluationConduct an alumni surveyGet feedback from other instructorsMake the use of data for tenure and promotion more consistentOffer incentives to studentsShare results with students
Faculty of Applied Sciences (School of Computing Science)	<ul style="list-style-type: none">Evaluate departments and programsCreate a set of best practices for the institutionHave an online evaluation instrument that's easy to useUse a 5-point scaleHave an instrument than can be tailored for each department, course and levelUse specific questions about coursesExplain to students the reasons for conducting evaluationsOffer incentives to studentsMake evaluations compulsoryEmploy different types of evaluationsInclude other evaluation methods and perspectivesConduct the evaluation at strategic times during the semesterAllow instructors to conduct their own evaluations

Faculty of Education	<p>Use the evaluation to identify teaching excellence</p> <p>Create a balance between quantitative and qualitative data</p> <p>Focus on learning</p> <p>Develop a bank of questions that will provide reliable data</p> <p>Flexibility in the selection of questions</p> <p>Ability to compare/analyze data more effectively</p>
Faculty of Health Sciences	<p>Explain to students the value of participating in the evaluation</p> <p>Offer incentives</p> <p>Make the results public</p> <p>Use an instrument that is flexible across departments, courses, levels and instructors.</p> <p>Provide the right tools, resources and facilities for instructors</p> <p>Peer evaluation and professional help with teaching</p> <p>Make evaluations mandatory for students</p> <p>Do not tolerate inappropriate comments from students</p> <p>Survey faculties on barriers to teaching</p>
Faculty of Science	<p>Use specific questions designed to help improve teaching</p> <p>Have a flexible instrument</p> <p>Tailor surveys for departments, courses and levels</p> <p>Explain to students the reason for the evaluation</p> <p>Conduct mid-term evaluations</p> <p>Making the results available to students as soon as possible</p> <p>Offer incentives like additional marks</p> <p>An easy-to-use instrument that doesn't increase instructor workload</p> <p>Opportunity for faculty to explain context of low scores</p> <p>Evaluations from educational consultants</p>
SFU Faculty Association	<p>Needs for evaluations include: student learning, tenure and promotion and a flexible instrument</p> <p>Define the purpose of student evaluations of teaching and courses for Senate</p> <p>Evaluate the merits of online versus paper evaluations</p> <p>Define how evaluation data should be used</p> <p>Design measurements of performance outside the lecture setting</p> <p>Add context to evaluation reports</p>
TPC Chairs	<p>Determine what students should be evaluating</p> <p>Give attention to valid data interpretation – provide guidance and training</p> <p>Use a flexible online instrument with a question bank</p> <p>Allow departments to create questions</p> <p>Use better questions that are in the right sequence</p> <p>Use a system that doesn't increase department workload</p> <p>Change departmental guidelines to include other forms of evaluation</p> <p>Consider more data than just evaluation results for tenure and promotion</p> <p>Publish evaluation results</p> <p>Look at the results over time and consider trajectories, not simply a moment in time</p> <p>Research students' ability to rate/discriminate across a scale</p> <p>Have an instrument or program that can analyze/correlate</p> <p>Encourage student participation by having instructors promote the evaluation</p>

5 DETAILED SUMMARIES

5.1 BEEDIE SCHOOL OF BUSINESS

CURRENT INSTRUMENT

The current teaching and course evaluation system asks students about teaching but provides no context – it doesn't distinguish between different types of classes/courses: small and large classes; seminar vs. lecture courses, first- vs. fourth-year courses; or experiential vs. traditional approaches. It also doesn't account for the use of technology.

The group was split in its opinion of the current instrument – some don't have a problem with it and are quite happy to continue using it; while others think that it's outdated and needs to be revised according to more up-to-date research. One participant pointed out that *“we have no evidence that the teaching and course evaluation is working well, but we have some evidence of problems.”*

Mostly the instrument is used to provide data for tenure and promotion. Though, the scales used to measure performance are inconsistent which yields results that aren't transparent for students and teachers.

USE OF DATA

The evaluation process is *“primarily about evaluating instructors.”* Typically the data is used for tenure and promotion or for making hiring decisions about new lecturers. *“The results are most important for non-tenure-track (teaching) faculty...if they don't stink as teachers, they will get promoted as long as their research is good.”* *“Another shortcoming is that instructors aren't rated on their own merit – they are compared to others in their cohort.”*

While this process is mainly focused on summative data, there are many cases where instructors consider the comments for formative purposes. Some instructors collect their own data early in the term to improve their own courses and others pay close attention to how the data combines with written comments. In this way individual instructors take the initiative to find trends, identify contradictions and predict what needs to change. Even a comment like there's *“too much colour in your PowerPoint slides”* can be helpful and one participant noted responding to feedback like this.

Other individuals that use the data include the Associate Dean and area coordinators (like department chairs) who use the data to assess instructors.

CONCERNS

Transparency and perception of how data used is a concern: *“The perception that results matter is false.”* It's also widely believed *“that demanding teachers get worse evaluations.”* Not everyone has an accurate understanding of how courses are evaluated because the current instrument just assesses teaching, not courses.

Access to data is an issue – it's not available to all stakeholders (e.g. course coordinators), instructors have to wait a long time to get the results, and sometimes results are just filed away and only a few instructors request to see it.

Another issue is that *“students don't know how to do evaluations...some come from a reactive perspective.”* Instructors don't understand how *“students could possibly know what good teaching is, given that they have only short-term exposure to teaching.”*

Several participants also stated that timing is a concern, too. Evaluations could be done at more appropriate times during the semester. Right now, the evaluation is conducted during the last week of the semester when students have high stress levels and before they know what the course outcomes are (i.e. what their final grade is).

Finally, concerns were discussed about the possibility of sharing data with students:

- “Public dissemination involves a shaming factor.”
- “The culture of shaming doesn't work.”
- “When you announce results to students, you are announcing them to the world.”

SUGGESTED IMPROVEMENTS

In order to effectively revise the evaluation process the institution *“needs to address a fundamental question: What problem are you trying to fix?”* This implies taking a holistic approach – *“look at how everything around the evaluation tool works...because everything is connected...the tool won’t matter if the context or the current evaluation system is broken.”*

Meaningful data – that has context and is relevant – along with a more consistent approach to tenure and promotion is also required. Help faculty and TPC chairs to take a more standardized approach to how they use data by working to bring them *“on board with how evaluations are done...proper use requires education (training) beyond just a manual that they have to read.”* Further to that the entire TPC process needs to fundamentally value teaching.

“Those, like tenured professors, who are not ‘at risk’ don’t care about teaching evaluations...if the data isn’t being used meaningfully, then it shouldn’t be collected; if it is used only for teaching faculty, then it shouldn’t be collected for research faculty.”

Another key recommendation was to provide plenty of communication and instruction about the evaluation process to both instructors and students. Being able to trust the process is important for instructors – they need to be supported for taking initiative to welcome student feedback. Starting with the TPC process, instructors want to know that teaching is valued – that a well-rounded approach is taken to evaluating each individual instructor. For example, including peer evaluations, allowing instructors to select their own questions, and creating opportunities to conduct mid-term evaluations would help instructors be more interested in receiving feedback.

Students also need to know how evaluations are used:

- “The data would be better if students took the process seriously.”
- “Some think the process is a joke until I tell them that this is the only way I’m assessed for raises, etc.”
- “When I tell them that [evaluations] make a difference, some are surprised.”
- “Students [think] that the process doesn’t make a difference because they see bad tenured teachers coming back again.”

One suggestion offered was to provide ongoing instruction to students about the importance of the evaluations. This kind of messaging would also need to be reinforced by instructors each time they introduce evaluations.

It’s also important to design and structure questions for the instrument in a way that would provide a better experience for students and better data for departments/instructors:

- Ask questions that instructors are not already asking
- Rather than placing the most important questions at the end, place them in order of importance
- Ask more objective questions that will produce concrete results (e.g. did course readings match assignments?).

Finally, the group thought that the timing and instrument used for evaluations is crucial because this can impact the results. One participant recalled that even when she gave students in-class time to do evaluations online, half did not complete them; in the next class she gave them paper forms and the response rate was much higher.

5.2 FACULTY OF APPLIED SCIENCES

CURRENT INSTRUMENT

The group shared the various evaluation methods used in their departments – some use a form that includes an introductory paragraph that explains what evaluation results are used for, the engineering department commented that they've never used an online instrument – always paper based forms with bubble and open-answer formats.

One participant comment that the SFU forms are “*useless*” and has created her own surveys using WebCT, with an emphasis on short-answer questions about likes, dislikes, methods, etc.

USE OF DATA

Both comments and numerical data are used for making tenure and promotion decisions. Though, it's felt that “*results are all anecdotal...there is always debate about the value of a number to quantify teaching effectiveness.*” There's definitely a lack of context for the data being considered.

Evaluations are also being used to improve teaching and curriculum. Instructors are typically interested in the comments that students make and sometimes they even conduct their own evaluations to get student feedback. One instructor said: “*I tell students I don't care about the bubble questions; I want their comments because narrative information tells me more.*”

CONCERNS

Some have focused their concerns on the way that data is being used for tenure and promotion – that it's being used out of context or in a “*simplistic*” fashion. Other measures, such as summaries of professional development activities should be more effectively incorporated into this process.

Related to this, instructors mentioned some items that can negatively impact the results:

- External factors beyond the instructor's control
 - Students don't bother to complete the form because “*everything was good,*” suggesting that negative perspectives tend to be overrepresented in the data
 - Class attendance affects the ability to get students to all complete the evaluation
 - Students don't know their final grade when they do the evaluation
 - Students may be concerned about anonymity with online forms
 - No transfer of results back to students which results in students using “RateMyProf”-type sites
-

SUGGESTED IMPROVEMENTS

The group agreed that the best approach is to gather formative and summative data at the same time, preferably online. Their requirements for an evaluation instrument that could do this include:

- Have the option to collect feedback on programs
- The ability to ask questions after each class
- Use a question bank
- Use core questions combined with customizable options
- Ask open-ended questions
- Have a tool that can compile results for sharing and provide context (i.e. demographic data, GPA)
- Indicate to students and instructors who haven't completed the evaluations.

Other ways of getting feedback could also include:

- Conducting an alumni survey (i.e. find out what students think about a course a few years down the road, maybe after they've started working)
- Getting feedback on courses from other instructors.

Suggestions were made about making the use of data for tenure and promotion more consistent. *"We need to go beyond providing user guidelines; maybe doing training sessions with directors or TPC chairs. They need clear guidance on how to use teaching and course evaluation results."* *"The idea of briefing TPC on the use of evaluations and the ground rules could standardize the process."*

Providing incentives to students was mentioned as a strategy, too. Offering bonus marks (e.g. 1-2%) or withholding a final grade could be plenty of encouragement for students to participate.

While faculty are generally not in favour of sharing results with students, there were some in this group who thought this is a good approach:

- "Certainly we need to balance and filter the data, but we need to show students that they are being listened to
 - "Can we present the results to show students and faculty how feedback transformed courses?"
-

5.3 FACULTY OF APPLIED SCIENCES (SCHOOL OF COMPUTING SCIENCE)

CURRENT INSTRUMENT

The existing instrument gathers student feedback on teaching and is used for tenure and promotion. The evaluation is paper based: *"It's always been done that way...it's archaic but keeps it consistent."*

Some mentioned that the instrument is too narrow and doesn't ask the right questions. However, this approach continues to be used because it meets SFU policy on course evaluation and also because *"we can never agree on a model."*

This has led some instructors to ask their own questions throughout the term"

- "I use formative evaluations more. I put questions in the final for students to evaluate the course content."
- "I don't doubt the wisdom in students. But the instrument is not helpful. I do my own surveys to improve my courses."

USE OF DATA

Data is almost exclusively used for teaching and promotion decision-making and *"not so much for the improvement of teaching."* Instructors view that it's not a bad indicator of teaching ability, but feel that it's definitely *"far from perfect."* One reason is because data is usually considered out of context (e.g. the size of a class).

Most instructors pay close attention to the written comments. Although, some commented:

- "Some questions are useless such as the punctuality question. Why should I care how students evaluate my ability to start and finish on time?"
- "Students are not always the best audience to answer the questions asked in course evaluations – questions knowledge, teach ability or expertise."

CONCERNS

General concerns about the evaluation process include:

- Students don't care about giving feedback which makes the response rate low
- There's a lack of accountability for students
- The process creates added workload for faculty and staff
- *"Sometimes I feel like I need to have discussion about the answers."*
- *"I have a moral concern with incentives...I don't like the idea of a student receiving a lower grade because he or she failed to fill out a course evaluation."*

Some concerns about questions used on the existing instrument are:

- All of the questions are too generic
 - There are too many characteristic-oriented questions
 - Questions are not worded so that all students can interpret them (considering that approximately 20% of students are international).
-

SUGGESTED IMPROVEMENTS

Recommendations from the group included:

- Evaluate departments and programs (e.g. *"Are we offering the right courses?"*)
 - Create a set of best practices for the institution
 - Have an online evaluation instrument that's easy to use (e.g. on smart phones, perhaps)
 - Use a 5-point scale (instead of 4)
 - Have an instrument than can be tailored for each department, course and level
 - Use specific questions about courses (e.g. labs, assignments). *"Questions like 'this course is too easy' or 'too difficult' are not the right ones to ask. How do they know if it's easy or difficult?"*
 - Explain to students the reasons for conducting evaluations
 - Offer incentives like bonus marks (1%) or early registration for filling out evaluations
 - Make evaluations compulsory – don't release grades unless evaluations have been completed
 - Employ different types of evaluations depending on how the results need to be used (e.g. one evaluation for TPC and another for teaching improvement). *"I have no interest in how difficult they find the course. Instead I want to know if the assignments and labs were useful. If they weren't, then tell me what was wrong with them. There should be two sets of questions or two instruments: one for TPC and one for course improvement."*
 - Include other evaluation methods and perspectives
 - Conduct the evaluation at strategic times during the semester
 - Allow instructors to conduct their own evaluations.
-

5.4 FACULTY OF ARTS AND SOCIAL SCIENCES

CURRENT INSTRUMENT

The group felt that the evaluation process is imposed – *“we do evaluations because we have to.”* Even still, participants indicated that most instructors take evaluations seriously. *“Teaching is hard and serious so evaluation has a complex set of considerations.”* Many instructors are doing formative evaluations on their own as a way to improve their teaching; however, even though pedagogy may be discussed with students, the student evaluation should not be the only method by which an instructor is evaluated.

Many feel that the current instrument is not effective at capturing accurate results. In fact, some felt that all forms of evaluation will have the same problem: that it’s not about improving teaching. However, open comments from students may yield some useful information and be helpful in a formative way.

The current instrument is also limited by the fact that it doesn’t help students separate learning outcomes from teaching performance – linking these together has *“terrible implications”* for the way people teach. The group noted that curriculum and desired learning outcomes need to be clearly defined in order for students to know what to evaluate so that faculty can identify any gaps in students’ understanding.

USE OF DATA

Most often student evaluation results are *“used pragmatically”* for tenure and promotion, salary review or for intervention when an instructor gets a lower than average score. Most of the time, *“instructors are given summary sheets with numbers [and] individual comments.”* However, it seems *“departments can be very different in how they use the results”* – only using them selectively and contingently. Some departments report average scores; some conduct mid-term evaluations; sometimes professors are evaluated as a TA; and a few instructors collect feedback after each class. Unless there seems to be any anomalies in the data, most departments and instructors focus on comments from students over the bubble responses.

The evaluation may or not may not alter teaching tremendously, especially since the primary audience of evaluation results is the TPC. Evaluations are seen by many as a tool to *“find flaws and deny tenure.”* Some commented that teaching evaluations have been given an *“unconscionable weight”* because they continue to be used as the only basis for making hiring and pay decisions. This puts a great deal of stress on some faculty while others have their teaching evaluations treated more informally.

Some reported: *“Summative evaluation should be about the course, curriculum, class, learning outcomes; while formative should be about teaching, so teachers can change their teaching.”* Instructors are more interested in students’ comments and ongoing forms of evaluation that can be used in a formative way to help improve teaching. One participant noted that *“many instructors do formative evaluations all the way through the semester...there are lots of feedback loops that aren’t part of the formal process but are very important in improving teaching.”*

There are also some cases where evaluation responses are not useful, for example:

- “I don’t want to know that there’s too much reading or that deadlines are too close to the exams.”
- “Certain evaluation results aren’t useful in practice. For example, asking how passionate or enthusiastic an instructor is – how can this rating be usefully employed? What do the results mean?”
- “Students evaluate from the gut without being qualified...much of what they say is of minimal or no value.”

CONCERNS

Instructors are feeling very pressured to share their evaluation data with students. However, this approach is met with great trepidation because the instrument hasn't proven effective, student responses often don't seem accurate and departments use data inconsistently.

The approach taken in some departments has caused many to distrust the evaluation process:

- "Teaching does not improve."
- "[Our] department only looks for the good things that come out in reports."
- "People have been denied tenure based only on the numbers [and] no other sources."
- "It's important for administration; it's not about teaching, but about things like accreditation."
- "I find it hard to imagine any new system that would be an improvement...it will only lead to the same morass."

There's also a wide array of concerns related to how students complete evaluations:

- "Students don't know how to evaluate. Lots of student comments are not helpful, useless, and a waste of time."
- Some students don't bother coming to class but they use the survey to express very strong opinions.
- "Some courses don't require attendance [so it's] difficult to get everyone to complete the evaluation. And often times the higher the [response] rate the lower the score."
- Evaluations don't accurately capture information about students' participation in the course or their relationship with the instructor. *"No matter how much effort you put into it, there are still some students say who you are never available."*
- Students don't read the syllabus and they don't understand how the evaluation process works – it's felt that students just want the *"bad teachers"* fired.
- The evaluation can be used by students as a tool for disseminating awful or cruel remarks. "There are biased and hateful evaluations, which will not benefit teaching practice." "Junior faculty may be bullied and intimidated."
- Timing, external influences and other people may impact how students complete the evaluation (e.g. students may take out their frustration about the strike during the evaluation).
- Students generally have no idea of course design or pedagogical structure.
- "Evaluations among students have a lot of dissonance; it's hard to believe they are from the same class."
- "Students are partly responsible for the classroom learning environment, but instructors are the only ones being evaluated."

Other concerns were about the instrument itself and the accuracy of the data that it produces:

- The instrument seems challenging to use which could have an impact on response rates. This, in turn, may mean that there's not a solid representation of data available: *"If only 70 percent answered a question, how representative is it?"*
- Some courses are just destined to receive negative evaluations because of things like time of day, sensitive course content, course accessibility, or optional versus mandatory courses
- Faculties all use the data for different reasons – some look at the numbers for teaching awards and study leave, other scientific departments use evaluation as a weapon against faculty to deny what they are working toward.

Some also commented that resistance to the learning outcomes could transfer into student evaluations of teaching and courses. The suggestion was made that learning outcomes should be settled first before evaluating teaching.

Incorporating a learner-centred approach to instruction was also raised as a concern. Many faculty members are onside with this approach, but students are viewed as a *"huge hurdle"* to implementing it because they're often extremely resistant to different instructional methodologies.

SUGGESTED IMPROVEMENTS

Everyone agreed the evaluation can make students feel included by giving them a valuable role to play, but sometimes their expectations are unfounded. For example, *“if they have terrible experiences, they want [their] instructor fired.”* More effective results would come from better communication with students – clearly letting them know what the expectations are for the process. One instructor shared: *“I let the class know how I teach, how evaluations are used and how I will respond to student feedback, but I maintain the right to teach the class.”*

Instead of forcing students to complete the evaluation at the end of class, positive results could come from providing the option for students to complete it online at their leisure. Having the opportunity to officially respond to student feedback given through mid-term evaluations or informal discussions throughout the term would also help students feel involved and engaged.

Gaining popularity with students doesn't necessarily mean that an instructor is an effective teacher. The university needs to recognize good teaching using methods other than just the evaluation; otherwise *“there is no incentive to be a good teacher.”* One way to do this is by using better questions – more detailed questions or ones that are appropriate and useful. Instead of asking about instructor enthusiasm, ask: *“Did you take this class specifically because of this professor?”*

Suggestions were made about how the evaluation survey could be structured:

- Offer options for small and large classes
- Collect information about the student's study behaviours (e.g. how many hours a student devotes to study) and why they may not be coming to class
- Create departmental-specific questions
- Use more specific questions that can help in a formative way (e.g. what was the most helpful part of the course? What was the least helpful part of the course? Was the syllabus clear?)
- Get feedback on instructor behaviour and classroom environment/space
- Have the option to include open-ended questions, not just the standard survey questions
- Explain on the evaluation form that inappropriate comments will not be passed on to the instructor
- Ask students how they would design the syllabus.

Getting data that is useful for both summative and formative purposes is really at the core of what the group shared. One participant mentioned that asking students for reflective papers and *“letters to successors”* was a useful feedback mechanism that is more meaningful for both students and instructors: *“Students are experts in their own experience; evaluations should ask about this and separate those inquiries from performance evaluations.”*

Other suggestions for improving the data included averaging course evaluation results for instructors, discarding extremely high / low scores and holding students accountable for unnecessary hurtful comments.

Offering a safe, non-threatening environment for evaluation from peers or education consultants could help foster more interest with instructors about making teaching improvements. One instructor said: *“I learn only 20% from survey results, but [when] consultants come in my class [they] add more to my teaching.”* Another commented: *“We all want to teach well. That's why colleagues' in-class observation is useful. We get information that we can never get from surveys or students but from people who know pedagogy.”* Further to this approach, an evaluation support unit or consultative group with teaching experience would prove to be an excellent resource for instructors.

Some other suggestions included:

- Considering external influences, class size or other unique factors that could impact results
- A complete separation of formative and summative evaluations (i.e. formative only for teaching evaluations and summative only for course evaluations).

Finally, the group recommended giving consideration to the *“inter-culturalization and internationalization of teaching.”* Many students are here from different countries and they bring with them expectations associated with different pedagogies. *“Students may think the teacher is not doing what they are supposed to do based on different expectations.”* Summative evaluation does not help address this – the best way is to give students the opportunity to *“talk about their experiences.”*

5.5 FACULTY OF EDUCATION

CURRENT INSTRUMENT

When asked how student evaluations of teaching and courses are used in their faculty, participants initially responded: *“Badly!”, “It’s hardly used because of the instrument.”*

Other comments indicated the reliability of the existing instrument is unverified which makes it challenging to use the data in a constructive way. *“Shoddy”* was used to describe the current data interpretation process – there’s not enough time or any best practices to interpret the data in a more rigorous fashion. This includes not being able to analyse the answers to open-ended questions for qualitative value.

Finally, issues like formatting and inflexibility with presentation of questions were also mentioned as challenges with the existing instrument.

USE OF DATA

When it comes to using the data, everyone agreed that a consistent method for data analysis needs to be a core component of any future evaluation instrument. *“Validity and inference are key”* – a validated instrument is required in order for data to be interpreted and used properly.

“The data is almost always used individually, but we need to look at data collectively, what’s working, what’s not working, looking for patterns.” While it would be *“interesting to look at patterns across instructors for a given course,”* it was felt by some that data shouldn’t be compared across departments – *“it’s a destructive way to use the data.”* Others felt that if the data is used validly and appropriately for comparison, then it might be okay.

“Clear guidelines on how to interpret the data are essential.” Systematic use of data that follows a set of standards or guidelines is necessary. Standardizing how data is collected or compared will really improve its accuracy. Right now, data is not analysed systematically – especially qualitative data, like answers to open-ended questions. Often numbers are just averaged and sometimes written comments can seem contradictory or even get omitted. It was also noted that TPCs routinely skip using quantitative data because it is undifferentiated. *“We need better questions that will differentiate”* – for example, using longer rating scales like 1 – 10.

“The ultimate test is whether the evaluation data is actually useful for making decisions.” When it comes to making serious decisions (e.g. firing), it’s imperative that the data is accurate. There could be situations where anomalies are found, like when an instructor receives a poor evaluation from an extremely large class – or outlier. Some felt that large class sizes definitely get lower ratings. Using data to deal with these kinds of non-typical situations makes it *“important to define if or when there is an obligation to take action.”*

CONCERNS

The group mentioned it’s important for students to understand the purpose behind the evaluation. Creating an awareness of the opportunity they have to effect change through the evaluation process could help foster a sense of responsibility towards it. *“The evaluation is an opportunity to ask students to describe their experience...students are more engaged when they know that their feedback will be applied.”*

Misunderstanding or misinterpreting the wording of items and temporal features of the evaluation make it less valuable for students and departments alike.

What’s at the heart of the evaluation process for most instructors is striving for teaching excellence. *“Some individuals do look at it to help improve their teaching.”* In some cases, the evaluation process is perceived as an aid to making improvements to teaching. However, because the evaluations are also used for tenure, promotion and review of faculty, it’s challenging for instructors to remain unbiased to the process. For example, instructors are able to create their own files for TPC review which means that they will often opt to only include “kudos” comments to sway the TPC’s opinion.

SUGGESTED IMPROVEMENTS

“We need to identify teaching excellence.” Feeling accurately represented is critical for instructors. Within the evaluation framework, it’s important to make a clear distinction between instructor characteristics and teaching effectiveness. Also, there needs to be opportunity for instructors to identify potential courses as outliers – classes that are extremely large or that cover difficult or sensitive content. The challenge for departments will be striking the reasonable balance between listening to what instructors say about a class versus responding to student input and taking action when necessary. Having solid, reliable data will make this easier.

A key recommendation from the group was to use question selection that would create a balance between qualitative and quantitative data – focusing more on learning, teaching and student satisfaction. These kinds of questions would need to offer flexibility, be clear for students to interpret, and provide reliable data. Using response theory would help create a bank of pre-defined questions that instructors or administrators could select from that would generate reliable data.

Taking this kind of standardized approach to using questions would also make it easier to use data for both formative and summative purposes. For example, offering the option for instructors to add their own items to a survey so that they can privately review student feedback and implement changes to their course delivery.

Flexibility in the development of questions would also include:

- Creating questions for different courses, fields or instructors
- Allowing departments and instructors to select and create their own questions
- Offering open-ended questions.

Finally, the group recommended creating a guide of best practices that could be used institution-wide for conducting evaluations and interpreting data consistently. A goal for the future instrument and its guidelines: *“people find the instrument / document useful.”* Some notes for the guide:

- “When changes are first implemented, evaluations tend to be lower”
- “When an instructor tries new things, ratings often drop”
- Surveys “shouldn’t be too long”
- “We want quality, not quantity”.

A more automated, online instrument could create *“lots of opportunity to contextualize data.”* In order for data to be put data into context, it needs to be analysed and compared. Some recommendations for improving evaluation data included:

- Compare and analyse data over time; before and after course completion (e.g. mid-term evaluations); or across instructors or departments
- Use available software to assist with data interpretation
- Triangulate multiple data sets
- Include more demographics like course descriptions or number of hours required
- Work with institutional experts to help with best practices for data interpretation.

5.6 FACULTY OF HEALTH SCIENCES

CURRENT INSTRUMENT

An online evaluation system is used to gather student feedback on teaching. Response data is used by instructors for making improvements and by committees responsible for tenure, promotion and hiring.

Moving to an online instrument has caused a dip in response rates – currently the average response rate is around 50%. An incentive is offered to students for participating: if the participation rate of the class is over 65%, the entire class is entered into a draw for a prize.

Instructors are sure to encourage that students complete the evaluations. Sometimes, for smaller classes, faculty members take their students to a computer lab to fill out the course evaluation.

Participants commented that the current instrument is too narrow – that it doesn't ask the right questions. There's also dissatisfaction concerning the fact that the online participation rate is lower than paper. *"I personally dislike the system. The one-size-fits-all model is not working. You can't evaluate a 500-person class the same way as a boutique course."*

USE OF DATA

Participants felt that summative data is only *"useful from an administrative point of view."* In fact, some commented that they weren't even certain the data is reviewed very closely: *"I don't think the Faculty of Health Sciences views the course evaluations in detail, otherwise they wouldn't keep hiring the really horrible teachers."*

Instructors prefer to have student responses to open-ended questions because these can provide useful information for making insightful improvements to teaching and course content. *"I do my own course evaluations with more open-ended questions. For example, I ask students to describe an 'aha moment' in the class."*

CONCERNS

Some key concerns raised by the group include:

- Students don't care about giving feedback
 - Institutional factors that have potential to negatively impact evaluations
 - The relationship between instructors' job satisfaction and course evaluations
 - That teachers are being hired despite bad reviews
 - That students can make derogatory or extremely negative comments about instructors
 - That the data isn't accurate because students are inclined to only complete evaluations if they enjoyed the course or if they felt it was really horrible.
-

SUGGESTED IMPROVEMENTS

Primarily suggestions were offered to help make the process of more value to students. Explaining the reasons for the evaluation, clearly defining expectations, offering incentives and making the results public were highlighted as options:

- *“Consider how students of different class, gender, race, etc. experience their courses. Is there difference? If so, how can we address issues?”*
- *“When students are registering for courses they need to know: What are the course requirements? What's the instructor's teaching style? How many hours should be spent on course assignments? If you make the results of course evaluations public then they can access this kind of information.”*

Further to this, a flexible instrument is required that can be tailored for each department, course, level and instructor. This would make it possible to ask specific questions about courses details like presentation formatting, labs, assignments or items that were specifically valuable/useful to student learning. Instructors felt strongly that forms should be designed to provide qualitative feedback by asking the right questions – open-ended, specific questions.

Ensuring that instructors have the right tools and resources is also essential: *“Things in the classroom should flow smoothly...students get frustrated and take it out on instructors in course evaluation, while a failing computer or microphone is not the fault of the instructor.”*

Peer evaluation and professional help with teaching (e.g. from the Teaching and Learning Centre) are welcome opportunities for improving teaching. This approach could help instructors to focus on the strategies of teaching: *“We are so content-oriented [that we] don't think about teaching strategy and pedagogy.”*

Some also felt that the completing evaluations should be mandatory for students to receive their grades – this could help raise the response rate for some classes. Additionally, the institution should:

- Develop a zero-tolerance policy for evaluations with inappropriate, derogatory comments
 - Create an annual survey for faculties on *“barriers to teaching”*.
-

5.7 FACULTY OF SCIENCE

CURRENT INSTRUMENT

The existing instrument is used to gather student feedback on teaching. The data is used for making decisions about tenure and promotion and sometimes instructors use it to make teaching and course improvements.

Many feel that the existing instrument is too narrow because it doesn't ask the right questions – too much emphasis is placed on close-ended questions and the open-ended questions used are too general. Further to this, the instrument is used too late in the semester to be effective – *“students care less at the end of the semester.”* An instructor who conducted their own mid-term evaluation said the responses were useful *“because [the students] have a stake in the course and care about giving comments.”*

USE OF DATA

Currently, evaluation data is used in a very limited way, mostly for tenure and promotion decision-making. One individual said: *“We can't only focus on numbers for TPC – the feedback we get from students is useful.”* The group agreed that numbers matter but they are not necessarily a reliable indicator of teaching quality.

The context of the data is also important – *“faculty consider the evaluation in terms of tenure and promotion, while students look at it differently.”*

Having data that is both flexible and accessible is also challenging:

- “There are lots of policy considerations in place.”
- “I find that the evaluation documents are too long to read.”

CONCERNS

The ethical and appropriate use of data was commented on as an issue. Currently there are no guidelines that govern who has access to the data, where it's stored or how it will be used. One participant commented: *“Hiring sessionals based on evaluation scores is problematic. For example, a sessional gives a hard exam, gets a low rating and doesn't get rehired.”*

Other issues with the existing instrument include the:

- Types of questions used
 - Added workload for faculty and staff
 - Fact that students don't care about giving feedback and faculty aren't interested in receiving it.
-

SUGGESTED IMPROVEMENTS

"You can't improve without good measurement. The question to be asked is: 'Are we asking the right questions?'" Suggestions for creating questions that would help improve teaching included:

- Ask specific questions about what the student learned from the course like key terms, theories, topics or skills
- Incorporate discipline-specific questions
- Include questions on organization: *"Is the instructor punctual? Organized?"*
- Have more open-ended questions that are specific: *"If you could change something, what would it be?"*

To that end, a flexible instrument is required so that surveys can be tailored for each department, course and level (1st year 2nd year, etc.).

Some comments focused on making the evaluation experience better for students. Explaining the reason for the evaluation (e.g. *"it's not a popularity contest"*), conducting mid-term evaluations and making the results available as soon as possible would all help students feel empowered. *"We need to consider that students have a sense of entitlement when they pay for their education."* Offering incentives like additional marks (e.g. 1%) for completing an evaluation would also encourage student participation.

Instructors also see the need to have an instrument that aligns with what they do. For starters, they suggested the evaluation process should be uncomplicated and streamlined. For example, efficiencies could be gained by having an online system that interfaced with Canvas. They'd also like to know what happens to the data as well – so that they have opportunity to respond to low scores and explain context.

Incorporating observation and feedback from educational consultants would help induce instructors' confidence in the evaluation process: *"I would pick a pedagogy expert's opinion about my teaching over student evaluations."*

5.8 SFU FACULTY ASSOCIATION

CURRENT INSTRUMENT

This group indicated that the existing instruments in their respective faculties place more emphasis on tenure and promotion than on enhancing teaching and learning. Some instruments are not considered obsolete and are valued, especially ones that are paper based because they only permit students who come to class to complete the evaluation. Though, one participant said the instrument used in their faculty is “*completely useless*” due to the lack of clarity about what it's evaluating.

USE OF DATA

Here are some points made by participants about how data is used:

- The data is not used to support faculty members
- The current instrument does not do a good job of measuring lab performance.
- “*The current data is useless...in our department, we don't use it at all.*”
- The timing of evaluations is important: “*formative evaluations need to be done early in the semester.*”

CONCERNS

Most concerns relate to how data is used – “*there's no clear idea of what we are looking for in evaluations.*” Departments are not consulting “*experts*” to help manage their instruments or analyse the results.

When it relates to tenure and promotion, the evaluation process isn't considered a means for positive use – it doesn't help advance careers. It's widely thought that TPCs “*always extract the negative information*” from evaluations. This has led to many faculty members not caring about the feedback they receive and are, therefore, not interested in taking an active role in effecting changes the system.

Thinking about potential of an online evaluation system triggered thoughts about:

- Lack of security
- Participation from students who don't attend class. “*With paper based evaluations you can get the opinion of those who regularly come to class and have some knowledge about the course. But with an online system, even people who do not come to class can evaluate.*”
- Students colluding or getting together to fill out evaluations

For some, moving to a new system in general causes concern:

- “Who decided that the current system is obsolete?”
 - “Who's designing the questions for evaluations?”
 - “What is the design process?”
-

SUGGESTED IMPROVEMENTS

The top three needs highlighted for student evaluations were:

- Student learning
- Tenure and promotion
- Flexible evaluations that measure different aspects of performance (e.g. labs) and allows faculty members to conduct formative evaluations (e.g. early or mid-semester evaluations)

One participant commented: *“It doesn't matter what instrument you are using...what's important is how the data is being used. For example, if you are interested in improving teaching and learning, you should be interested in formative evaluation.”*

The following recommendations were made for moving forward:

- Define the purpose of student evaluations of teaching and courses for Senate
 - Evaluate the merits of online versus paper evaluations – don't assume that paper based is obsolete
 - Define how evaluation data should be used (e.g. for what purpose)
 - Design measurements of performance outside the lecture setting (e.g. labs)
 - Add *“context”* (i.e. demographic information) to evaluation reports
-

5.9 TPC CHAIRS

CURRENT INSTRUMENT

Typically the Tenure and Promotion Committee (TPC) uses teaching and course evaluation results to review salaries or make decisions about instructor merit, tenure or promotion.

One participant commented that the TPC used question 22 as a red-flag indicator for problems with instructors. This question was something like: *“How would you evaluate the effectiveness of this person’s teaching?”* One participant mentioned that in Education, they use a form with just three questions: *“What are the best and worst features of the course? What would you change? How would you rate your instructor (based on a rating scale)?”*

Comparison of results for similar courses (classified by size, format, etc.) and student comments are also important factors. It was interesting to note that the online survey provided more comments to work with than bubble forms – this was useful because the TPC looks at a wider range of factors for tenure and promotion.

For salary assessment, TPC members look at overall evaluation results for instructors. When there are extreme results (good or bad), they drill down to see what students think is very well done or very poorly done (e.g., attitudes toward students, communication skills, etc.). In general, students’ views correlate with TPC members’ own experiences with particular individuals.

USE OF DATA

“The results are not an indication of whether a person is a good teacher – they are a factor, but there’s a lot more.”

The way teaching and course evaluation results are used depends on the department rather than the faculty. Any inconsistencies could be related to the lack of training received by TPC chairs.

CONCERNS

The following are some issues and concerns raised by the group:

- Lower response rates and instructor ratings using online instrument – instructors want to return to paper based system
 - *“Faculty members do not respect the evaluations and if they convey this attitude to students, students won’t take them seriously either.”*
 - Giving the evaluations out at the end of a class can show a lack of value which impacts responses
 - Validity and reliability of the instrument and data is poor – cannot make effective decisions based on it
 - Peers without expertise shouldn’t evaluate peers
 - Cannot force new forms of evaluation (e.g. classroom visits) on instructors
 - Students are not qualified to comment on some aspects (e.g. *“Is your instructor knowledgeable about the topic?”*)
 - No standards for extracting and analysing data / student comments
 - The context of questions on the survey impact the data
 - Evaluation data is often used *“sloppily”*
-

SUGGESTED IMPROVEMENTS

The following were improvements suggested by the group:

- Determine what students should be evaluating and consider why particular questions are being asked
 - Instead of focusing on whether an instrument is valid or not, give more attention making sure interpretations of data are valid
 - Use an online instrument (potentially interfaced with Canvas) that has a question bank but also allows some latitude for departments to create questions
 - Use a clear standard for comparison in questions – for example, *“if you say ‘compared to other courses,’ you need to specify which courses”*
 - Consider question sequence to avoid bias
 - Develop guidelines on what questions to use and how to interpret data
 - Use a system that doesn’t increase department workload
 - Change departmental guidelines to include other forms of evaluation (e.g. peer reviews)
 - Consider more data than just evaluation results for tenure and promotion
 - Publish evaluation results
 - Look at the results over time and consider trajectories, not simply a moment in time
 - Do some psychometric studies on students’ ability to rate/discriminate across a scale
 - Provide guidance and training on statistical interpretation and data analysis
 - Have an instrument that can track student demographics and make correlations using this data
 - Use data analysis programs that can suggest accuracy of results by particular profiles / contexts
 - Encourage student participation by having instructors promote the evaluation
-

5.10 STUDENTS

STUDENT PERCEPTION

Students commented that they *“did not think the course evaluations were actually used.”* In place of evidence, rumours have created the notion that evaluations are simply used for evaluating sessionals / TAs and making decisions about rehiring them. If evaluations were used for making improvements to teaching or revising curriculum, students indicate that they *“would like to be able to give feedback on other issues, such as pre-requisites.”*

Overall, the experience with evaluations depends on the class and the instructor. There are some instructors who encourage that students complete evaluations and it's obvious that they take feedback seriously; while others show little or no interest in the process at all. Some small classes are having positive outcomes using qualitative evaluations that focus more on what students are thinking – one student commented that this *“feels more democratic and instructors take it seriously.”*

The question was raised, *“should evaluations be optional or mandatory?”* Because the existing process is not mandatory, the sample data is probably very limited.

REASONS FOR LOW EVALUATIONS

Behaviour or triggers that could cause students to provide a low evaluation include:

- No evidence of changed teaching behaviour despite student expectations
- Instructors that are rude or negative to students
- Classes that are boring or seem like there's no point in attending (e.g. when an instructor just reads off slides)
- Instructor's lack of interest in the course or students
- Instructors that are unclear or don't communicate well
- Unapproachable instructors or instructors who lack attention during office hours
- Instructor acts like it's a chore to come to class
- Instructors that aren't competent or equipped to teach concepts
- Using dated materials
- Instructors with extremely narrow/rigid viewpoints and who are not open to other ideas.

REASONS FOR HIGH EVALUATIONS

Behaviour or triggers that could cause students to provide a high evaluation include:

- Very original class format/delivery
 - Instructors who know their material in great depth and can explain topics well
 - Instructor demonstrates a passion for subject
 - Courses that effectively use media and technology
 - Instructors who *“go beyond call of duty”* and are flexible for the needs of students (e.g. extending office hours when needed)
 - Instructors who give a lot of feedback to help students improve
 - Instructors who clear communicators – even when there may be a language barrier
 - Instructors who are positive, cheerful and nice.
-

CONCERNS

Student concerns about the teaching and course evaluation were varied. The majority of comments focused on how the evaluation process could impact course selection. Sometimes there's no choice but to take a course when it's offered, though, if students have the option, they'd like to be selective about which courses they take. Currently, students inform themselves by word of mouth and online forums (e.g. Rate My Professor), or by using course outlines. Having evaluation results available or an SFU-specific online forum for course discussion would help this process a great deal.

Other concerns were about the method and timing used for conducting evaluations. For example, the group held the opinion that *"students won't take time outside of class to fill [the evaluations] out"*. This could lead to skewed results if the evaluation forms are moved to an online format due to lower response rates. Sometimes students feel that the evaluation is completed too early – that there hasn't been enough time to evaluate the course fairly.

"The harder the course, the lower the evaluations" – some courses are just harder to give a high evaluation for because they're challenging. There's also a fear that tenured faculty don't want to make improvements to their teaching and, therefore, aren't interested in receiving feedback from students.

SUGGESTIONS FOR IMPROVEMENTS

Even if *"instructors make a promise to read them,"* students feel that the evaluation process doesn't make a difference. Since this is the case, it's important for the institution to establish credibility in the evaluation system. Expectations need to be clearly defined and communicated. Instructors need to be held accountable for making changes – *"it's a moral hazard when there are no accountability mechanisms."*

One example was given of an instructor who distributes cue cards every class to ask what was clear, what was not and when to hold office hours. Students felt that this example highlighted the importance of ongoing evaluation processes like:

- Introducing mid-term evaluations to allow opportunities for instructors to enhance their teaching
- Conducting pre and post student expectation surveys.

Other suggestions for improvement included:

- Look at common themes across departments or certain types of courses
 - Ask for feedback on exam formats/questions
 - Ask for feedback on departmental issues
 - Eliminate the GPA question and focus on more relevant biographical information such as learning style, study habits, language background or major
 - Ask for comments on the way the course was taught
 - Seek more feedback on course content
 - Split questions into sections: course content, teaching style, textbook.
-



SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

THE TEACHING AND COURSE EVALUATION PROJECT
FINAL REPORT

APPENDIX X: PROOF OF CONCEPT

NOVEMBER 26, 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

DOCUMENT CONTROL

DOCUMENT INFORMATION

	INFORMATION
Document Owner	<i>Corinne Pitre-Hayes</i>
Project /Organization Role	<i>Project Leader</i>

DOCUMENT HISTORY

<i>1.0</i>	<i>Nov 14, 2013</i>	<i>Detail of Proof of Concept moved to appendices</i>
<i>1.1</i>	<i>Nov 17, 2013</i>	<i>Revisions from core team review</i>
<i>1.2</i>	<i>Nov 26, 2013</i>	<i>Revisions from SCUTL feedback</i>

TABLE OF CONTENTS

1	OBJECTIVES	1
2	THE FRAMEWORK	1
3	PARTICIPANTS	2
4	THE PROCESS	2
5	RESPONSE RATE	3
6	VOLUNTEER INSTRUCTOR FEEDBACK.....	3
7	STUDENT FEEDBACK	4
8	CONCLUSION.....	4

1 OBJECTIVES

With the research, consultation with other institutions and engagement with the SFU community completed, the next step was a small scale demonstration of the emerging recommended approach. With the help of volunteer instructors, a proof of concept was conducted with a small number of actual student evaluations during the Summer Term of 2013. The objective of the proof of concept was to obtain pragmatic feedback from SFU instructors and students on a small scale trial of the approach. Table 1 provides a summary of the objectives of the proof of concept.

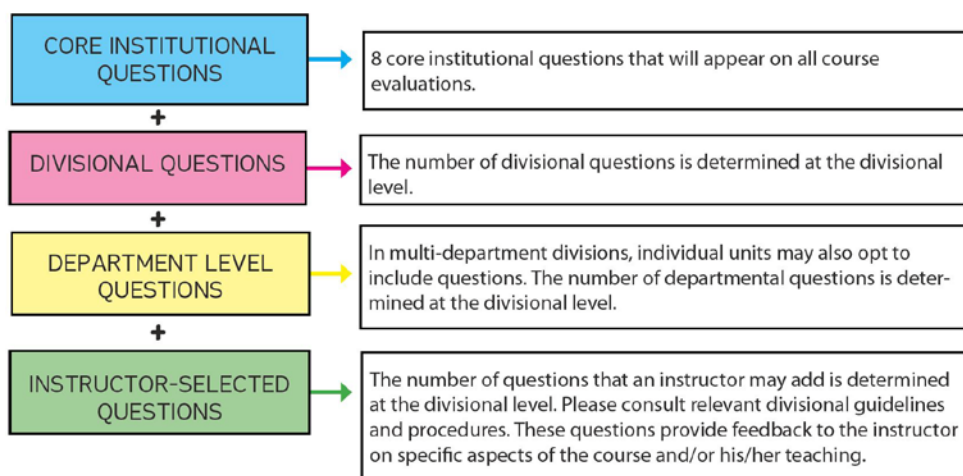
Table 1: Objectives of the Proof of Concept

Category	Feedback
Representative framework	<ul style="list-style-type: none"> • Sufficiently flexible? • Faculty, School, Department level questions? • Instructor questions for formative purposes?
Support	<ul style="list-style-type: none"> • Materials valuable? • Additional materials? • Other forms of support?
Reports	<ul style="list-style-type: none"> • Distributing the results?
Online versus Paper	<ul style="list-style-type: none"> • Response rates?
Process	<ul style="list-style-type: none"> • Any general comments, feedback, suggestions?

2 THE FRAMEWORK

A multi-level framework developed at the University of Toronto was utilized as the representative approach that best matched the key requirements highlighted in the project team's research and internal/external consultation. Figure 1 illustrates U of T's multi-level framework.

Figure 1: Multi-level Framework from the University of Toronto



Source: Step-by-Step Faculty Guide, University of Toronto

For the proof of concept, the University of Toronto framework was adapted to permit volunteer instructors to create their own specific questions at the instructor level. In total, up to 20 teaching and course evaluation questions were included as summarized in Table 2 below.

Table 2: Proof of Concept Questions

Framework level	Number of questions
Institution-wide	8
Faculty level	Up to 4
Department level	Up to 4
Instructor from question bank	Up to 4*
Open questions created by instructor	Up to 2*

* Total number instructor questions = 4

In addition to the 20 questions above, three questions related to the students' experience with the evaluation tool were also included at the end of each evaluation.

3 PARTICIPANTS

There were six faculties represented in the proof of concept. These included:

- Faculty of Applied Sciences
- Faculty of Arts and Social Sciences
- Beedie School of Business
- Faculty of Education
- Faculty of Health Sciences
- Faculty of Science.

A total of 14 volunteer instructors participated. There were 18 courses, including 14 lecture or seminar courses and four courses offered through the Centre for Online and Distance Education. The courses range in size from 13 to 329 students. A total of 1,329 invitations to complete evaluations were sent as part of this small scale demonstration. A compendium of support materials provided to participants is included in Appendix XI.

4 THE PROCESS

The proof of concept process consisted of the following eight steps:

- 1) Volunteer instructors asked to select questions from the question bank provided
- 2) Instructors invited to include up to two of their own open-ended questions
- 3) Instructors asked to encourage student participation
- 4) The student evaluation period opened via email, with periodic email reminders
- 5) Instructors asked to encourage participation from students who had not yet completed an evaluation
- 6) The evaluation period closed
- 7) Reports designed and distributed
- 8) Feedback sessions conducted with volunteer instructors, and follow up survey conducted with participating students.

A representative online tool was utilized for this small scale demonstration. The tool, BLUE, is an offering of a Canadian software company, eXplorance Inc. The tool met the requirements of the proof of concept, was recommended by other institutions, and the vendor was amenable to working with the project team on a demonstration basis.

Care was taken to ensure that the privacy rights of instructors and students were protected. The vendor signed a Protection of Privacy Schedule with the University. The proof of concept was in full compliance with University Policy | 10.08, Collection and Disclosure of Instructor and Course Evaluations. As a follow up to the project, the team will be working with the University Archivist/Coordinator of Information and Privacy to document this compliance in the newly created Privacy Impact Assessment form.

5 RESPONSE RATE

According to a summary of the literature (Rawn 2008) regarding online response rates published in 2010 by the Student Evaluation of Teaching (SEoT) Committee at the University of British Columbia, response rates for online student evaluation of teaching typically range from 30 to 60%. Published studies generally show response rates for online evaluations are 20 to 30% lower than those for paper based evaluations. However, there is evidence that online response rates increase over time. Of significant interest is that research consistently shows no meaningful differences in instructor ratings between online and paper based evaluations. The research also shows that students are more likely to provide qualitative comments when responding to online evaluations. Finally, studies show that better online response can be achieved through careful timing of email reminders and instructors indicating the value of the feedback to their students.

Overall, the response rate for the proof of concept was 72%. This was considerably higher than the typical range of 30-60% reported in the UBC summary. When response rates for lecture/seminar courses were considered separately the response rate was a very strong 82%. CODE courses, however, exhibited a low response rate of 27% which was quite similar to the rate experienced with the current system. One of the main differences between the proof of concept process for lecture/seminar courses versus CODE courses were the lecture/seminar courses included a direct appeal from the instructor to the students with regard to the value of their feedback, and in a number of cases the incentive of a small bonus mark. The stark difference in response rates appears to support the impact of instructor appeal in motivating students to complete evaluations. Table 3 provides a brief summary of the response rate for the proof of concept.

Table 3: Summary of Proof of Concept Response Rate

Course Type	SFU Proof of Concept Response Rate
Lecture and Seminar	82%
CODE	27%
TOTAL	72%

6 VOLUNTEER INSTRUCTOR FEEDBACK

In addition to collecting feedback from individual participants during the proof of concept, two review sessions were held with a total of seven volunteer instructors after the completion of the proof of concept. On the whole, the feedback was very positive. Participants also provided valuable constructive feedback and suggestions.

Key themes in participant feedback included:

- The flexibility of the multi-level framework is great, but it will be very important and could be challenging to reach agreement on the right questions to ask at the different levels
- The support materials provided were good, but personal support is strongly recommended during the transition to the new system
- The reports were valuable and useful, but more sophisticated reporting capabilities would be very helpful

- With a strong focus on encouraging student participation, the online approach seemed to produce good response rates and the open-ended responses were longer and more thoughtful.

Table 4 provides a summary of volunteer instructor feedback.

Table 4: Summary of Volunteer Instructor Feedback

Category	Feedback
Representative Framework	<ul style="list-style-type: none"> • Multi-level framework very positive, however noted that this would need to be adapted for faculties with no "department" level • Could be difficult to reach agreement on questions at different levels in some faculties/schools/departments; this may be especially challenging with CODE courses • Question bank generally positive; suggested language occasionally unnecessarily complex • Ability to ask students specific questions in the instructor level very positive; emphasis on the importance of being able to "guarantee" that results will only be available to the instructor • Suggestion to offer evaluations in multiple languages
Support	<ul style="list-style-type: none"> • The number of questions in the questions bank initially a bit overwhelming • How-to documents helpful, but recommended having people to guide instructors when first implementing the system • Links from the evaluation system to online resources preferred to pdfs • Centrally accessible online resources preferred as it was sometimes hard to locate the emails generated by the system
Reports	<ul style="list-style-type: none"> • Multiple levels of reporting positive • Greater statistics, comparison and trend data would assist with valid interpretation of results • More flexibility with demographics and exploring the data with "drill-down" capabilities seen as very valuable • Interest in having flexibility to design and create custom reports
Online versus Paper	<ul style="list-style-type: none"> • Noted that responses rates were comparable or better than historical rates of existing systems; in some cases best ever response rate for a given instructor • Students in a number of courses were offered a bonus mark to complete an evaluation - in some cases this was tied to peer pressure; appeared to have a positive impact on response rate; may not be practical for the full scale implementation • Open-ended responses were longer and more thoughtful; the assumption is that students take more time when completing the evaluation online outside of class • Noted that the online approach saves a lot of data entry of responses to open-ended questions • Easier to manage, administer and get information out of an online tool
Process	<ul style="list-style-type: none"> • Really appreciated the auto-reminder emails that only went to students that had not yet completed an evaluation

7 STUDENT FEEDBACK

Details of student feedback on the proof of concept can be found in the summary of survey results contained in Appendix XII.

8 CONCLUSION

Overall, the project team felt that the proof of concept was a valuable exercise. It demonstrated that the multi-level framework adapted from the University of Toronto appears to be sufficiently flexible and provided insight into some of the challenges that will need to be overcome for implementation. It highlighted the importance of providing adequate support during the transition to the new system. The importance of clearly defining reporting requirements up front and ensuring strong reporting capabilities during system selection was strongly underscored. Finally, it demonstrated that, with careful attention to encouraging student participation, it is possible to achieve a strong response rate with online evaluations at SFU.



SFU

SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

THE TEACHING AND COURSE EVALUATION PROJECT
FINAL REPORT

APPENDIX XI: PROOF OF CONCEPT: SAMPLE SUPPORT
MATERIALS

NOVEMBER 17, 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

TABLE OF CONTENTS

- 1 INFORMATION FOR INSTRUCTORS
- 2 INSTRUCTIONS FOR SELECTING AND CUSTOMIZING YOUR QUESTIONS
- 3 QUESTIONS AND ANSWERS FOR STUDENTS
- 4 TALKING POINTS FOR INSTRUCTORS
- 5 THE "BIG 6" PREDICTORS OF RESPONSE RATE
- 6 YOUR COURSE EVALUTION REPORT: A GUIDE FOR THE VPA OFFICE
- 7 YOUR COURSE EVALUTION REPORT: A GUIDE FOR FACULTY DEANS
- 8 YOUR COURSE EVALUTION REPORT: A GUIDE FOR SCHOOL/DEPARTMENT HEADS
- 9 YOUR COURSE EVALUTION REPORT: A GUIDE FOR STUDENTS

1 INFORMATION FOR INSTRUCTORS

SFU Teaching and Course Evaluation Project

Summer 2013 proof of concept: Information for instructors

Introduction

Thank you for your participation in this project!

SFU has been conducting student evaluations of teaching and courses for more than 30 years. During that time, SFU's evaluation instrument and processes have remained basically unchanged even though there has been [extensive research](#) on more effective ways to conduct evaluations and use the results.

The [Teaching and Course Evaluation \(TCE\) Project](#) was launched by the VP Academic in 2012 to make recommendations for replacing SFU's instrument and processes for student evaluation of instructors and courses and to develop a best-practices guide for interpretation and use of the data.

Your role

You are one of 14 instructors who have volunteered to participate in the summer [proof of concept](#) (PoC). The PoC is an important part of the TCE project. It is not intended to test a particular product. Rather, it will help the project team obtain pragmatic feedback from SFU instructors on a small scale trial of the emerging recommended approach to conducting student evaluations. Specifically, you will be helping to answer four questions:

1. Is the emerging recommended framework sufficiently flexible to meet the needs of the SFU community?
2. What kinds of support are the most meaningful and helpful for instructors when participating in conducting evaluations using the emerging recommended framework?
3. What can we learn about the distribution of evaluation results enabled by the representative system?
4. What can we learn about the effects on response rates of using the framework with the representative system?

The feedback that you and your students provide will help inform the project team's recommendations to Senate, and in particular, highlight challenges and issues that will need to be addressed as part of any implementation effort.

The information below will provide you with an overview of the PoC process and a timeline of key activities. You will receive more detailed information, including instructions for choosing the questions you would like to include on the evaluation instrument, next week. Thank you again for your willingness to participate.

Corinne Pitre-Hayes
Project Director

PoC questions and answers

1. What is the purpose of the PoC?

The primary purpose is to test the feasibility of an online teaching and course evaluation system at SFU. In particular, it will seek to do four things:

- Gauge the flexibility of the emerging recommended framework
- Assess the types of support most helpful to instructors
- Learn about the distribution of results enabled by the representative system
- Learn about response rates with the framework and representative system

2. How many instructors and students will be involved?

For this trial, there will be 14 instructors and 1297 students in 18 courses (including four online courses). All instructors are voluntary participants.

3. What will the evaluation instrument consist of?

The TCE form will include up to 23 questions:

- 8 institution-wide questions common to all forms
- 4 Faculty-wide questions
- 4 department/school-level questions
- Up to 4 questions determined by the course instructor
- 3 questions related to the PoC evaluation experience

This multi-level format is intended to facilitate the evaluation form addressing the needs and priorities of various constituents. Most questions will ask students for a scaled response. However, you will also have the opportunity to include some open-ended questions.

4. How will the questions be generated?

For this trial, questions will be drawn from a bank of 184 questions developed by the University of Toronto. To avoid duplication, questions selected at a higher level (for example, school/department) will not be visible at a lower level (for example, instructor).

At the instructor level, you will have the opportunity to choose questions from the question bank, but you will also have the option of creating your own open-ended questions.

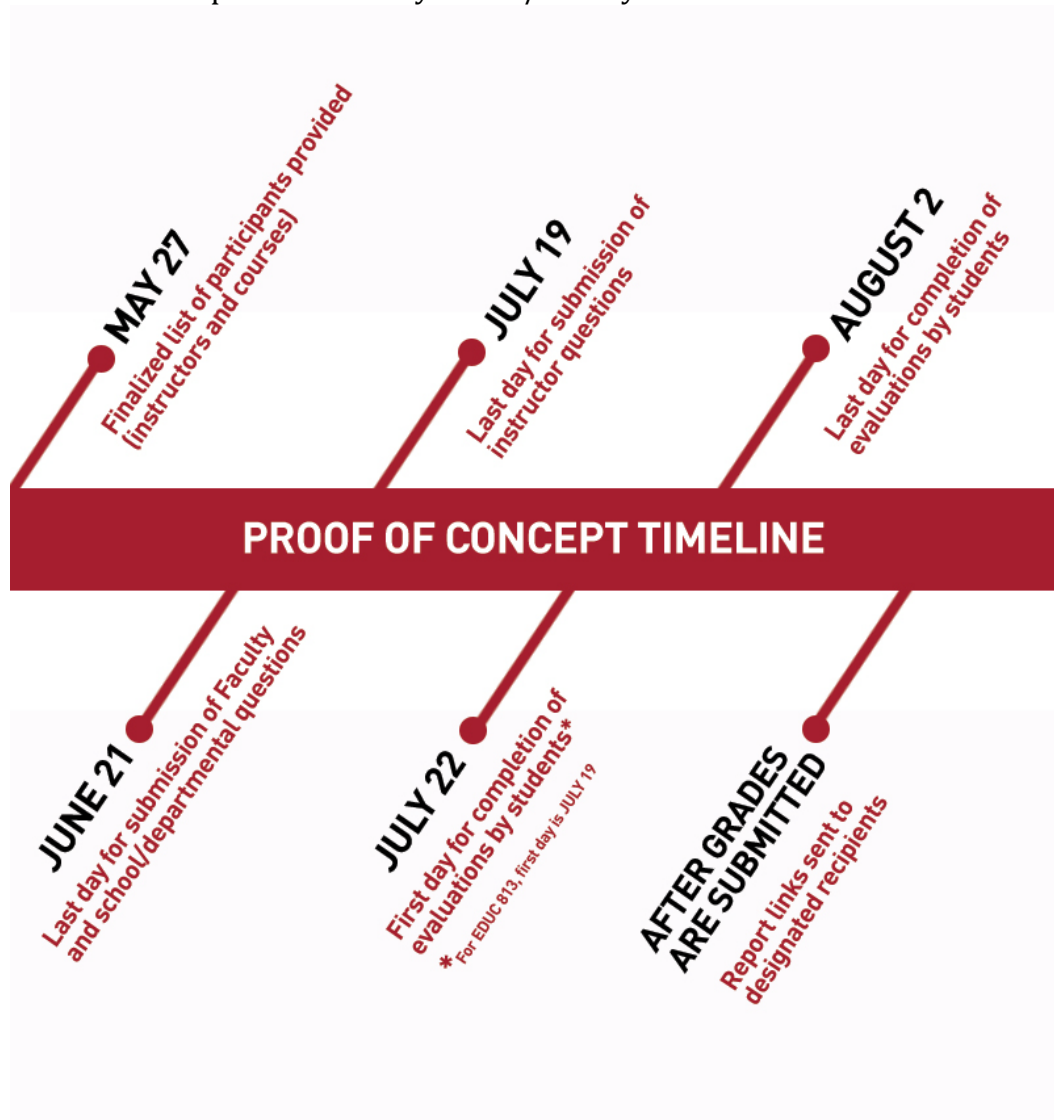
5. Who will see the results?

For this trial, chairs/directors, deans and the VP Academic will see all results above the instructor level. Only instructors will see the responses to instructor questions. Students will receive a general summary of the PoC results.

What will happen next?

The timeline graphic below outlines the PoC process.

- Next week you will receive an email invitation to submit up to four questions for each of your courses in the PoC. The email will include a link to the online question bank.
- You will have until July 19 to submit your questions.
- Your students will have from July 22 (July 19 in one case) to August 2 to complete their evaluations. During that time they will receive automated email reminders from the evaluation system.
- Once grades have been submitted for all courses, you and your students, as well as selected academic administrators, will receive an email with links to the data and reports to which you and/or they will have access.



2 INSTRUCTIONS FOR SELECTING AND CUSTOMIZING YOUR QUESTIONS

Summer 2013 – Proof of concept Instructions for selecting and customizing your questions

This document outlines the steps for selecting, customizing and submitting your questions for the teaching and course evaluation form.

Overview

As an instructor, you will be able to submit up to four questions for each of your courses. You may select all your questions from an online question bank, or you may substitute up to two open-ended questions that you create. You may copy questions selected for one course for use in another course. You also have the option of submitting no questions. Instructions for all these options are listed below.

How to select questions from the question bank

Step 1. Go to the online question bank.

You will receive an email inviting you to select your questions. Click on the link in the email to go to the online question bank.

Step 2. View the questions in the question bank.

Find the section titled “Select your questions from the question bank” and click the “Show Section” button. You will see an expanded section listing question categories. To see the questions inside any category, click the “Show Section” button in that category.

Step 3. Preview any questions that interest you.

Click the “Preview” button to the left of any question that interests you. This will show you the question along with the answer options that your students will see.

Step 4. Select the questions you want to add to your form.

To select a question, click the “Select” button to the right of the question. You can easily deselect a question you have selected by clicking the “Deselect” button to the right of the question. Note that selected questions are not added to the evaluation form until you click on “Submit.”

Step 5. Submit the questions you have selected.

Once you have finalized your questions, click the “Submit” button at the bottom of the page to add your questions to the evaluation form. Remember that you can submit a maximum of four questions.

If you are interrupted before you can complete the submission process, just click the “Save” button at the bottom of the page to store your work without submitting your questions. You will be able to return to your work by clicking on the email link you received previously.

If you want to change your selections later, you can do so by selecting new questions and clicking the “Update” button at the bottom of the page.

If you don’t want to select any questions, simply click the “Submit” button without selecting any questions. Doing this will spare you from reminder emails asking you to pick your questions.

How to create an open-ended question

Step 1. Go to the online question bank.

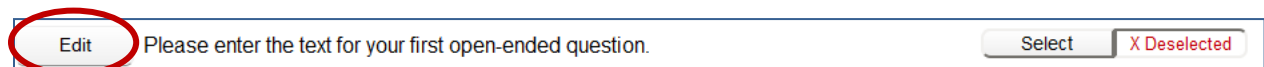
You will receive an email inviting you to select your questions. Click on the link in the email to go to the online question bank.

Step 2. Go to the section for creating custom questions.

Find the section titled “Create custom open-ended questions” and click the “Show Section” button. You will see an expanded section with an option for creating up to two questions.

Step 3. Create your question(s).

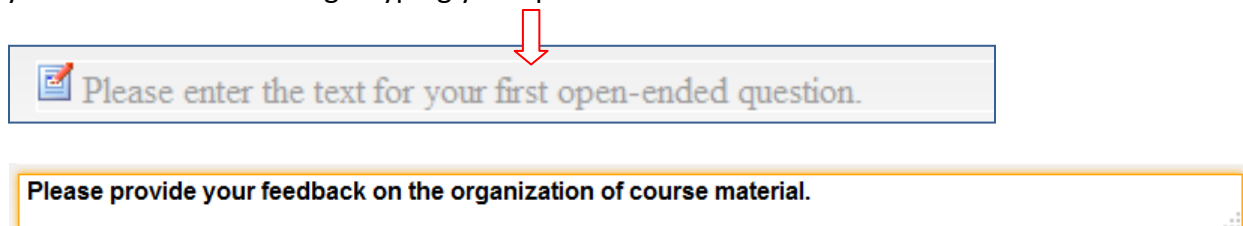
Click the “Edit” button beside the first question:



Once clicked, the question field will expand as shown:



Click on the phrase “Please enter the text for your first open-ended question” to activate the yellow edit box. Then begin typing your question.



To include your question on the evaluation form, click the “Select” button to the right of the question. The question will now show as “Selected.” You can easily deselect a question you have selected by clicking the “Deselect” button to the right of the question. Note that selected questions are not added to the evaluation form until you click on “Submit.”



If you wish, repeat the steps above to create another open-ended question.

Step 4. Submit the questions you have selected.

Once you have finalized your questions, click the “Submit” button at the bottom of the page to add your questions to the evaluation form. Remember that you can submit a maximum of four questions, including those you select from the question bank.

If you are interrupted before you can complete the customization process, just click the “Save” button at the bottom of the page to store your work without submitting your questions. You will be able to return to your work by clicking on the email link you received previously.

If you want to change your selections later, you can do so by repeating the edit process and clicking the “Update” button at the bottom of the page.

How to review and preview your questions

Near the top of the question selection/creation page, you will see a “View” search box:



Change the “All options” drop-down menu to show “Selected”:



The question bank will refresh to display only the subsections that contain your selected questions. Expand those subsections to view your questions. If you created any custom questions, these will also be displayed as long as you have selected them for your questionnaire.

To the far left of the “View” search box is a “Preview” button. Click this button to see the entire evaluation form as it will appear to your students.

How to save selected questions before submitting them

If you are interrupted before you can complete the customization process, just click the “Save” button at the bottom of the page to store your work without submitting your questions. You will be able to return to your work by clicking on the email link you received previously.



Once you receive the confirmation message that your settings have been saved, you may close the window.

Note that the “Exit” button at the bottom of the page can be used to refresh the page with your latest changes and prepare for the closing of the browser. Once you submit your questions, however, you can simply close the browser.

How to submit selected questions

Once you have finalized your questions, click the “Submit” button at the bottom of the page to add your questions to the evaluation form. Remember that you can submit a maximum of four questions, including those you select from the question bank.



Once you click “Submit,” the “Save” and “Submit” buttons are replaced by an “Update” button. At this point you may close the window since the questions have been submitted.

If you want to change your questions later, you can do so by repeating the edit process and clicking the “Update” button at the bottom of the page.

Note that once the question selection deadline is met the update option will no longer be available.

How to copy questions to/from one course to/from another

If you teach multiple courses and wish to apply the selected and/or custom questions from one course to another course, you can use the copy function.

First locate the “Copy” box at the bottom of the screen:

A screenshot of a user interface element for copying questions. It consists of a light blue rounded rectangle containing the text 'Copy' on the left, followed by a small 'to' label and a downward-pointing arrow icon. To the right of this is a larger, empty dropdown menu box with a downward-pointing arrow icon on its right side. Further to the right is a light blue button with the text 'Apply' in a darker blue font.

Using the drop-down box, select the source (from) or destination (to) course to copy the questions from or to:

A screenshot of the same 'Copy to' interface as above, but with the dropdown menu box now containing the text 'AD140012012F RESEARCH:METHODS&DATA'. The text is highlighted with a yellow border. The 'Apply' button remains to the right.

Click the “Apply” button to copy the questions. You will then see a message similar to this one:

The questionnaire has been copied to 'AD140012012F RESEARCH:METHODS&DATA (GREG KESWICK)' successfully.

Note that performing this operation will overwrite any question customization that might have been done for the destination course.

3 QUESTIONS AND ANSWERS FOR STUDENTS

SFU Teaching and Course Evaluation Project

Summer 2013 proof of concept: Questions and answers for students

1. What is the proof of concept (PoC) all about?

The proof of concept in which your class is participating is a small-scale trial of a new approach to conducting student evaluations. The trial is part of SFU's Teaching and Course Evaluation (TCE) Project, which is looking for more effective ways to collect and use student evaluations of instructors and courses.

2. How many people will be involved in the PoC?

The trial will involve 14 instructors and 1297 students in 18 summer courses.

3. How is the PoC different from the existing approach at SFU?

There are at least three differences:

- First, the evaluations provide a lot of flexibility to ask more relevant and appropriate questions. Faculties, departments and instructors have customized the questionnaires for their specific courses.
- Second, the results will be distributed more widely. For the proof of concept, students will receive aggregated results.
- Third, the evaluations will be done online instead of on paper.

4. How will the PoC work?

On July 22 you will receive an email invitation with a link. The link will take you to an online questionnaire for your course. The questionnaire will contain up to 23 questions:

- 8 institution-wide questions common to all forms
- 4 Faculty-wide questions
- 4 department/school-level questions
- Up to 4 questions determined by the course instructor
- 3 questions related to the evaluation experience

Most of these questions will use a multiple-choice format. A few questions will be open-ended so that you can comment in greater detail.

5. How long will it take me to complete the questionnaire?

You should be able to finish the evaluation in 15 minutes or so.

6. Will the answers I provide be used to do an actual evaluation of the course, or is this just a test of the process?

Your answers will definitely be used to evaluate the course. The instructors involved in the PoC are very interested in your feedback.

7. Will my instructor be able to identify my responses?

No. Instructors will not know which answers have been submitted by which students. In addition, instructors will not see the evaluation results until they have submitted final student grades.

8. Who else will see the results and how will they be used?

For the PoC, chairs/directors, deans and the VP Academic will see all results above the instructor level. Instructors will be the only ones to see the responses to instructor questions. Students will receive a general summary of aggregated evaluation results.

The results will be used by academic units to improve their courses and programs. They may also be used to evaluate the teaching performance of instructors. Instructors will be able to use the results to adjust their teaching practices.

9. What does the project team hope to learn from the PoC?

The trial has four goals:

- To gauge the flexibility of the new evaluation approach
- To assess the types of support most helpful to instructors
- To learn about the distribution of results enabled by the new approach
- To learn about response rates with the new approach

10. Where can I find out more about the PoC and the TCE Project?

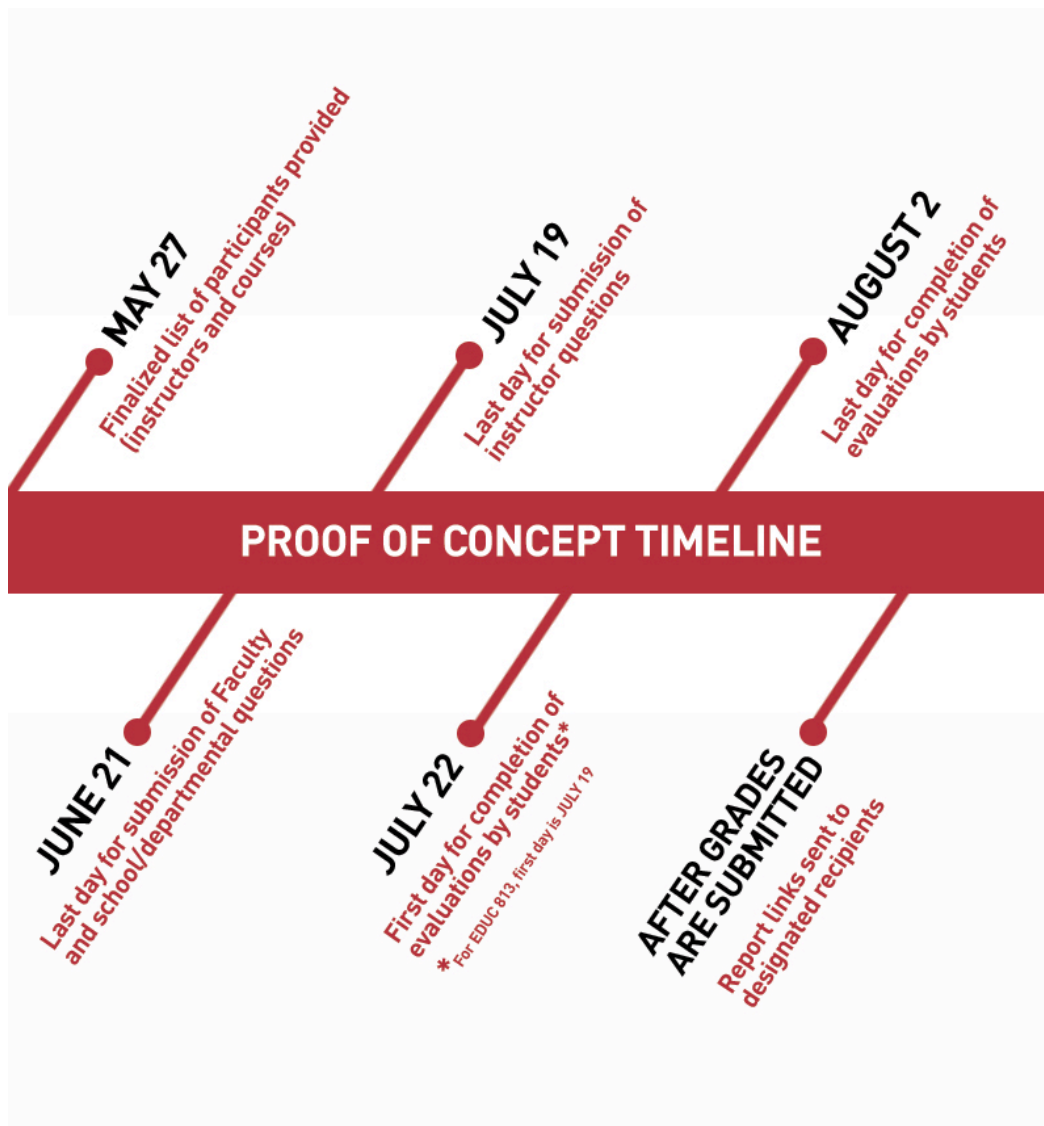
Basic information is available at www.sfu.ca/teachingandcourseeval.html. You may also contact Sarah Bolduc of the project team at sarahbolduc@gmail.com.

What will happen next?

The timeline graphic below outlines the PoC process.

- On July 22 (July 19 in one case) you will receive an email with a link inviting you to complete an online course evaluation.
- Between July 22 and August 2 you will receive reminder emails prompting you to complete the evaluation.
- On August 2 the link will be deactivated.

- Once grades have been submitted for all courses, you along with the course instructor and selected administrators will receive emails with links to the data and reports to which you and/or they will have access.



4 TALKING POINTS FOR INSTRUCTORS

SFU Teaching and Course Evaluation Project

Summer 2013 proof of concept | Talking points for instructors

Students—your participation will make a difference

It's very important and worthwhile for you to participate in course evaluations, and especially in this trial evaluation, which will help to improve the teaching and course evaluation process for all students at SFU. Your responses will help us to:

- Identify the most and least helpful parts of the course
- Determine which course components are working, and which should be revised
- Deliver compliments and constructive feedback to the appropriate people within the department/school and faculty

Ultimately, you will be helping to make SFU a better place for everyone!

What the proof of concept is

- A trial involving the students and instructors in 18 summer courses
- Part of a project to update SFU's course evaluation system, which hasn't changed much in 30 years
- An attempt to make course evaluations more useful for students, instructors and the university
- A test of how course evaluations could be done, not a finalized system

How the course evaluations will work

- On July 22, each student will receive an email invitation with a link to an online questionnaire
- The link will remain active until August 2
- Students will be able to answer the questions from any connected device, including smartphones and tablets
- Most questions will use a rating scale; some questions will be open-ended to allow more in-depth feedback
- The entire questionnaire should take about 15 minutes
- The results of the questionnaire will be used to evaluate both the course and the new evaluation format

How the trial system benefits students

- Questionnaires are online and can be accessed anytime during the period when they are active
- Questionnaires can be accessed on mobile devices
- Questions are customizable by departments and instructors, making them more relevant to the course
- Students will receive a report summarizing the survey results for their courses

Additional information

Visit the Teaching and Course Evaluation Project site at www.sfu.ca/teachingandcourseeval/proof.html

5 THE "BIG 6" PREDICTORS OF RESPONSE RATE

The "Big 6" Predictors of Course Evaluation Response Rate

Presenter: Matthew Champagne, Ph.D.

Moderator: Michael Edwards

Facilitator: Carey Watson



LOTA SOLUTIONS About today's Presenter

- ❑ Served as Director of Assessment, Founder of the IDEA Lab, Chair of University Evaluation Committee & Asst Professor of Psychology
- ❑ Senior Research Fellow and Evaluator for Army Research Institute, National Science Foundation & U.S. Department of Education
- ❑ Author of 30+ refereed articles and 50+ invited presentations on evaluation, embedded assessment, and learning
- ❑ Evaluator and PI on 12 federal and philanthropic grants of distance learning projects of national impact (\$8 million)
- ❑ Advisor to dozens of Fortune 100 & non-profits; composed 5000+ surveys for 900+ learning organizations

EVALUATION AND FEEDBACK TECHNOLOGY

LOTA SOLUTIONS National Survey Results

Top 5 Concerns:

1. Low Response Rate (46%)
2. Content (10%)
3. Results of evaluation not applied (9%)
4. Results not timely, not informative (9%)
5. <4% = Cost, disaster stories, resistance from faculty, etc.

EVALUATION AND FEEDBACK TECHNOLOGY

LOTA SOLUTIONS National Survey Results

Response Rate Concerns:

- Initial paper-to-online RR plunge
- Initial bump followed by the plunge
- Time taken each term to focus on new incentives

EVALUATION AND FEEDBACK TECHNOLOGY

LOTA SOLUTIONS "Big 6" Predictors

- Meaningfulness of Content
- Perceived Importance
- Go where the students are
- Closing the Feedback Loop
- Knowledge of Results
- "Meaningful" Incentives

EVALUATION AND FEEDBACK TECHNOLOGY

LOTA SOLUTIONS 1. Meaningfulness of Content

- Unambiguous Questions
- Robust Questions

EVALUATION AND FEEDBACK TECHNOLOGY



Unambiguous Questions

- Course Evaluation replaces the student interview. If they say "what?", you've lost them.
- Identify and remove the "stoppers"

EVALUATION AND FEEDBACK TECHNOLOGY



Unambiguous Questions

- A. "The course content was well organized and presented clearly."
- B. "The instructor maintained office hours and encouraged students to seek help when needed."

If they say "what?", you've lost them.

... extreme drop out rate of 26%

EVALUATION AND FEEDBACK TECHNOLOGY



Robust Questions

- Are the evaluation items comprehensive?
Are we asking too many questions?
Are we asking the right questions?

In 2000, 78 colleges reduced to 16 criteria

EVALUATION AND FEEDBACK TECHNOLOGY



Example Criteria & Questions

1. Course Organization
 - The objectives for this course are clear
 - My instructor has made expectations clear to the students
2. Communication
 - My instructor communicates effectively
 - My instructor responds effectively to student questions in class
3. Fairness of Grading
 - My instructor is fair with grading procedures

EVALUATION AND FEEDBACK TECHNOLOGY



Putting Predictor #1 in Action

1. Review your evaluation forms from student perspective
 2. Choose the 10-15 MOST important criteria for teaching
 3. Craft unambiguous targeted questions to reflect those criteria
- (Faculty can always ask additional questions not found in the "core")

EVALUATION AND FEEDBACK TECHNOLOGY




2. Perceived Importance

- Faculty Appeal – best predictor of response rate – responsible for largest variability
- Administrative Appeals
- Reminders

EVALUATION AND FEEDBACK TECHNOLOGY

LOTA SOLUTIONS Putting Predictor #2 in Action


1. Give your instructors a template or suggestions for explaining importance to students
2. Tell students something beyond "this is important"



EVALUATION AND FEEDBACK TECHNOLOGY

LOTA SOLUTIONS 3. Go where the Students are

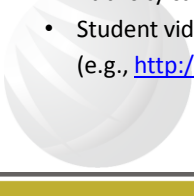
- Peer Encouragement
- Mobile Devices



EVALUATION AND FEEDBACK TECHNOLOGY


LOTA SOLUTIONS 3. Go where the students are

- Peer Encouragement
- Student Government
- Publicity campaigns
- Student videos of encouragement (e.g., <http://vimeo.com/11495435>)

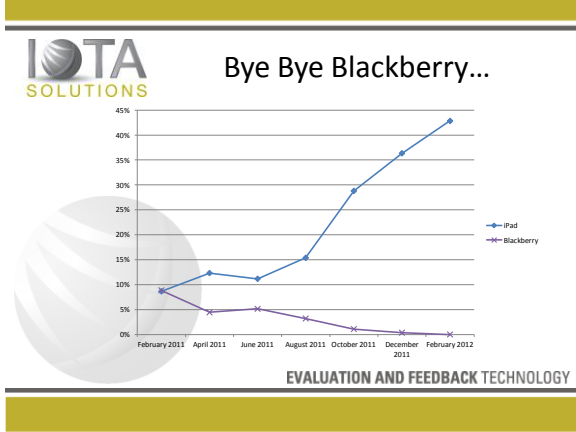
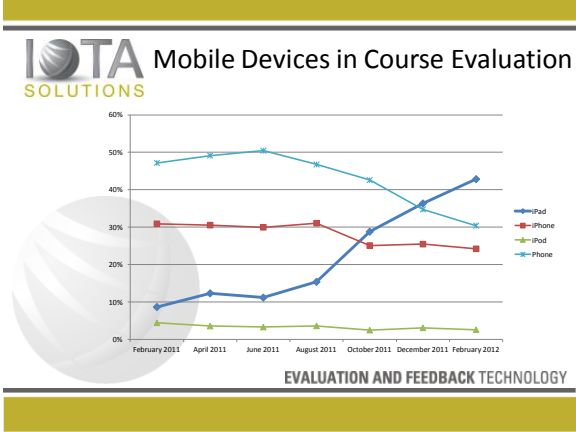


EVALUATION AND FEEDBACK TECHNOLOGY

Go where the students are:
Mobile Devices



Course Organization	Strongly Agree	Agree	Neither Agg. Nor Disagg.
1. This course well organized			
2. The course/teacher's clear			
3. The objectives for this course are clear			
4. My instructor connects the course to the real world			
5. Please rate your confidence in suggestions/feedback provided and implemented by the organization as part of this course			





Putting Predictor #3 in Action

1. Encourage students to fill out evaluations via mobile devices in and outside of class
2. Track their use and modify accordingly
3. Get talented and creative students involved

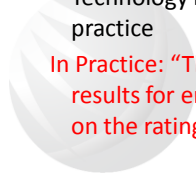


EVALUATION AND FEEDBACK TECHNOLOGY



4. Closing the Feedback Loop

- Provide students feedback to their feedback
- More important today than ever
 - Technology makes it easier to put this into practice



In Practice: "Those who complete it can see results for entire campus on 4 specific items on the rating form."

EVALUATION AND FEEDBACK TECHNOLOGY



4. Closing the Feedback Loop

Midterm Evaluation



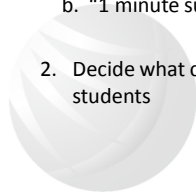
"I used some of the suggestions and made changes to my course."

EVALUATION AND FEEDBACK TECHNOLOGY



Putting Predictor #4 in Action

1. Midterm Evaluation! Suggestions:
 - a. Same as End-Of-Term
 - b. "1 minute survey" = pro, con, 1 question not asked
2. Decide what chunk of feedback can be provided to students



EVALUATION AND FEEDBACK TECHNOLOGY

5. Knowledge of Results

myclass evaluation
HANDS-FREE COURSE EVALUATION FOR HIGHER EDUCATION

ACTIVE EVALUATIONS | RECENT EVALUATIONS | HISTORICAL EVALUATIONS

LOGO

Summer Term 2011

COURSE	#STUDENTS	#RESPONSES	RESPONSE RATE%	ROSTER
MI101-846-SEM10	29	5	17.2%	
MI102-847-SEM10	30	15	50%	
MI103-846-SEM10	10	4	40%	

BENCHMARKING >>> YOUR DEPARTMENT AVERAGE: 56.00%
INSTITUTION AVERAGE: 50%

IMPORTANT! Your evaluation is still active and we are still calculating your responses. Results are not available until after the evaluation ends. Complete access to navigation will resume when your evaluation closes.

Student Response Rates for Your Courses (HTML)

Click here to download pictures. To help protect your privacy, Outlook prevented automatic download of some pictures in this message.

From: yourusername@universityof.edu
To: yourname@domain.com
Cc:
Subject: Student Response Rates for Your Courses

Sent: Thu 3/3/2011 7:25 AM

Dear Professor Nolan:

The course evaluation period ends this Sunday, December 16. Below is the "midstream" feedback on the student response rate for your courses.

Please encourage your students to visit www.schooleval.com and complete the evaluation form for your courses. If you have any questions or concerns about the evaluation process, do not hesitate to contact me directly at mudlyn@lotasolutions.com. Thank you!

Course	Students	Responses	Response Rate
Critical Thinking and Problem Solving	21	4	19.0%
Ethical and Legal Issues in the Professions	10	2	20.0%
Intro to Psychology	7	2	28.6%



Putting Predictor #5 in Action

- Comparisons with colleagues
- Advice for boosting rates
- Push same information via email to faculty multiple times during evaluation period

EVALUATION AND FEEDBACK TECHNOLOGY



6. Incentives

From National Survey:

- Most successful: Bonus points / extra credit
- Least successful: Giveaways by lottery
- Incentive inflation:
 - Amazon gift card, MP3 player, iPod, Kindle, Netbook, iPad, ... ?? ...
- Cost + Perceptions + Loss of Focus

EVALUATION AND FEEDBACK TECHNOLOGY



Incentives with a Lesson

- Purpose: Education-focus versus external-focus
- Early view of grades for students
- Survey says: 68% had no monetary incentive
- **Of course, Midterm, close the loop, knowing their voice was heard**

EVALUATION AND FEEDBACK TECHNOLOGY



Takeaways & Questions

- Online evaluation is a difficult problem with KNOWN solutions
- Not alone – same conversation at every school
- Apply the Big 6 to your own evaluation practices

EVALUATION AND FEEDBACK TECHNOLOGY

6 YOUR COURSE EVALUATION REPORT: A GUIDE FOR THE VPA OFFICE

Your course evaluation report: A guide for the VPA Office

I. General guidelines for reviewing the course evaluation institutional report

1. Read, understand and think about the questions on the evaluation before reviewing the student responses to better understand what the results may entail.
2. Some demographic information is provided in the first section of your report. These demographic results break down the respondents' gender, age, student origin, cumulative GPA and program information. Since the proof of concept has a limited number of participating students, some course evaluation did not include demographic information to avoid infringing on the anonymity of the respondents. Therefore, the available demographic information may not be fully representative of all the students/courses that have participated in the evaluation.
3. Review the quantitative data in your report. The following pages of this guide outline an approach to understanding the statistical summary of students' responses.
4. Take note of any extreme results in your report. For example, you might notice a few students strongly disagree that the course was intellectually stimulating. Although it is important to pay attention to all responses, it is also important to note a few extremes are not cause for worry. For example, in a faculty of 200 students with a response rate of 100 students, two or three comments suggesting the course was not intellectually stimulating are not reflective of the majority of students' responses.
5. Apart from the institutional report, you will receive faculty and school/department reports. The same guidelines apply to a faculty and school/department reports, but keep in mind that the statistical results will represent only the faculty/department addressed in that specific report.
6. Please note: We have included optional open-ended questions only in the individual (instructor) reports. The TCEP team determined that, for reporting purposes, the comments collected in response to these open-ended questions would be most useful at the course level. It would, however, be possible to provide them at the institutional, Faculty or school/department level upon request.
7. Remember: If you would like consultation on any aspect of your course evaluation report, please contact the Teaching and Course Evaluation Group at course_evaluations@sfu.ca.

II. Understanding overall results

Scale used: 1=Strongly Disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree

Question	Response
	Mean
The course provided me with a deeper understanding of the subject matter.	3.82
The course covered the content that was described in the course outline.	3.86
The course assignments, projects, tests and exams allowed me to demonstrate my understanding of the course material.	3.86
The instructor and/or the course material stimulated me to think in new ways.	3.95
The instructor created an atmosphere that was conducive to learning.	4.23
Core Institution-wide Questions Composite Mean	3.95

Scale used: 1=Poor, 2=Fair, 3=Good, 4=Very Good, 5=Excellent

Question	Response
	Mean
Overall, the quality of my learning experience in this course was:	3.27

The response scale for the listed items required the students to indicate the extent to which they agree with the statements. This is one of the two scales used throughout the course evaluation. The second scale ranges from poor to excellent.

Please note that other scales may be used for some questions, depending on how they were designed at the faculty, departmental or individual level. The type of scale used will always be noted for each result.

The mean shows the institutional average response to each listed item.

The Core Institution-wide Questions Composite Mean is the average of mean responses to the listed question. This is the mean of means for the listed questions on the table.

III. Understanding the details and summary pages of your report

Statistics	Value
Mean	2.72
Median	3.00
Mode	2, 3
Standard Deviation	+/-1.16

The mode, median and mean are measures of the center of the distribution of student ratings. All three are useful summaries of the results but some may be more appropriate than others depending on the shape of the distribution. For example, the median may be more appropriate than the mean when the distribution is skewed.

Mean: average of all responses to a question
Median: represents the middle response of a question's rating when ordered from smallest to largest
Mode: represents the response that occurs most frequently for that question.

Please note that the calculated institutional averages reflect only the courses that were part of the Teaching and Course Evaluation Project's summer 2013 proof of concept.

The standard deviation represents the average spread of each response from the mean. Small standard deviation suggests that the ratings are close to the mean, whereas large standard deviations suggest ratings are distant from the mean.

7 YOUR COURSE EVALUATION REPORT: A GUIDE FOR FACULTY DEANS

Your course evaluation report: A guide for faculty deans

I. General guidelines for reviewing the course evaluation for the Faculty report

1. Read, understand and think about the questions on the evaluation before reviewing the student responses to better understand what the results may entail.
2. Some demographic information is provided in the first section of your report. These demographic results break down the respondents' gender, age, student origin, cumulative GPA and program information. Since the proof of concept has a limited number of participating courses and students, some reports may not include demographic information to avoid infringing on the anonymity of the respondents.
3. Review the quantitative data in your report. The following pages of this guide outline an approach to understanding the statistical summary of students' responses.
4. Take note of any extreme results in your report. For example, you might notice a few students strongly disagree that a particular course was intellectually stimulating. Although it is important to pay attention to all responses, it is also important to note a few extremes are not cause for worry. For example, in a faculty of 200 students with a response rate of 100 students, two or three comments suggesting the course was not intellectually stimulating are not reflective of the majority of students' responses.
5. Apart from the Faculty report, you will receive school/department and individual reports. However, these reports do not include instructor-level questions and results. The same guidelines apply to an school/department or individual report, but keep in mind that the statistical results will reflect only the department/course addressed in that specific report.
6. Please note: We have included optional open-ended questions only in the individual (instructor) reports. The TCEP team determined that, for reporting purposes, the comments collected in response to these open-ended questions would be most useful at the course level. It would, however, be possible to provide them at the institutional, Faculty or school/department level upon request.
7. Remember: If you would like consultation on any aspect of your course evaluation report, please contact the Teaching and Course Evaluation Group at course_evaluations@sfu.ca.

II. Understanding overall results

Scale used: 1=Strongly Disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree

Question	Response
	Mean
The course provided me with a deeper understanding of the subject matter.	3.82
The course covered the content that was described in the course outline.	3.86
The course assignments, projects, tests and exams allowed me to demonstrate my understanding of the course material.	3.86
The instructor and/or the course material stimulated me to think in new ways.	3.95
The instructor created an atmosphere that was conducive to learning.	4.23
Core Institution-wide Questions Composite Mean	3.95

Scale used: 1=Poor, 2=Fair, 3=Good, 4=Very Good, 5=Excellent

Question	Response
	Mean
Overall, the quality of my learning experience in this course was:	3.27

The response scale for the listed items required the students to indicate the extent to which they agree with the statements. This is one of the two scales used throughout the course evaluation. The second scale ranges from poor to excellent.

Please note that other scales may be used for some questions, depending on how they were designed at the faculty, departmental or individual level. The type of scale used will always be noted for each result.

The mean shows the faculty average response to each listed item

The Core Institution-wide Questions Composite Mean is the average of mean responses to the listed question. This is the mean of means for the listed questions on the table.

III. Understanding the details and summary pages of your report (Part 1 of 2)

Statistics	Value
Mean	2.72
Median	3.00
Mode	2, 3
Standard Deviation	+/-1.16

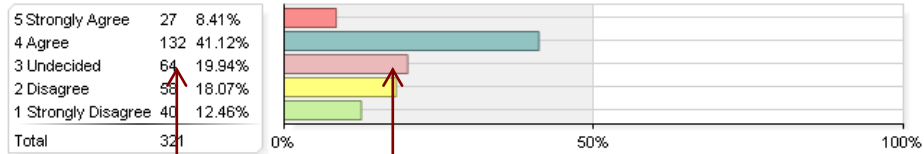
The mode, median and mean are measures of the center of the distribution of student ratings. All three are useful summaries of the results but some may be more appropriate than others depending on the shape of the distribution. For example, the median may be more appropriate than the mean when the distribution is skewed.

Mean: average of all responses to a question
Median: represents the middle response of a question's rating when ordered from smallest to largest
Mode: represents the response that occurs most frequently for that question.

The standard deviation represents the average spread of each response from the mean. Small standard deviation suggests that the ratings are close to the mean, whereas large standard deviations suggest ratings are distant from the mean.

III. Understanding the details and summary pages of your report (Part 2 of 2)

3. The course assignments, projects, tests and exams allowed me to demonstrate my understanding of the course material.



Comparison of Mean Scores

Question	Department Average			Faculty Average			Institutional Average		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
The course assignments, projects, tests and exams allowed me to demonstrate my understanding of the course material.	3.15	3.00	1.19	3.15	3.00	1.19	3.84	4.00	1.07

Each question breaks down the responses according to the number of students who endorsed that specific scale rating, along with the rating's percentage in relation to the other ratings.

The Comparison of Mean Scores table shows the school/department's mean, median and standard deviation in relation to the school/department's respective faculty and institutional averages. If more than one school/ department participated within a faculty, the department average will display NRP (no response). Otherwise, the department and faculty averages would be identical.

Please note that the calculated departmental, faculty and institutional averages reflect only the courses that were part of the Teaching and Course Evaluation Project's summer 2013 proof of concept.

There is a bar graph or response distribution available for each question. Each bar represents the percentage of student respondents who endorsed each rating on the scale.

When looking at this visual representation, it may help to reflect on the following questions: Do responses fall to one side of the scale more so than the other side? Do students' responses show highly divergent responses or two peaks in the distribution? Is there any large percentage of students who selected the extreme scale responses?

**8 YOUR COURSE EVALUTION REPORT: A GUIDE FOR
SCHOOL/DEPARTMENT HEADS**

Your course evaluation report: A guide for school/department heads

I. General guidelines for reviewing the course evaluation school/department report

1. Read, understand and think about the questions on the evaluation before reviewing the student responses to better understand what the results may entail.
2. Some demographic information is provided in the first section of your report. These demographic results break down the respondents' gender, age, student origin, cumulative GPA and program information. Since the proof of concept has a limited number of participating courses and students, some reports may not include demographic information to avoid infringing on the anonymity of the respondents.
3. Review the quantitative data in your report. The following pages of this guide outline an approach to understanding the statistical summary of students' responses.
4. Take note of any extreme results in your report. For example, you might notice a few students strongly disagree that a course was intellectually stimulating. Although it is important to pay attention to all responses, it is also important to note a few extremes are not cause for worry. For example, in a school/department of 200 students with a response rate of 100 students, two or three comments suggesting the course was not intellectually stimulating are not reflective of the majority of students' responses.
5. If a majority of students felt an item was not at all or only somewhat part of their learning experiences, instructors are encouraged to take note and be proactive for the next time they teach the course. They are also encouraged to take advantage of resources and support through the Teaching and Learning Centre at www.sfu.ca/tlcentre.
6. As mentioned, apart from the school/department report, you will receive individual reports for your participating instructors. However, these reports will not include instructor-level questions and results. The same guidelines apply to individual reports, but keep in mind that the statistical results will reflect only the course addressed in that report.
7. Please note: We have included optional open-ended questions only in the individual (instructor) reports. The TCEP team determined that, for reporting purposes, the comments collected in response to these open-ended questions would be most useful at the course level.
8. Remember: If you would like consultation on any aspect of your course evaluation report, please contact the Teaching and Course Evaluation Group at course_evaluations@sfu.ca.

II. Understanding overall results

Scale used: 1=Strongly Disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree

Question	Response
	Mean
The course provided me with a deeper understanding of the subject matter.	3.82
The course covered the content that was described in the course outline.	3.86
The course assignments, projects, tests and exams allowed me to demonstrate my understanding of the course material.	3.86
The instructor and/or the course material stimulated me to think in new ways.	3.95
The instructor created an atmosphere that was conducive to learning.	4.23
Core Institution-wide Questions Composite Mean	3.95

Scale used: 1=Poor, 2=Fair, 3=Good, 4=Very Good, 5=Excellent

Question	Response
	Mean
Overall, the quality of my learning experience in this course was:	3.27

The response scale for the listed items required the students to indicate the extent to which they agree with the statements. This is one of the two scales used throughout the course evaluation. The second scale ranges from poor to excellent.

Please note that other scales may be used for some questions, depending on how they were designed at the faculty, departmental or individual level. The type of scale used will always be noted for each result.

The mean shows the school/department average response to each listed item.

The Core Institution-wide Questions Composite Mean is the average of mean responses to the listed question. This is the mean of means for the listed questions on the table.

III. Understanding the details and summary pages of your report (Part 1 of 2)

Statistics	Value
Mean	2.72
Median	3.00
Mode	2, 3
Standard Deviation	+/-1.16

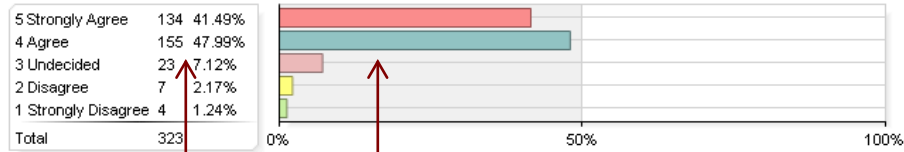
The mode, median and mean are measures of the center of the distribution of student ratings. All three are useful summaries of the results but some may be more appropriate than others depending on the shape of the distribution. For example, the median may be more appropriate than the mean when the distribution is skewed.

Mean: average of all responses to a question
Median: represents the middle response of a question's rating when ordered from smallest to largest
Mode: represents the response that occurs most frequently for that question.

The standard deviation represents the average spread of each response from the mean. Small standard deviation suggests that the ratings are close to the mean, whereas large standard deviations suggest ratings are distant from the mean.

III. Understanding the details and summary pages of your report (Part 2 of 2)

3. The course assignments, projects, tests and exams allowed me to demonstrate my understanding of the course material.



Comparison of Mean Scores

Question	Department Average			Faculty Average			Institutional Average		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
The course assignments, projects, tests and exams allowed me to demonstrate my understanding of the course material.	4.26	4.00	0.79	4.25	4.00	0.78	3.84	4.00	1.07

Each question breaks down the responses according to the number of students who endorsed that specific scale rating, along with the rating's percentage in relation to the other ratings.

The Comparison of Mean Scores table shows the school/department's mean, median and standard deviation in relation to the school/department's respective faculty and institutional averages.

Please note that the calculated departmental, faculty and institutional averages reflect only the courses that were part of the summer 2013 proof of concept.

There is a bar graph or response distribution available for each question. Each bar represents the percentage of student respondents who endorsed each rating on the scale.

When looking at this visual representation, it may help to reflect on the following questions: Do responses fall to one side of the scale more so than the other side? Do students' responses show highly divergent responses or two peaks in the distribution? Is there any large percentage of students who selected the extreme scale responses?

9 YOUR COURSE EVALUTION REPORT: A GUIDE FOR STUDENTS

Your course evaluation report: A guide for students

I. General guidelines for reviewing your course evaluation report

1. Some demographic information is provided in the first section of your report. These demographic results break down the respondents' gender, age, student origin, cumulative GPA and program information. Please note that some reports may not include demographic information to avoid infringing on the anonymity of the respondents.
2. You are receiving the evaluation results for institution-wide questions and for your own faculty and school/department. If your instructor chose to share his or her evaluation reports, you may also receive course-specific results. The following pages of this guide outline an approach to understanding the statistical summary of the results.
3. Review the quantitative data in your report.
4. Take note of what report you are reading. You will notice that a numeric result for the same question may differ depending on whether it is an institutional, faculty or departmental result. For example, for questions that provide a mean, the results in the departmental report will be based on the responses of students in your department, whereas the institutional report will be based on the responses of all participating students at SFU.
5. Please note that the calculated departmental, faculty and institutional averages reflect only the courses that were part of the summer 2013 proof of concept.

II. Understanding the overall results

Scale used: 1=Strongly Disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree

Question	Response
	Mean
The course provided me with a deeper understanding of the subject matter.	3.82
The course covered the content that was described in the course outline.	3.86
The course assignments, projects, tests and exams allowed me to demonstrate my understanding of the course material.	3.86
The instructor and/or the course material stimulated me to think in new ways.	3.95
The instructor created an atmosphere that was conducive to learning.	4.23
Core Institution-wide Questions Composite Mean	3.95

Scale used: 1=Poor, 2=Fair, 3=Good, 4=Very Good, 5=Excellent

Question	Response
	Mean
Overall, the quality of my learning experience in this course was:	3.27

The response scale for the listed items required students to indicate the extent to which they agree with the statements. This is one of the two scales used throughout the course evaluation. The second scale ranges from poor to excellent.

Please note that other scales may be used for some questions, depending on how they were designed at the faculty, departmental or individual level. The type of scale used will always be noted for each result.

The mean shows the students' average response to each listed item

The Core Institution-wide Questions Composite Mean is the average of students' mean responses to the listed question. This is the mean of means for the listed questions on the table.

III. Understanding the details and summary pages of your report (Part 1 of 2)

Statistics	Value
Mean	2.72
Median	3.00
Mode	2, 3
Standard Deviation	+/-1.16

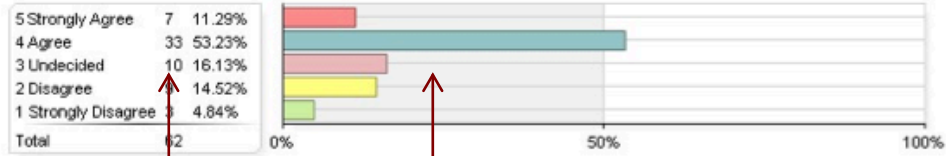
The mode, median and mean are measures of the center of the distribution of student ratings. All three are useful summaries of the results but some may be more appropriate than others depending on the shape of the distribution. For example, the median may be more appropriate than the mean when the distribution is skewed.

Mean: average of all responses to a question
Median: represents the middle response of a question's rating when ordered from smallest to largest
Mode: represents the response that occurs most frequently for that question.

The standard deviation represents the average spread of each response from the mean. Small standard deviation suggests that the ratings are close to the mean, whereas large standard deviations suggest ratings are distant from the mean.

III. Understanding the details and summary pages of your report (Part 2 of 2)

3. The course assignments, projects, tests and exams allowed me to demonstrate my understanding of the course material.



Comparison of Mean Scores

Question	Course Average			Department Average			Faculty Average			Institutional Average		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
The course assignments, projects, tests and exams allowed me to demonstrate my understanding of the course material.	3.52	4.00	1.04	3.15	3.00	1.19	3.15	3.00	1.19	3.84	4.00	1.07

Each question breaks down the responses according to the number of students who endorsed that specific scale rating, along with the rating's percentage in relation to the other ratings.

The Comparison of Mean Scores table shows the course's mean, median and standard deviation in relation to the course's respective departmental, faculty and institutional averages. Please note that the calculated departmental, faculty and institutional averages reflect only the courses that were part of the Teaching and Course Evaluation project's summer 2013 proof of concept.

There is a bar graph or response distribution available for each question. Each bar represents the percentage of student respondents who endorsed each rating on the scale.



SFU

SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

TEACHING AND COURSE EVALUATION PROJECT FINAL REPORT

APPENDIX XII: PROOF OF CONCEPT, RESULTS FROM STUDENT SURVEYS

Prepared by Joanne Quan, Analyst; Jessica Tilley, Analyst
Institutional Research and Planning, Simon Fraser University

OCTOBER 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

TABLE OF CONTENTS

1	RESULTS FROM STUDENT SURVEY: TEACHING AND COURSE EVALUATION INSTRUMENT	1
1.1	INTRODUCTION	1
1.2	METHODOLOGY	1
1.3	RESULTS	1
1.3.1	STUDENT USE OF ELECTRONIC DEVICES FOR THE EVALUATIONS	1
1.3.2	FEEDBACK ON THE EVALUATION QUESTIONS AND INTERFACE	1
1.3.3	STUDENT COMMENTS	2
1.4	SURVEY INSTRUMENT	3
1.5	DISTRIBUTION OF STUDENT RESPONSES	3
2	RESULTS FROM SURVEY OF STUDENTS: TEACHING AND COURSE EVALUATION REPORTS	5
2.1	INTRODUCTION	5
2.2	METHODOLOGY	5
2.3	RESULTS	5
2.3.1	FEEDBACK ON THE PRESENTATION OF RESULTS	5
2.3.2	FEEDBACK ON THE STUDENTS' INTERPRETATION OF RESULTS	5
2.3.3	FEEDBACK ON WHAT SHOULD BE INCLUDED IN THE EVALUATION REPORT	6
2.4	SURVEY INSTRUMENT	6
2.5	DISTRIBUTION OF STUDENT RESPONSES	7

1 RESULTS FROM STUDENT SURVEY: TEACHING AND COURSE EVALUATION INSTRUMENT

1.1 INTRODUCTION

The purpose of this survey was to get feedback from students about their experience with the Summer 2013 proof-of-concept online teaching and course evaluation. The survey asked students about the evaluation questions, the evaluation instrument's interface, and other comments they have. This report presents a summary of the survey's outcome.

1.2 METHODOLOGY

This survey was administered by eXplorance Inc. as part of the online teaching and course evaluations. The survey questions appeared at the end of the evaluation form and were optional. A total of 1,297 students¹ who received invitations to fill out their course evaluations were also invited to participate in the survey. Of these students, 960 participated in the survey², representing a response rate of 74%. Assuming the sample is representative, proportions calculated on all respondents are accurate within $\pm 1.6\%$, 19 times out of 20.

1.3 RESULTS

The general feedback was very positive and many students agreed that the online evaluation was easy to use. This section summarizes the student responses and highlights the common themes from the open-ended comments.

The survey instrument and distribution of responses to the survey questions can be found in sections 4 and 5 respectively.

1.3.1 Student Use of Electronic Devices for the Evaluations

Students were asked about the device(s) they used to complete the online course evaluation. Among respondents:

- 84% used a desktop and/or laptop to complete the evaluation, and
- 21% used a smartphone and/or tablet to complete the evaluation.

1.3.2 Feedback on the Evaluation Questions and Interface

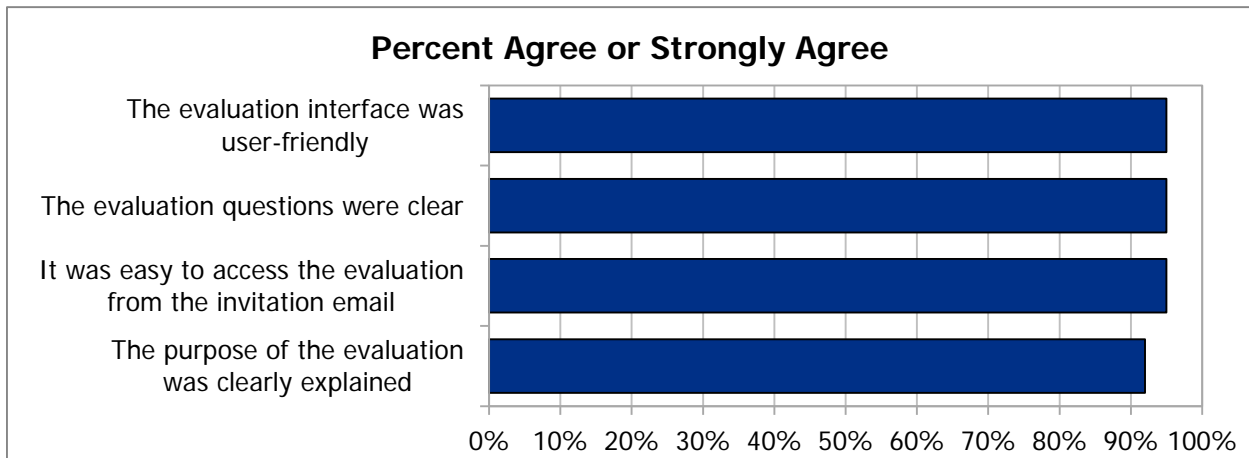
Students were then asked to rate their agreement with various aspects of the course evaluation (Figure 1). Among respondents:

- 93-95% said that they agree or strongly agree that
 - the purpose of the evaluation was clearly explained,
 - it was easy to access the evaluation from the invitation e-mail,
 - the evaluation questions were clear, and
 - the evaluation interface was user-friendly.

¹ The survey appeared on every evaluation form, so students who were enrolled in more than one course could have filled out the survey more than once. This total represents a unique count of students.

² This is not a unique count of students. The results presented in this report may contain multiple responses from the same student. Since the results were anonymous, there was no way to identify duplicates. The number of students who were enrolled in more than one course is very small.

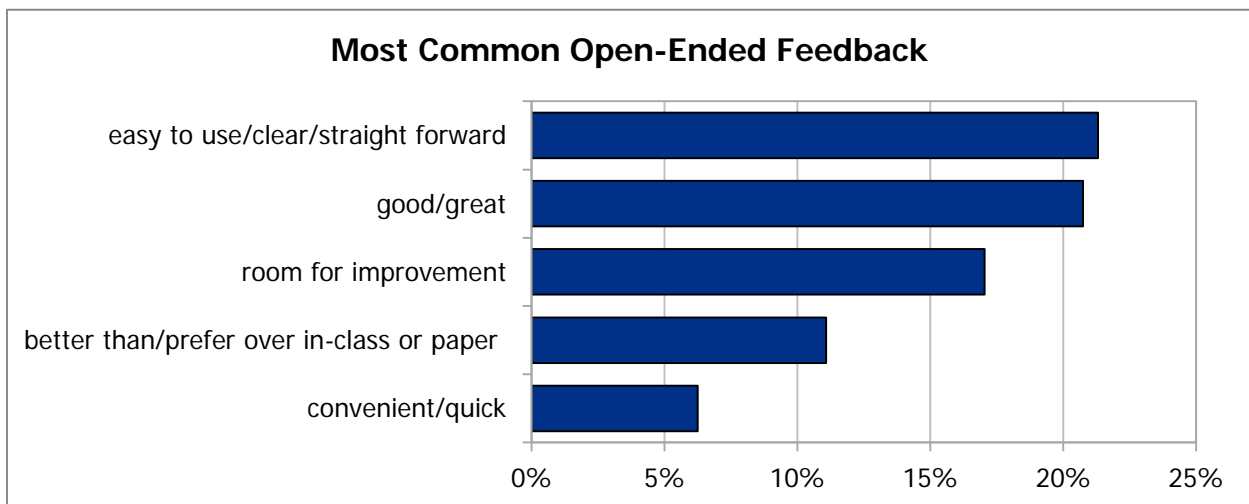
Figure 1



1.3.3 Student Comments

Finally, students were asked if they had any comments about their overall experience with the online evaluation. Figure 2 displays the top 5 common themes from the student comments.

Figure 2



Notes: The percentage represents the percentage of respondents who made a comment. The categories are not mutually exclusive; students could have made comments that fall under more than one category.

Some common reasons why students preferred the online evaluation over the current paper-based evaluation include:

- the online evaluation saves class time and paper,
- students can take their time to think about their responses when there is no pressure to complete the evaluation quickly, and
- the online evaluation questions were more thorough than the paper-based evaluation and allowed students to give more thorough feedback about the course.

While most students complimented the success of the online evaluation tool, many students also provided constructive feedback to improve the evaluation process for future implementations. Student suggestions for areas of improvement are listed below:

- Include more questions on the evaluation
 - Include questions about student assessments, such as midterms, assignments, grading, etc.
 - Include a question about the course instructor’s strengths and weaknesses.
 - Include open-ended comment boxes at the end of each scaled question so students can elaborate on their responses.
 - Conclude the evaluation with a question that asks students for any further comments or concerns about the course or instructor.

- Improve the evaluation interface
 - Increase the size of the open-comment text boxes for mobile devices.
 - Place the ‘next’ button in a more intuitive location.
 - Reduce the width of the evaluation form so it fits better on the computer screen.
 - Keep the ordering of scales consistent (e.g. scales went from positive to negative then switched and went from negative to positive).

- Improve the invitation process
 - Make in-class announcements to avoid invitation e-mails going unread (e.g. students may not check their inboxes regularly).

1.4 SURVEY INSTRUMENT

- 1) What device did you use to complete this evaluation?
- 2) Please state your level of agreement with the following.
(Scale: Strongly Agree, Agree, Undecided, Disagree, Strongly Disagree)
 - a. The purpose of the evaluation was clearly explained.
 - b. It was easy to access the evaluation from the invitation e-mail.
 - c. The evaluation questions were clear.
 - d. The evaluation interface was user-friendly.
- 3) Please write any comments you have about your overall experience with the evaluation.

1.5 DISTRIBUTION OF STUDENT RESPONSES

This section displays the distribution of student responses² to the survey questions. The tables that follow show the number and percentage of students selecting each response to the questions. In cases where students could select more than one response, the percentage of all respondents (percentage of all those who answered the question) is provided. Please note that percentages in these tables are rounded to one decimal place. Summaries that group categories (e.g. where “Strongly Agree” and “Agree” are grouped) may result in percentages that do not exactly match the sum of the percentages displayed in the tables. This is an artifact of rounding.

1) What device did you use to complete this evaluation?		
Laptop	492	52.1%
Desktop	299	31.6%
Smartphone	147	15.6%
Tablet	50	5.3%
Other	1	0.1%
Total Respondents	945	

Notes: Students were allowed to select more than one device.
The percentage represents the percentage of respondents.

2) **Please state your level of agreement with the following:**

The purpose of the evaluation was clearly explained.		
Strongly Agree	446	47.2%
Agree	430	45.5%
Undecided	58	6.1%
Disagree	8	0.8%
Strongly Disagree	3	0.3%
Total Responses	945	100.0%
Missing cases	15	

It was easy to access the evaluation from the invitation e-mail.		
Strongly Agree	551	58.6%
Agree	337	35.9%
Undecided	31	3.3%
Disagree	17	1.8%
Strongly Disagree	4	0.4%
Total Responses	940	100.0%
Missing cases	20	

The evaluation questions were clear.		
Strongly Agree	484	51.8%
Agree	400	42.8%
Undecided	36	3.9%
Disagree	11	1.2%
Strongly Disagree	3	0.3%
Total Responses	934	100.0%
Missing cases	26	

The evaluation interface was user-friendly.		
Strongly Agree	513	54.9%
Agree	374	40.0%
Undecided	34	3.6%
Disagree	10	1.1%
Strongly Disagree	3	0.3%
Total Responses	934	100.0%
Missing cases	26	

Summary		% Strongly Agree / Agree
The evaluation interface was user-friendly.		94.9%
The evaluation questions were clear.		94.6%
It was easy to access the evaluation from the invitation e-mail.		94.5%
The purpose of the evaluation was clearly explained.		92.7%

3) **Please write any comments you have about your overall experience with the evaluation.**

Student comments are not provided here, but were provided to the Teaching and Course Evaluation Project Committee to inform recommendations.

2 RESULTS FROM SURVEY OF STUDENTS: TEACHING AND COURSE EVALUATION REPORTS

2.1 INTRODUCTION

At the beginning of the Fall 2013 term, students who were enrolled in courses that participated in the Summer 2013 Proof-of-Concept were invited to view the reports that contained aggregated results from the online teaching and course evaluations³. The purpose of this survey was to get feedback from students about the evaluation reports.

2.2 METHODOLOGY

The survey, which was administered by eXplorance Inc. from October 18th to October 29th, was sent to all students who were enrolled in courses that participated in the Summer 2013 Proof-of-Concept. Of the 1,297 students invited to participate in the survey, 180 responded. This represents a response rate of 14%. Assuming the sample is representative, proportions calculated on all respondents are accurate within $\pm 6.8\%$, 19 times out of 20.

2.3 RESULTS

The survey instrument and distribution of responses to the survey questions can be found in sections 4 and 5 respectively.

2.3.1 Feedback on the Presentation of Results

Students were asked about their satisfaction with the quality of presentation and the amount of information presented in the report. Among respondents:

- 88-89% said that they are very or somewhat satisfied with the readability, layout, format and organization in the course evaluation report.
- 69% feel that the amount of information presented in the evaluation report was just right.

2.3.2 Feedback on the Students' Interpretation of Results

Students were then asked about how easy they thought it was to interpret the different summaries presented in the report and whether they found the evaluation report guide⁴ to be useful.

- 70% of respondents read the evaluation report guide. Of these students, 86% found it to be useful in helping them understand the evaluation report.
- Over 80% of respondents said it was very or somewhat easy to interpret the bar chart and median,
- 78-79% said it was very or somewhat easy to interpret the frequency table and mean,
- 72-74% said it was very or somewhat easy to interpret the comparison of scores and mode, and
- 68% said it was very or somewhat easy to interpret the standard deviation.

³ Only certain reports were made available to students.

⁴ The report guide contains general guidelines for reviewing the course evaluation report. It contains definitions and explanations of the various components presented in the evaluation report.

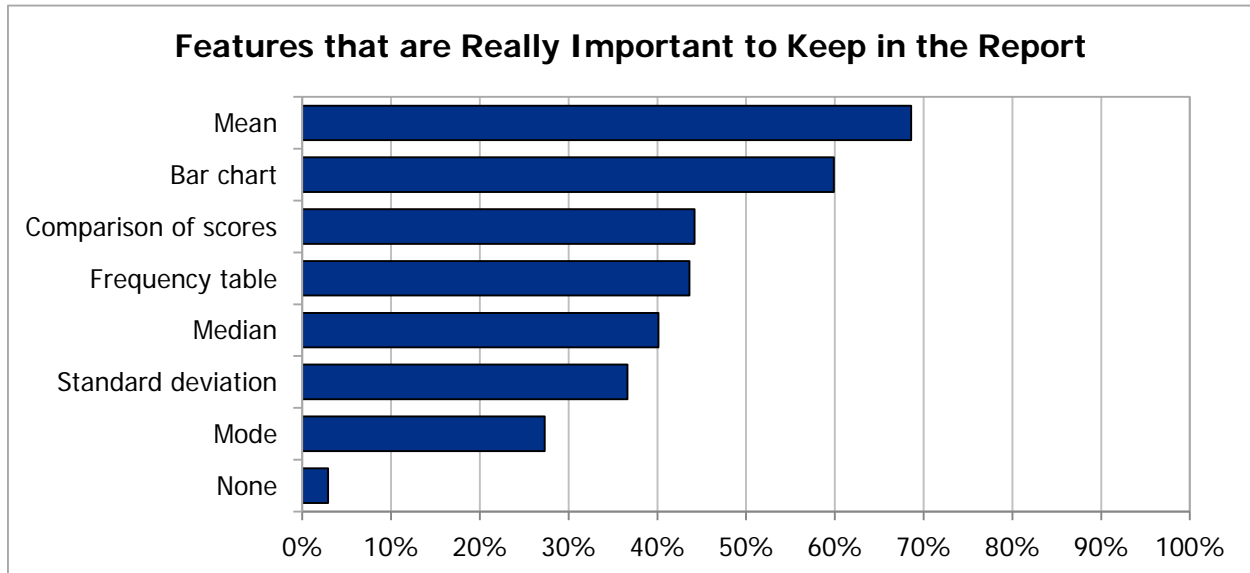
2.3.3 Feedback on What Should be Included in the Evaluation Report

Students were asked which types of summaries should be included in the report (Figure 1). The top two choices were:

- Mean (69%), and
- Bar chart (60%).

Fewer than 50% of respondents said that the other summaries are really important.

Figure 1



Students commented that they would like to see the results broken down by discipline and that the report should contain navigation links at the top for quick access to the section of interest.

2.4 SURVEY INSTRUMENT

- 1) How satisfied were you with the quality of each of the following characteristics in the course evaluation report?
(Scale: Very Satisfied, Somewhat Satisfied, Not Very Satisfied, Not at all Satisfied, Undecided)
 - a. Readability
 - b. Layout (position of tables, graphs, texts, etc.)
 - c. Format (colours, fonts, etc.)
 - d. Organization
- 2) Did you find the guide useful in helping you to understand the evaluation report?
 - a. Yes
 - b. No
 - c. N/A: I didn't read it
- 3) How do you feel about the amount of information that is presented in each question?
 - a. Too Much
 - b. Just Right
 - c. Too Little
 - d. Undecided

- 4) How easy did you think it was to interpret each of the following?
(Scale: Very Easy, Somewhat Easy, Somewhat Difficult, Very Difficult, Undecided)
- Frequency table
 - Bar chart
 - Mean
 - Median
 - Mode
 - Standard deviation
 - Comparison of scores
- 5) Which of the following do you think are really important to keep in the report? Please select all that apply.
- Frequency table
 - Bar chart
 - Mean
 - Median
 - Mode
 - Standard deviation
 - Comparison of scores
 - None
- 6) Is there any additional information you would like to see in the course evaluation report?
- 7) Do you have any further comments about the course evaluation report?

2.5 DISTRIBUTION OF STUDENT RESPONSES

This section displays the distribution of student responses to the survey questions. The tables that follow show the number and percentage of students selecting each response to the questions. In cases where students could select more than one response, the percentage of all respondents (percentage of those who answered the question) is provided. Please note that percentages in these tables are rounded to one decimal place. Summaries that group categories (e.g. where "Very Satisfied" and "Somewhat Satisfied" are grouped) may result in percentages that do not exactly match the sum of the percentages displayed in the tables. This is an artifact of rounding.

1) How satisfied were you with the quality of each of the following characteristics in the course evaluation report?

Readability		
Very Satisfied	65	36.7%
Somewhat Satisfied	93	52.5%
Undecided	6	3.4%
Not Very Satisfied	12	6.8%
Not at all Satisfied	1	0.6%
Total Responses	177	100.0%
Missing Cases	3	

Layout (position of tables, graphs, text, etc.)		
Very Satisfied	72	41.4%
Somewhat Satisfied	83	47.7%
Undecided	6	3.4%
Not Very Satisfied	12	6.9%
Not at all Satisfied	1	0.6%
Total Responses	174	100.0%
Missing Cases	6	

Format (colours, fonts, etc.)		
Very Satisfied	80	45.7%
Somewhat Satisfied	75	42.9%
Undecided	6	3.4%
Not Very Satisfied	13	7.4%
Not at all Satisfied	1	0.6%
Total Responses	175	100.0%
Missing Cases	5	

Organization		
Very Satisfied	68	39.8%
Somewhat Satisfied	85	49.7%
Undecided	6	3.5%
Not Very Satisfied	11	6.4%
Not at all Satisfied	1	0.6%
Total Responses	171	100.0%
Missing Cases	9	

Summary		% Very / Somewhat Satisfied
Organization		89.5%
Readability		89.3%
Layout (position of tables, graphs, text, etc.)		89.1%
Format (colours, fonts, etc.)		88.6%

2) **Did you find the guide useful in helping you to understand the evaluation report?**

Yes	106	85.5%
No	18	14.5%
Total Responses	124	100.0%
Not Applicable: I didn't read it	53	
Missing Cases	3	

3) **How do you feel about the amount of information that is presented in each question?**

Too Much	23	13.0%
Just Right	123	69.5%
Too Little	13	7.3%
Undecided	18	10.2%
Total Responses	177	100.0%
Missing Cases	3	

4) **How easy did you think it was to interpret each of the following?**

Frequency table		
Very Easy	59	33.7%
Somewhat Easy	79	45.1%
Undecided	12	6.9%
Somewhat Difficult	20	11.4%
Very Difficult	5	2.9%
Total Responses	175	100.0%
Missing Cases	5	

Bar Chart		
Very Easy	65	37.4%
Somewhat Easy	79	45.4%
Undecided	12	6.9%
Somewhat Difficult	16	9.2%
Very Difficult	2	1.1%
Total Responses	174	100.0%
Missing Cases	6	

Mean		
Very Easy	66	37.9%
Somewhat Easy	70	40.2%
Undecided	12	6.9%
Somewhat Difficult	23	13.2%
Very Difficult	3	1.7%
Total Responses	174	100.0%
Missing Cases	6	

Median		
Very Easy	56	31.8%
Somewhat Easy	86	48.9%
Undecided	12	6.8%
Somewhat Difficult	20	11.4%
Very Difficult	2	1.1%
Total Responses	176	100.0%
Missing Cases	4	

Mode		
Very Easy	53	30.1%
Somewhat Easy	74	42.0%
Undecided	15	8.5%
Somewhat Difficult	32	18.2%
Very Difficult	2	1.1%
Total Responses	176	100.0%
Missing Cases	4	

Standard deviation		
Very Easy	45	25.9%
Somewhat Easy	73	42.0%
Undecided	14	8.0%
Somewhat Difficult	36	20.7%
Very Difficult	6	3.4%
Total Responses	174	100.0%
Missing Cases	6	

Comparison of scores		
Very Easy	44	25.3%
Somewhat Easy	85	48.9%
Undecided	16	9.2%
Somewhat Difficult	23	13.2%
Very Difficult	6	3.4%
Total Responses	174	100.0%
Missing Cases	6	

Summary

	% Very / Somewhat Easy
Bar chart	82.8%
Median	80.7%
Frequency table	78.9%
Mean	78.2%
Comparison of scores	74.1%
Mode	72.2%
Standard deviation	67.8%

5) **Which of the following do you think are really important to keep in the report? Please select all that apply.**

Mean	118	68.6%
Bar chart	103	59.9%
Comparison of scores	76	44.2%
Frequency table	75	43.6%
Median	69	40.1%
Standard deviation	63	36.6%
Mode	47	27.3%
None	5	2.9%
Total Respondents	172	

6) **Is there any additional information you would like to see in the course evaluation report?**

Student comments are not provided here, but were provided to the Teaching and Course Evaluation Project Committee to inform recommendations.

7) **Do you have any further comments about the course evaluation report?**

Student comments are not provided here, but were provided to the Teaching and Course Evaluation Project Committee to inform recommendations.



SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

TEACHING AND COURSE EVALUATION PROJECT FINAL REPORT

APPENDIX XIII: INCORPORATING CONTEXT INTO THE INTERPRETATION OF RESULTS

Prepared by:

Prepared by Jessica Tilley, Analyst; Joanne Quan, Analyst; Kiran Bisra, Analyst
Institutional Research and Planning, Simon Fraser University

NOVEMBER 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

TABLE OF CONTENTS

1 INTRODUCTION 1
2 PROPOSED SOLUTION 1
3 METHODOLOGY 1
4 RESULTS 2
5 FUTURE MODELS 3

1 INTRODUCTION

Teaching and course evaluation results are often presented in aggregated summaries, such as an overall average of student ratings to a set of questions. Aggregated scores alone may be difficult to interpret, and the comparability of scores, over time or between instructors or faculties for example, may be questionable in the absence of context - that is, the consideration of characteristics that describe the course and students. For example, research has shown that class size, class time of day and gender of the instructor can influence teaching evaluations (see Appendix II). Many of the factors that affect evaluation results are outside of the instructor's control, which leaves the results open to criticism and doubt.

The purpose of this exercise, undertaken by SFU's Office of Institutional Research and Planning, was to explore a method for taking these contextual factors into account when interpreting evaluation results. The final goal was to suggest a way of determining whether an instructor's evaluation results are in line with expectations, given the type of course taught, and the type of students in the course.

2 PROPOSED SOLUTION

One solution to the problem described above is to create a model that predicts what score an average instructor would get, given the set of contextual factors that are outside of their control. For example, what score would be expected by an average instructor, if the course was a required lab course, taught in person, with 80 students in it, of which 40% are lower level students and 60% are upper level? If the answer to that question were known, we would be able to determine whether an instructor teaching such a course in a particular term had evaluation results that were above expectation, below expectation, or within expectation (within a certain error of the predicted result).

The best way to create such a model would be to use existing course evaluation data. For example, once a new instrument has been in use for about a year, the data generated from evaluations in that time can be used to create and validate a predictive model. That model could then be used going forward. This approach would ensure that SFU is meeting the overall principle that evaluation results must be considered in the context of the course characteristics (see Appendix II).

3 METHODOLOGY

To illustrate this approach, an exploratory model was developed using the summer 2013 proof of concept data. There are several reasons why this example model cannot be used moving forward, but should instead be considered a prototype. The most important reasons are that constraints of time and data (the proof of concept data only included 18 classes and 943 evaluations, which is not enough information to build a robust model) necessitated some simplifying assumptions, and created a situation in which the results cannot be generalized. Nevertheless, this exercise demonstrates how future models could be helpful in providing a context for results.

The prototype model was constructed by using seventeen classes, and the efficacy of the model was tested on one class (Class X) to demonstrate how the model would operate in the future. A multiple regression was performed between overall student learning experience as the dependent variable and student demographics and classroom characteristics as independent variables. The analysis was executed using the software package Statistical Product and Service Solutions (SPSS).

A set of available factors was identified and considered for inclusion in the model. This set is not exhaustive, and is expected to grow in future iterations of this exercise. Two main categories of data emerged: student demographics and classroom characteristics.

Student demographics include:

- Age
- Gender
- Cumulative Grade Point Average (CGPA)
- Faculty
- Year level
- International status
- Whether the student was required to take Foundations of Academic Literacy (FAL; this variable acts as a proxy for English as an Additional Language status).

Classroom characteristics include:

- Instruction mode (i.e. in person, online)
- Campus
- Class time of day
- Class level (i.e. graduate versus undergraduate)
- Class size
- Class component (i.e. lecture, tutorial).

The outcome measure is Question 5 from the proof of concept institution-wide core questions. This question asked students to rank their agreement with the following phrase: “The instructor created an atmosphere that was conducive to learning.”¹ Responses were ordered on a 5-point scale ranging from “Strongly Agree” (value 5) to “Strongly Disagree” (value 1). For the purposes of the prototype model, this was treated as a linear scale².

Prior to analysis, factors which highly correlate with one another (e.g. age and year level) were identified and one of the two variables was excluded. Also, some variables had to be excluded due to the small size of the data set relative to the number of available factors. It is expected that larger data sets will allow for the inclusion of a greater set of variables. These decisions were made based on prior research and may not reflect the final model in future iterations of this exercise.

4 RESULTS

Four independent variables were found to have a statistically significant relationship with evaluation results. These are:

- FAL requirement
- Instruction mode
- CGPA
- Class size.

¹ Note that the original intention had been to use Question 6 from the institution-wide core questions: “Overall, the quality of my learning experience in this course was ... Excellent; Very Good; Good; Fair; Poor”. In the end, this question was not used, because its scale was presented in the wrong direction on the evaluation seen by students, and there was concern that the validity of the results may be compromised as a result.

² There are arguments for and against this approach. However, since this is how the data are used in practice (the scores are averaged as though they are from a linear scale), it was determined that this approach would be best.

The prototype model was applied to predict the score of Class X, which had been excluded from the model. The overall score on Question 5 (averaging over all proof of concept classes) is 3.95, which is higher than the average achieved score of 3.87 from Class X. This may initially give the impression that the instructor's teaching evaluation score is below expectation for Class X. However, the model, when applied to Class X, indicates that the achieved score falls within the 95% confidence interval of the predicted score. Thus we can conclude that given the specific classroom context, the instructor's evaluation score is within expectation.

5 FUTURE MODELS

The methodology used to build the prototype model was constrained by a short timeline and the small set of responses. Future models may have fewer constraints, which should result in a more robust solution. This section discusses some issues to consider in future models.

In the future, with larger data sets, instead of building a linear regression model, we can explore constructing a Multilevel Linear model. Multilevel or hierarchical models are those in which data are collected at different levels of analysis (e.g. students in a classroom). For example, the fact that individual students respond together and have the same exposure within a classroom means their evaluations are not independent of one another. Multilevel modeling can take these dependencies into account, enabling us to better assess the effects of student demographics and classroom characteristics on the level of overall student learning experience.

It is important to keep in mind that the dependent variable in such a model would have to be one of the institution-wide core questions, since it would need to be a question that appears on every evaluation. Care should be taken to include institution-wide core questions that are appropriate for this purpose.

The issue of missing responses will need to be addressed in future iterations of this exercise. By definition, students who do not respond to the survey cannot contribute to evaluation scores, but their presence in the class may still have an effect on the class dynamic, and the experiences of other students. A systematic statistical approach to this problem will have to be developed as part of final model construction.

Finally, the prototype model took into account a small set of student demographics and classroom characteristics. Future models could expand the set of variables and explore new relationships. For example, such a model could incorporate the impact on course learning experience of a student taking a course outside his/her major. Nevertheless, it is important to remember that there will always be some important factors that cannot be measured, and no model can achieve perfection. Instead this modeling process represents our best chance of incorporating contextual factors into the evaluation scores.



SIMON FRASER UNIVERSITY

STUDENT EVALUATION OF TEACHING AND COURSES

THE TEACHING AND COURSE EVALUATION PROJECT

BEST PRACTICES ON INTERPRETATION AND USE OF
EVALUATION DATA

NOVEMBER 26, 2013

Contact: Corinne Pitre-Hayes
604.765.4610
cpitreha@sfu.ca

DOCUMENT CONTROL

DOCUMENT INFORMATION

	INFORMATION
Document Owner	<i>Corinne Pitre-Hayes</i>
Project /Organization Role	<i>Project Leader</i>

DOCUMENT HISTORY

VERSION	ISSUE DATE	CHANGES
<i>1.0</i>	<i>Nov 20, 2013</i>	<i>First synthesis of best practice principles for SET at SFU</i>
<i>1.1</i>	<i>Nov 22, 2013</i>	<i>Revisions from core team review</i>
<i>1.2</i>	<i>Nov 26, 2013</i>	<i>Revisions from SCUTL feedback</i>

TABLE OF CONTENTS

1	INTRODUCTION	1
2	RECOMMENDED APPROACH.....	1
3	BEST PRACTICE PRINCIPLES FOR SET DATA AT SFU	1
3.1	POLICIES.....	1
3.2	DESIGN	1
3.3	IMPLEMENTATION.....	2
3.4	INTERPRETATION OF EVALUATION DATA.....	2
3.5	RESPONSIBLE USE OF EVALUATION DATA	3

1 INTRODUCTION

A key objective of the Teaching and Course Evaluation (TCE) project was to develop a best practices guide on the interpretation of data from student evaluations. The objective of this guide is to support the responsible use of student evaluation data. Supporting responsible use of evaluation data was a primary goal of the TCE project.

2 RECOMMENDED APPROACH

The original approach for developing a best practices guide was to create a single document that would be used by the entire SFU community. As a result of our findings during the project, it is recommended that a slightly different approach be taken. Rather than creating a generic technical manual for all stakeholders that may or may not be used, adoption of a set of best practice principles for using student evaluations of teaching and course data is recommended.

Additionally, it is recommended that these principles, once approved, be embedded in the implementation of a new system of student evaluations at SFU. The approach will be to ensure that coordinated, locally appropriate procedures that are aligned with these principles are put in place by each faculty, school/department, or common interest group as part of the implementation process. The emphasis will be on developing support materials that are meaningful and of practical use in the context of the rollout to each constituent group. It is expected that this approach will be of greater practical value and is likely to be more effective in supporting responsible use of evaluation data.

Section 3 below outlines 14 recommended best practice principles in five different areas for the use of data from student evaluations.

3 BEST PRACTICE PRINCIPLES FOR SET DATA AT SFU

3.1 POLICIES

Student evaluation of teaching and courses will be conducted based on consistent, standardized policies.

Effective policies are essential to ensure that, overall, evaluations are conducted in a consistent and equitable manner. These characteristics of consistency and equitability are necessary for valid and useful evaluation data. Additionally, consistent institutional policies will be adopted for storage of, ownership of, and access to the raw data and reports.

Institutional policies will address design, implementation and interpretation.

Institutional policies on student evaluations often focus primarily on implementation. However, many issues affecting validity are introduced during design and interpretation of results. For example, questions that are unclear or improperly constructed can be introduced at the design stage. Effective policies can be employed to mitigate such issues from arising. SFU policies will cover all stages of evaluation including design, implementation and interpretation.

3.2 DESIGN

Clear evaluation goals for student evaluation of teaching and courses will be developed and maintained at the institution-wide level, the faculty level, the school or department level, and at the instructor level.

Evaluation design should accurately reflect current pedagogical priorities and current research on student learning. A multi-level approach will enable a distinct focus on these different priorities at the institution-wide, faculty and school or department level. Faculty and school or department level questions will be designed by appropriate members of the given faculty or school/department.

The evaluation instrument will be designed according to rigorous theoretical and psychometric standards.

The questions in the evaluation instrument will be carefully designed and tested with the aid of structured support from appropriate groups such as the Teaching and Learning Centre (TLC) and Institutional Research and Planning (IRP). This will ensure that both the questions and the scales are understandable and answerable by students. "Noise" and bias will be reduced through careful design of questions that will be made available in a formal SFU evaluation question bank. The question bank will be utilized at all levels of the multiple-level approach. Support will be provided to instructors to assist with designing appropriate questions that will provide the desired results. A manual will be made available to support good practice in selecting/designing questions for the instrument at all levels.

3.3 IMPLEMENTATION

A consistent approach will be used for distributing, collecting, processing and storing evaluation results.

In order to ensure validity, evaluations will be implemented in a consistent manner across the institution. This consistency will extend to the process of distributing evaluations and collecting, processing, and storing results and comparable statistics.

Instructors and students will be provided with sufficient information regarding the evaluation process, an efficient means of providing feedback, and easy access to evaluation results.

There will be a coordinated effort by the institution to explain the value of course evaluations both to instructors and to students. Student response rates will depend, in part, on an effective appeal from the instructor to the student regarding the importance of evaluations. Promoting ways in which the evaluation results have been used to make improvements will also be utilized.

Evaluations will be administered using appropriate and standardized procedures.

Appropriate and standardized procedures will be established for the administration of course evaluations. The research suggests that comparisons - if made at all - should only be between instructors in the same or similar disciplines and between courses that share similar characteristics. Practices will be standardized regarding how comparisons between instructors or courses (if employed) are made. The privacy of instructors and students will be protected. The appropriate amount of data will be distributed to appropriate audiences. Appropriate and consistent procedures will be in place for storage of and access to data.

3.4 INTERPRETATION OF EVALUATION DATA

Issues affecting interpretation of evaluation data will be factored into the presentation of results.

Issues affecting the interpretation of evaluation data can be related to course demographics. Demographics will be included in evaluation reports and where possible, factored in to the presentation of data. A model will be created by IRP that accounts for potential areas of bias and this information will be incorporated into the presentation of evaluation data. Care will be taken to encourage the comparison of scores of comparable courses (e.g. comparable disciplines, similar course types, and similar student types). Care will also be taken when including comparisons with aggregate scores (with Department, Faculty or University scores) in the presentation of the data.

Training and support will be provided to assist administrators and faculty with interpretation of results.

Guidelines and education will be provided to assist administrators to understand statistical data (for example, the standard deviation, range or frequency of reported scores) to facilitate interpretation of evaluation results. Structured support will be provided to assist faculty with interpretation of the data and with the application of results to improve/enhance teaching. Structured resources will be provided to

support training in the interpretation of evaluation data. Support and materials will address different audiences.

3.5 RESPONSIBLE USE OF EVALUATION DATA

Student evaluation of teaching and courses data will not be used as the only source of data to evaluate teaching performance.

Student evaluation of teaching and courses will not be the sole source of data for decision-making around teaching performance. Other sources may include peer evaluation (classroom observation) and/or teaching portfolios. One additional source of data will include the option for instructors to include a brief narrative that would provide context to the evaluation and make note of any circumstances they think should be considered. This narrative will be included in reports used for summative purposes. Decisions based on student evaluations as one source of data will include trends over time.

Evaluation data will be a viable means of providing diagnostic or formative feedback to instructors.

The results of instructor level questions will be private to the instructor. These instructor-only questions will provide an opportunity to include questions of a purely formative nature that are very specific to the instructor and the course.

Evaluation results will be issued promptly at an appropriate time.

Evaluation results will be distributed **after** all grades have been submitted in a given term. However, once the grades have been submitted, results will be issued promptly.

Evaluation results will not be issued unless they meet minimum requirements.

Reports based on an insufficient number of students (or an insufficient response rate) will not be produced. A set of rules regarding when to issue reports will be created and followed.

Evaluation results will be shared with students.

Students will be given access to evaluation results. The details of which results are communicated to students may vary across faculties.