REPORT OF THE COURSE EVALUATION PROJECT TEAM

April 27, 2017

1.0 Introduction

In 2014, the Associate Vice-President, Academic established the Course Evaluation Project Team (CEPT) to explore the potential for a new, campus-wide, course evaluation[1] model (see Appendix 1). This initiative arises from a commitment to (a) update the mechanism to capture Waterloo student feedback about the quality of the student educational experience, and to (b) move toward a system where student feedback is one of several metrics for evaluating instructor performance. The CEPT's mandate was to work on the first initiative: updating the student feedback mechanism so that it aligns with current teaching and learning practices (most Waterloo course evaluation tools were developed in the 1980s). Recognizing that current course evaluation tools are measures of student perceptions, the proposed assessment tool is described as "student course perceptions" (SCP)[2].

Since 2014, the project team has reviewed the literature on course evaluation and conducted consultations across campus with representative stakeholders regarding the possible development and implementation of a new course evaluation model. A draft report was produced on November 7, 2016. In Fall 2016, the team sought opinions from the campus community about its recommendations. A survey was run, with several open-ended questions about the proposed recommendations (see Appendix 2) and the preliminary question set (see Appendix 3). The literature review, extensive project team discussions, and results of the Fall 2016 consultation process have culminated in the recommendations in this report.

Important context

The project team recognizes the limitations of SCPs while also acknowledging the ways in which they serve an important function for university operation and success. Data from SCPs represent one source of evidence to be considered for promotion and tenure, and for annual performance review purposes. While it is beyond the mandate of the CEPT, the team strongly advocates for a subsequent university team being struck that continues the discussion about how methods such as peer evaluation, teaching dossiers and other approaches can be applied in a consistent, systematic manner campus-wide to evaluate teaching, course design and delivery. These other sources of evidence of teaching and course quality should take on a substantially enhanced role (see Policy 77). In order for SCPs to be credible sources of information, they must be validated and recognized as student perceptions of teaching effectiveness and the learning experience in a course.

---

[1] The term "course evaluation" is commonly used in the research literature and on many Canadian university campuses.

[2] Alternative names include Student Course Feedback, Student Course Evaluations, Student Evaluation of Teaching and Course, or Student Perceptions of Learning (SPLs), (among others). The final name would be determined in Phase 2.
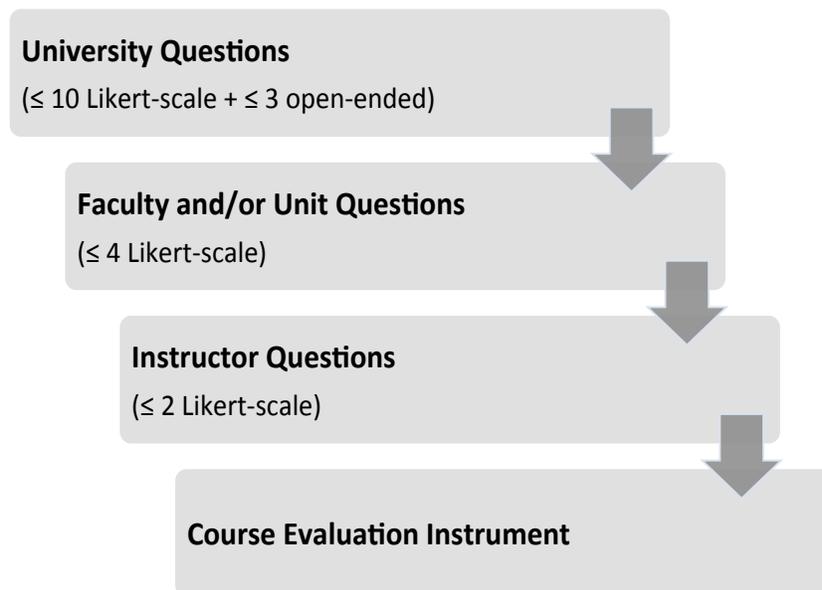
2.0 Key Research Themes and Findings
This section provides a summary of key research findings, organized by themes.

2.1 Reasons for a Cascaded Course Evaluation Model
The project team recommends the adoption of a cascaded course evaluation model. In this multi-level model, all Faculties include a common set of standard questions, complemented by optional additional questions chosen by each Faculty, academic unit, and instructor from an established, vetted, question bank. (See Figure 1).

Figure 1. Cascaded Course Evaluation Model

University Questions
(≤ 10 Likert-scale + ≤ 3 open-ended)

Faculty and/or Unit Questions
(≤ 4 Likert-scale)

Instructor Questions
(≤ 2 Likert-scale)

Course Evaluation Instrument

Course evaluation practices and instruments are varied at Waterloo. Adopting a common set of university-wide course evaluation questions would enable us to report institutionally on a key component of our mission – student perceptions of their learning experience.

Institutional reporting is fully consistent with the growing expectation from government and by the public for transparency and accountability from our post-secondary, publicly-funded universities. The Ontario Undergraduate Student Alliance has called for increased student access to course evaluation data, and Waterloo's Federation of Students has publicly advocated for access to aggregate data. A December 2015 report by the Ministry of Training, Colleges, and Universities (MTCU – renamed "Ministry of Advanced Education and Skills Development" in 2016) identified course evaluation data in its list of additional metrics that could be used in advancing an outcomes-based funding model for post-secondary education in Ontario.[i]

This cascaded model also gives Faculties, departments, and individual instructors the ability to select additional questions for more customized feedback. Instructors may select different additional questions over time, such as when instructional practices are changed in a course. University-wide questions are common at other universities, as noted in a recent survey carried out for an MTCU-funded research project on evaluations of teaching. More than 90% of the Ontario universities surveyed (n=20) had institution-wide student evaluations of teaching[ii]. In addition, Canadian universities of comparable size and prominence have already moved to a cascaded course evaluation model (e.g., Toronto, McGill, Simon Fraser).

2.2 Evaluation Instrument Design Principles
The evaluation model is structured on a set of guiding principles. The primary principle is that SCP questions need to connect to a well-grounded, empirically informed definition of effective teaching.

The project team's review of research into the elements of effective teaching[iii] shows that effective instructors design and deliver courses that result in meaningful student learning[iv]. While course evaluations do not measure student learning (that is the role of tools such as assignments, tests, and exams), students can provide useful feedback about how well the design and delivery of a course facilitated their learning (or not) and affected their learning experience.

This understanding - together with reviewing course evaluation instruments[v] used elsewhere and the current literature about such instruments[vi] - allowed the project team to identify three main dimensions through which students provide feedback by way of the instrument: Course Design, Course Delivery, and Learning Experience[3].

An analysis of Waterloo's in-use course evaluation instruments has revealed that few questions explicitly focus on the student learning experience. The questions also privilege lecture-based instructional practices[vii], and there is considerable variety in the wording and number of questions asked. Instructional and assessment practices have shifted over the past few decades to embrace an expanded repertoire of options (e.g., collaborative learning, active- and problem-based learning, authentic assessments). There has also been increased focus on learning outcomes and the use of educational technologies[viii]. Evaluation instruments that capture this evolution in course design, delivery, and the student learning experience are considered necessary and valuable.

Finally, we wish to reinforce the important message that student course perceptions (SCPs) represent one of several lines of evidence, each of which plays an important and complementary role in establishing a complete picture of effective teaching and the learning experience. (See Figure 2).

---

[3] See Appendix 3, which demonstrates the link between these three principles and a preliminary question set.

Figure 2. Three complementary lines of evidence
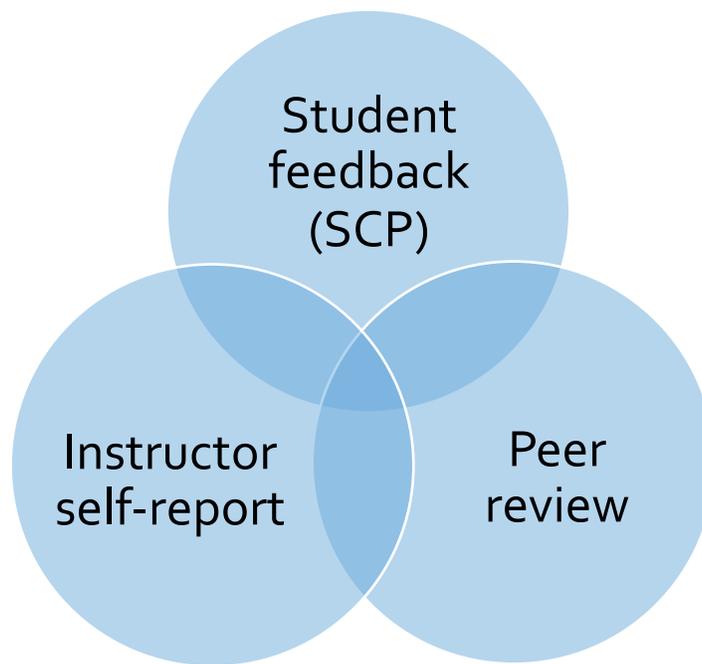
Accordingly, the project team recommends using the following principles to guide the development of evaluation questions:

1. SCPs should focus on students' perceptions of the quality of course design, course delivery, and the learning experience
2. SCPs should be designed to provide instructors with helpful, timely student feedback
3. unit chairs/directors should be able to use trends evident in successive SCPs as one means to help ensure high quality teaching for their academic programs
4. results from scaled questions should be viewed as student perceptions of teaching and a reflection of their learning experiences that may be further illuminated by open-ended comments
5. the selection of indicators of effective teaching and the wording of instrument items should be guided by the research literature as well as by ongoing assessment of evaluation instruments
6. evaluation questions should focus on instructional elements that students can reliably evaluate and avoid ones they cannot reliably evaluate[ix]
7. institution-wide questions should transcend course delivery formats and disciplines, and
8. the instrument should allow for the assessment of diverse teaching approaches with a combination of open-ended and Likert scale questions.

2.3 New and existing model use

Waterloo's [Policy 77](#) states that "student evaluations are an important source of information" in the assessment of teaching. However, teaching at Waterloo is assigned to a broader community than faculty and therefore the SCP process needs to consider the entire instructional community. "Instructor", as used in this report, includes all tenured and tenure-track faculty, adjunct appointments, lecturers, sessional instructors, and teaching assistants (TAs) who are in independent instructional roles.

Tenured professors and continuing instructors would use the new evaluation model. Instructors whose start date is after the commencement of a new evaluation process should be assessed with the new evaluation instrument. Faculties should offer instructors whose start date was prior to the commencement of a new course evaluation process the option to be assessed with (a) the previous Faculty-based instrument or (b) the new campus-wide instrument, until they have been awarded tenure and promotion to associate professor or attained continuing status.

2.4 Supportive Online Platform

The project team recognizes the benefits of online delivery of course evaluations. These include:

1. lowering resource costs when compared with paper-based approaches
2. easing the work to analyze, share and post data
3. adding flexibility to accommodate a cascaded evaluation model
4. increasing security of student access, and
5. enhancing accessibility by campus stakeholders to the evaluation process and its outcomes.

A locally developed online delivery system – *eValuate* – has been under development in the Faculty of Science for several years. Five of our six Faculties (Applied Health Sciences, Engineering, Environment, Mathematics and Science) have fully deployed *eValuate* using their existing instruments, and the Faculty of Arts has conducted an extensive pilot. Faculties have reported that *eValuate* has largely met expectations, and this software has effectively become the *de facto* campus solution. The project team concluded that the most reasonable and beneficial course of action would be for all Faculties to adopt *eValuate*.

An advisory committee has been struck to provide input to priorities for the technical development of *eValuate*. This committee is accountable to the office of the Associate Vice-President, Academic (AVPA) and the Associate Provost, Graduate Studies (APGS), and provides reports to the University Committee on Information Systems and Technology (UCIST). If the cascaded model of the instrument is adopted, a full review of software requirements to support the model will be initiated. The technical advisory committee would assist with that review.

2.5 Management of the SCP System

This section of the report addresses several high-profile issues that have been raised in project team discussions and through discussions with campus stakeholders.

2.5.1 The Issue of Bias

The team recognizes that every opportunity must be taken to enhance the clarity of each question's intent, and to minimize the potential for inappropriate comments. Research acknowledges that socio-cultural variables, biases, the "halo effect" and other influences can affect course evaluation results[x]. For example, student participation in SCPs can be compromised by factors such as bias (e.g., gender and race) in perceptions of course and instructional quality; indifference to the exercise by students and/or instructors; immaturity of respondents; misunderstandings of the purpose and application of course evaluation results; and instrument questions that are inappropriate or simply cannot be answered in an informed manner by students, among other factors and variables. The inherent bias in evaluation tools is a strong reason for instructor evaluation to be multi-pronged (i.e., SCPs are conceptualized as one evaluation tool).

Institutional and individual bias regarding specific groups is a challenge that we face in our society and in higher education. There is no question that biases (e.g., sexism, racism, ageism) exist on any campus, and that these biases can be expressed in SCPs. For example, in a study carried out at the University of Waterloo, when students received low grades, they gave statistically lower overall ratings of quality (course and instructor quality ratings were combined) to female instructors than male instructors.[4]

These are serious issues that the project team has discussed extensively. While it is not possible to control individual behaviours and responses to SCPs, it is possible to reduce the potential for bias, in its many potential forms, through careful design of the instrument. In addition, a common set of questions, and data in electronic form, can provide tools to investigate, recognize and address bias, better than is possible with different sets of instruments that cannot be aggregated for broader trends and outliers.

Similarly, we note that bias can be a factor in the interpretation of SCP data by university administrators (e.g., academic chairs/directors, staff). A realizable action will be to provide educational opportunities for those who use the data. (See Section 2.5.2.)

As noted already, the project team advocates for a multi-pronged evaluation of instructor performance and an investigation of the impacts of bias. Finally, a follow-up investigation of other assessment and evaluation methods, in addition to SCP information, would be a worthy undertaking. Other methods likely will require significant resources to scale up for department, Faculty, and or campus-wide use.

---

[4] See Endnote x.

2.5.2 Designing Support for Evaluation Instrument Users

While it is impossible to anticipate every potential factor that could compromise the quality, validity and fairness of evaluation responses, a properly designed and implemented training and orientation program can enhance the utility and validity of these evaluations.

Many universities have designed and implemented training and education programs for students, staff and instructors to support and guide the course evaluation process. Accordingly, all University students, faculty, staff (Faculty and departmental) administrators and system administrators should be trained in, and oriented to, the SCP and the use and interpretation of results. In addition, there should be orientation to the *eValuate* platform. Training and orientation content should comprise a generic core of information, plus material that meets the information needs of specific evaluation users. These information needs should be determined following consultation with each evaluation user group. Showcasing, and potentially sharing, the data analysis already occurring in the Faculties (e.g., trends against class sizes) could be another beneficial part of the education program.

With regard to the SCP, training and orientation content must address issues such as the intent of this evaluation tool, how and why these evaluations are used, how to interpret the results, the need to acknowledge the importance/role of bias (especially concerning gender and race) when completing and interpreting evaluations, and ethical obligations generally. In terms of the *eValuate* platform/technology, training and orientation content should explain the key features of the *eValuate* system and provide links to useful on-line resources (e.g., FAQs, instructional videos) that meet the needs of different user groups.

Training and orientation content should be accessible "on demand" via a single, dedicated on-line portal which would also enable access to the *eValuate* SCP and useful resources. Mandatory training and orientation content should be presented as a "toolkit," with online sub-sites dedicated to specific SCP user group information needs.

2.5.3 Testing, Monitoring and Evaluation: Instrument and Toolkit

The project team recognizes that validation of the SCP is needed. The testing of the instrument results will determine the reliability and validity of the instrument, including the influence of variables that could bias results at Waterloo. The results of this testing should be used to revise the instrument and/or the educational toolkit as appropriate before and following implementation to determine the influence of such variables at Waterloo.

Refinements to the SCP instrument should be made as necessary, following consultation with key campus stakeholders (including FAUW, GSA, and Feds) and regular expert review of operations and instrument design and performance. Further, a full assessment of the instrument and platform should take place after five years of campus-wide application, with monitoring and evaluation findings reported to Senate annually.

The Office of the Associate Vice-President, Academic (for undergraduate courses) and the Associate Provost, Graduate Studies (for graduate courses) should be responsible for oversight, coordination and reporting of campus-wide SCP assessment through the Quality Assurance Office, with consultation as required from the Centre for Teaching Excellence (CTE). Support for the technical use of the *eValuate* software would be provided by Information Systems and Technology (IST), and when required by the Centre for Teaching Excellence (CTE). Quality Assurance Office staff, along with the developers of *eValuate* (Science Computing) and IST staff, should determine an optimal strategy to ensure appropriate resourcing (sufficient capacity and operational support, user training and support) for *eValuate* for campus-wide use.

Quality Assurance Office staff should also monitor the performance of the SCP instrument and platform on a term-by-term basis, and report findings annually to Senate via the Senate Undergraduate Council (SUC), the Senate Graduate and Research Council (SGRC), and the Course Evaluation Advisory Group co-led by Science Computing and IST.

2.5.4 Data Management

The ownership of SCP data is an important issue. As such, information generated by the SCPs must be managed carefully. The collection, analysis and dissemination of student SCP data must be carried out in accordance with best practices concerning privacy of information, transparency and accountability. The collection, analysis and dissemination of evaluation data must adhere to privacy of information, transparency, and accountability in accordance with Policy 46 (Information Security).

Numeric information should be made accessible after authentication by the WatIAM system and should be available at the individual course level. These data should provide information generated by the core questions. The SCP data should present information to facilitate comparison with Faculty-wide ratings and program-specific ratings as determined to be statistically appropriate. Instructors should have access to all of their numeric information. The numeric results from these evaluations should be part of the instructor's record for annual performance review, and for tenure and promotion purposes. Written comments from students are intended for the instructor's use only.

Optional questions regarding TAs should be shared with the TAs; instructors are encouraged to engage in discussions about these results with TAs. Instructors, at their sole discretion, may use the written comments when seeking feedback and improvement. For example, they may show some or all of the comments to members of the Centre for Teaching Excellence (CTE) or colleagues when seeking advice about improving their teaching technique or course design.

2.5.5 SCP Administration Process
The project team believes that best practice for administering the SCP instrument includes the following obligations:

1. Provide students with information about the instrument at the outset of each term in each course so that they are aware of the type of feedback that will be requested. This information should be included in the course outline
2. Orient students to the purpose and applications of the SCP with reference to the toolkit
3. Conduct the SCP during the last two weeks of classes each term
4. Set aside approximately 15 minutes for in-class evaluation (for face-to-face courses); and
5. Close access to the SCP before the start of the exam period.

3.0 Recommendations:[5]
This set of recommendations reflects (a) the evolving analysis carried out by the project team, as represented in successive draft reports; (b) considerable debate amongst team members about key issues and responses; (c) the perspectives of stakeholder groups who were briefed about the project; and (d) the suggestions provided by individual respondents and groups in the Fall 2016 campus consultation program.

3.1 Teaching Evaluation
3.1.1 Student course perceptions (SCP)
- All UW course-based learning experiences, in all formats, should be evaluated.

- Students have a unique perspective to contribute regarding the course learning experience and, as such, their feedback should be solicited as part of the evaluation of teaching.

- The recommended nomenclature for this exercise is *Student Course Perceptions (SCP)*.

3.1.2 Use of complementary evaluation methods
- SCP results should be considered one of several potential data sources for annual performance appraisals, and for tenure and promotion purposes.

- As a priority, the university should explore the potential uses of additional, complementary teaching evaluation methods.

- The university should promote the use of additional teaching evaluation methods (e.g., peer evaluations, teaching dossiers, etc.). These complementary methods must be used in a consistent manner across campus.

---

[5] Note that full consensus by team members was not possible on all recommendations. It is understood that these Recommendations may be refined following testing in Phase 2.

- Faculties should decide which and how often complementary evaluation methods should be used.

- Significant investment in training for instructors, chairs/directors, and relevant staff should be allocated to ensure consistent and effective use of all evaluation tools.

3.1.3 Teaching Quality Improvement
- Triangulation of teaching evaluation methods (i.e., student course perceptions (SCP), peer evaluations, teaching dossiers, and/or other methods) should be used.

- The resources and expertise of the university's Centre for Teaching Excellence (CTE) should be promoted and endorsed as a valuable and effective resource to help instructors enhance their teaching effectiveness (e.g., through workshops, individual consultations, etc.).

3.1.4 Tenure and tenure-track status
- Tenured professors, continuing instructors, and instructors whose start date is after the commencement of a new evaluation process should be assessed with the new SCP instrument.

- Instructors whose start date pre-dates the commencement of a new evaluation instrument could decide to use the new evaluation instrument, or the previous instrument used in the respective Faculty.

3.2 Information Management
3.2.1 Information Management – Instructors
- Instructors should have access to all numeric information generated for their courses by SCP exercises.

- The numeric results from SCP results (i.e. core questions, Faculty and instructor-selected questions) should be part of the instructor's record for annual performance review, and for tenure and promotion purposes.

- Instructors should have the opportunity to place into context SCP results when used for annual performance reviews and for tenure and promotion files.

- Instructors, solely at their discretion, may share SCP data (for example, when seeking feedback and advice).

- Individual Faculties should follow standard and uniform protocols established by the university concerning interpretation of SCP information, and about levels of access by Deans, chairs and directors to SCP comments data.

3.2.2 Information Management – Student Access
- Summary/overall ratings about individual courses (i.e. core questions) should be made available to all students.

- The names of individual instructors should not be listed or accessible.

- Access to SCP results should be limited to those members of the university community with WatIAM credentials.

3.2.3 Managing Offensive Comments
- The *eValuate* system should be designed to screen the Comments section for potentially offensive words or phrases. These comments should be eliminated from the Comments content that is reviewed by the instructor. The numerics (data) would also be deleted in these cases.

- In the event that *eValuate* would not have the capacity to scan for offensive comments, alternative means would need to be identified and implemented.

- A bank of offensive words or phrases that could be present in the Comments section should be developed with advice provided by FAUW, Feds, GSA and SWEC.

- Anonymity of responses and comments should be ensured, with the exception of comments considered threatening, in which case *eValuate* should be used to filter these comments and to identify the student. Relevant university policies would then be applied (e.g., Policy 33, 46 and/or 71; Guidelines for Managing Student Information) to be administered by the appropriate Faculty associate dean (undergraduate studies).

- Quality Assurance Office staff should be the responsibility centre for this oversight role.

3.3 Instrument Design and Analysis
3.3.1 Cascaded Model Design
- The three-level cascaded model (i.e., core, course-based and Faculty questions) should be implemented campus-wide.

- Core questions in the student course perceptions (SCP) instrument should identify common elements of effective instruction (i.e., the three dimensions of effective instruction: course design, course delivery, and learning experience).

- Decisions about the selection of complementary questions should take place at the Faculty, program and/or instructor level.

- Complementary questions should be drawn from a bank of validated questions maintained by the Quality Assurance office.

- The cascaded model should be mandatory for summative (end-of-term) evaluations, and used at the instructor's discretion for formative (mid-term) evaluation purposes.

3.3.2 Number and Types of Questions

- The SCP question set should be finalized following extensive testing and refinement with a representative sample of students and instructors.

- Testing should include examination of the potential for bias in question choice, phrasing and sequence.

- The potential for, or evidence of, bias in instrument question design and responses should be a key element of pre-launch testing and of future assessments.

- The university's Survey Research Centre (SRC) should be engaged to provide advice and help to design the instrument and to manage the testing process.

- An optional question regarding experiential learning should be available in the bank of complementary questions.

3.3.3 Compatibility with existing instruments

- The cascaded student course perception (SCP) model structure should support and extend past and current data capture efforts used in previous course evaluation instruments.

3.3.4 Differentiation between Course and Instructor

- The instrument design should distinguish between course design and delivery elements to address cases when instructors teach courses they did not design.

3.3.5 Analysis of Numeric Data

- Statisticians and data visualization experts on campus should be consulted to determine how best to analyze and represent numeric data.

- Numeric data analyses should include reports on trends once sufficient data have been collected.

3.4 Instrument Implementation

3.4.1 Training and Orientation Toolkit

- A toolkit should be developed to support the SCP instrument . When additional evaluation methods of teaching are implemented, the toolkit should also support these methods.

- The toolkit should emphasize how to interpret the numeric data and comments in the context of the SCP's limitations.

- All students and faculty should be required to complete an on-line training and orientation module before use of the instrument.

- Experts (i.e., educational psychologists) should design the toolkit.

- Instructors (and the toolkit) should convey to students that course contexts differ (e.g., the time to return machine-scored tests will not be the same as for essays, tests or assignments).

- In-class explanations of course design and intent should be provided to promote clarity and understanding.

3.5 Monitoring and Evaluation
3.5.1 Role of Quality Assurance Office
- The Quality Assurance office should monitor the SCP instrument and toolkit on an annual basis.

- The Quality Assurance office should provide annual reports to Senate about the status of the SCP instrument and toolkit. These reports should also be provided to Senate Underground Council (SUC) and the Senate Graduate and Research Council (SGRC).

- The Quality Assurance office should carry out a full assessment of the SCP instrument and toolkit on a 5-year cycle.

- Refinements to the SCP instrument and toolkit should be made in consultation with key campus stakeholders (i.e., FAUW, Feds, GSA, SWEC).

4.0 Next Steps: Phase 2

The Vice-President Academic and Provost will determine next steps as they relate to SCPs at Waterloo. If the decision is made to proceed to fully develop the SCP framework, then the following major tasks would need to be accomplished:

- Develop, test, refine and validate a question set (both core and optional)
- Design and test the training/orientation toolkit
- Test the *eValuate* software and platform to ensure delivery capability using a cascaded model
- Pre-launch test of the prototype in its entirety (i.e., question set, toolkit and *eValuate* platform.

This work would require the creation of a new project team and sub-groups. It is likely that these project elements would require at least one year to complete. The project could require hiring a project leader and possibly staff resources to conduct research, develop, and test the question set and toolkit.

If there is approval to pursue a cascaded evaluation model, a sub-committee should be struck and user testing and survey validation should be undertaken on the core questions. Items should also be developed for the additional question bank (i.e., the Faculty/academic unit and instructor questions). The prototype instrument should be field tested through pilots, the results of which would be used to change, refine, and finalize the question set.

The *eValuate* project team and IST will need to work closely with Phase 2 project colleagues to identify and explore issues and opportunities for system design.

There will also be a need to keep the campus community informed regularly about project progress. This communication could include regular briefings for Senate and for campus interest groups. The prototype SCP framework would need to be reviewed by key university-level decision-making bodies (e.g., SUC, etc.) and Senate for their review and approval.

Report submitted by:
Mark Seasons
Chair
Course Evaluation Project Team

April 27, 2017

Appendix 1 – Course Evaluation Project
Project Mandate and Process

In 2014, the Associate Vice-President, Academic established the Course Evaluation Project Team (CEPT) to explore the potential for a new, campus-wide, course evaluation model. For the past two years, the project team has reviewed literature on course evaluation and conducted consultations across campus with representative stakeholders regarding the possible implementation of a new course evaluation model. Specifically, the project team was mandated to accomplish the following:

1. Examine the various administrative, logistical, technological, and cultural issues pertaining to course evaluations at the University of Waterloo
2. Establish best practices concerning all aspects of course evaluations based on a review of the literature
3. Consider the implications of adopting changes to current course evaluation procedures in relation to Policy 77, the MOA (Memorandum of Agreement with the Faculty Association), and faculty annual performance evaluations
4. Assess the feasibility of designing a common institutional survey instrument, with customizable sections at the Faculty, department, or instructor level (referred to here as a "cascaded" model)

The project team is composed of representatives from the major stakeholder groups at the University: faculty representation (academic Faculties; Faculty Association of the University of Waterloo – FAUW); undergraduate students (Federation of Students - Feds); graduate students (Graduate Students Association – GSA); academic support units (Centre for Extended Learning – CEL; the Centre for Teaching Excellence – CTE); and the University's Information Systems and Technology group – IST.

The project team and its subgroups have met regularly since May 2014. Recommendations have been informed by the review of appropriate literature, consultations with colleagues at other universities, and the review of a number of peer university websites to identify best practices and factors to consider when designing, implementing, and interpreting course evaluations.

In addition, the team has carefully considered perspectives and advice offered by the university's AccessAbility Services (AAS), the Office of the President (Special Advisor on Women's and Gender Issues), as well as subject matter specialists, including social psychologists, survey design methodologists, and teaching fellows at Waterloo. A consultation process was undertaken throughout 2015 with the Senate, Deans Council, FAUW, FEDS, GSA and all six Faculties (Applied Health Sciences, Arts, Engineering, Environment, Mathematics, Science). Three key concerns emerged from these consultations: (1) inherent biases in course evaluation, (2) the advisability of university-wide questions, and (3) the privacy of and access to data. The recommendations in this report address these and related concerns.

Appendix 2 – 2016 Survey Process and Results
Survey Design and Management:
In Fall 2016, the course evaluation project team decided that it was time to seek opinions from the campus community about the project team's recommendations. Accordingly, a survey was designed with the following open-ended questions that related to key aspects of the proposed approach to course evaluation:

1. What are the advantages and disadvantages of Waterloo adopting a cascaded model for course evaluation?
2. How well do the sample questions align with the instrument design principles outlined in this report?
3. What are the advantages and disadvantages of access to course evaluation information as presented in this report?
4. What other comments do you have about the recommendations and information presented regarding course evaluations at Waterloo?

The survey launched on November 8, 2016. Two emails were sent from the office of the Associate Vice-President (Academic) to introduce the survey and encourage responses (November 8, 2016 and December 9, 2016). In addition, the survey was highlighted in the university's Daily Bulletin (see November 25, 2016) and by the Registrar's Office (December 9, 2016). All communications to the campus community included a hot link to the survey that was posted on the Associate Vice-President (Academic)'s website (see: https://uwaterloo.ca/associate-vice-president-academic/course-evaluation-project). All survey responses were collected anonymously.

The survey was available on-line for campus stakeholder response from November 8, 2016 to January 20, 2017. More than 90 individual responses to the survey have been received to date, as well as written submissions from several academic units and campus organizations. These responses have been organized for analysis in an Excel spreadsheet managed by Quality Assurance staff in the office of the Associate Vice-President (Academic).

General Impressions:
Overall, the comments and suggestions are supportive of the recommendations presented in the Draft Report. However, positions vary widely and some appear fixed regarding specific issues. We can state that, in general, students want as much information as possible about the learning experience and teaching effectiveness. We note that specific groups, such as FAUW, seek strict controls on student course perception data use and access.

The majority of critical comments concerned the issues of bias (e.g., gender, race, etc.) in the context of teaching evaluation; the capacity of students to assess teaching quality; the proposed question set (i.e. number and types of questions); and access to student course perception data, specifically the written comments, by students and department chairs/directors. Comments were made about the validity of the concept of teaching evaluation and the utility of an orientation/training toolkit.

Areas of commonality/consensus included:
- The cascaded model (tri-level: core, course and Faculty-level questions)
- Use of the *eValuate* course evaluation software and platform
- The merits of a comprehensive system of evaluation of teaching beyond the sole source of student feedback
- The need to test and validate the question set before launch
- Recognition of bias, in its many forms, and implications for teaching evaluation
- The need to build upon historical databases of past course evaluations
- Restrictions on access to SCP data

Areas of difference of opinion/divergence of perspectives included:
- The proposed question set – number and types of questions
- Whether the complex issue of bias could be addressed effectively through instrument design and user training
- Whether an orientation/training toolkit would be an effective way to deal with bias
- Whether students should evaluate teaching effectiveness
- Whether instructor names should be accessible in evaluation databases
- Whether students and/or academic department chairs/directors should be able to access students' written comments

Summary of Organization/Group/Departmental Submissions:

Please see Table 1.

Table 1: Summary of Group/Departmental Submissions

Organization/ Group/Dept.	Comments
English	Concerns about the phrasing of draft questions Comments about what constitutes "reasonable amount of time" re: assignment return Concerns about the concept of "clear communicator" Bias: need to map out questions and assess re: bias; ensure that toolkit designed with bias as key issue; how might bias be addressed by chairs? Data management: are we taking sufficient steps to ensure privacy?
FAUW	Cascaded model: the value of campus-wide core questions is questionable, given Faculty cultures Sample questions: need to clarify use and purpose of evaluations before finalizing questions Position: valid to have student perceptions; not for course design or quality of teaching Access to information: agrees that evaluation comments for instructor alone; Faculties should decide whether to make numeric scores more widely available Additional comments: bias remains a significant concern; rejects position that potential for bias can be managed Recommendations: clarify that this is about student experiences; should not be used to evaluate teaching for merit or T+P; support consistent use of any new model
Feds	Overall, Feds is supportive of draft report (November 7, 2016); while areas of concern, happy with proposed processes and strategies Bias issue: acknowledges potential; notes student evaluations part of overall assessment process; students' assessments are an integral and necessary aspect of course evaluation Cascaded model supported Numeric data: recommends campus community should have access to these data Mid-Term evaluations: Feds policy calls for both formative and summative course evaluations TAs: encourages greater attention to role of TAs, important role played; should be properly evaluated

Feke et al.	Agreed with much in report – e.g., cascaded model Issue: systemic biases (explicit and implicit) – grave concerns; concerned that no concrete solutions provided in report, so broader conversation is needed Opinion that student training program would not properly address potential for bias Student course evaluation information should not be used in merit, T+P; a new, more equitable method is needed (e.g., peer teaching evaluations)
GSA	In agreement with CEPT report (November 7, 2016) Acknowledges commitment and effort of project team, "extraordinary effort" to reach consensus Cascaded model: could provide meaningful information re: cross-campus; must watch for biases and students' reliability in responses Sample questions: notes that province's outcomes based funding model will likely use course evaluation data Course design: noted that students not in position to comment knowledgeably about course design; rather, should be interpreted as expression of students' perceptions of course Course delivery: some elements are universal across campus – e.g., timeliness, clarity of communication, etc. – while others could be context-specific Learning experience: open-ended questions (comments) are important, a valuable tool Training: support for mandatory training plus in-class messages/explanations of importance of evaluation Privacy of information: expect that UW privacy policies and report recommendations re: access to data applies to TAs as well Ongoing monitoring and evaluation of instrument will be essential; each stakeholder group must feel respected; academic freedom and integrity must be upheld Belief is that students and their organizations want access to student perceptions of the quality of their course experiences; use to compare with experiences elsewhere. Data should exclude reference to ratings/rankings of instructors Evaluation of courses/teaching quality should be based on multiple evaluation methods

MATH	Support for use of evaluations to provide meaningful feedback Cascaded model generally supported Choice of questions is a key issue. Should have 4-6 core questions maximum. MATH wants role in choice of core questions; these questions should be applicable across disciplines Strong desire to maintain historical trend data – any new instrument would need questions that retain, or relate clearly to, existing ones Data management/access: more clarity required re. internal vs. external use of data; numeric should widely accessible, while comments for instructor's use Student evaluations: the instrument should be part of overall teaching evaluation process
PSYCH	Extraneous, "biasing" factors make student questionnaires invalid for summative evaluation Summative use of student evaluations harms student learning and instructors' integrity and academic freedom Proposed remedies for bias will not be effective Student evaluations could be useful for formative feedback Experiences on other campuses may not be relevant at Waterloo; need to be careful about "best practices" Alternatives to student questionnaires can generate less bias and do more to promote effective instruction Questionnaire design should be informed by on-campus expert advice
SWEC	Focus should be on student perceptions of course and their learning Student feedback data should not be published Should investigate whether discrimination apparent in past course evaluation results Future instruments should be evaluated regularly (i.e., annually) to support refinement of questions Training to minimize opportunities for bias should (a) be used if proven to be effective and (b) designed by experts Mechanisms should be in place to mitigate impact of sexist, racist, other inappropriate comments There is a need to (a) examine methods that could be used to assess teaching effectiveness of instructors and student learning and (b) review weighting/importance of student evaluations for merit and T+P

Appendix 3 - Dimensions of effective teaching and sample questions

Course Design Dimension	Course Delivery Dimension	Learning Experience Dimension
• I knew what I was expected to learn in this course • The graded work assessed what I was expected to learn • The course activities prepared me for the graded work • The coursework demands were…(Likert scale answer choices will reflect workload intensity – for example, very light to very heavy)	• The instructor returned graded work in a reasonable amount of time • The instructor was a clear communicator • The instructor created a supportive environment that helped me learn • The instructor stimulated my interest in this course	• The most important thing I learned in this course was* • Overall, I learned a great deal from this instructor • Overall, the quality of my learning experience in this course was excellent • What helped me to learn in this course?*
• What changes, if any, would I suggest for this course?*		

*Denotes an open-ended question.

References

i   Ministry of Training, Colleges, and Universities. (2015). Focus on Outcomes, Centre on Students: Perspectives on Evolving Ontario's University Funding Model. Toronto, ON: Queen's Printer for Ontario.

ii   Wright, A.W., Hamilton, B., Mighty, J., Scott, J., & Muirhead, B. (2014). The Ontario Universities' Teaching Evaluation Toolkit: Feasibility Study. University of Windsor: Centre for Teaching and Learning. 37

iii   Chickering, A.W. & Gamson, Z.F. (1987). Seven Principles for Good Practice in Undergraduate Education. AAHE Bulletin; Feldman, K.A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R.P. Perry and J.C. Smart (eds.), The scholarship of teaching and learning in higher education: An evidence-based perspective, pp.93-143. Springer Online.; Gravestock, P. & Gregor-Greenleaf, E. (2008). Student Course Evaluations: Research, Models and Trends. Toronto: Higher Education Quality Council of Ontario; Hativa, N., Barak, R., & Simhi, E. (2001). Exemplary university teachers: Knowledge and beliefs regarding effective teaching dimensions and strategies. The Journal of Higher Education, 72, 699-729.; National Survey of Student Engagement (NSSE). (2014). From Benchmarks to Engagement Indicators and High Impact Practices. http://nsse.iub.edu/pdf/Benchmarks%20to%20Indicators.pdf; Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. Review of Educational Research, 83, 598-642; Undergraduate degree level expectations for the University of Waterloo. Retrieved from https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/curriculum-development-and-renewal/program-review-accreditation/8-degree-expectations; Young, S. & Shaw, D.G. (1999). Profiles of effective college and university teachers. The Journal of Higher Education, 70, 670-686.

iv   Weimer, M. (2010). Inspired college teaching: A career-long resource for professional growth. San Francisco, CA: Jossey-Bass. 75

v   McGill University. (2014). Recommended pool of questions for courses and instructors. Retrieved from the world wide web on March 11, 2015: http://www.mcgill.ca/tls/files/tls/recommended_pool_of_questions_2014-final.pdf; Student Evaluation of Educational Quality (SEEQ) – example used was employed at the University of Manitoba. Retrieved from http://intranet.umanitoba.ca/academic_support/catl/media/seeq.pdf; University of British Columbia. (2008). University questions. Retrieved from http://teacheval.ubc.ca/introduction/university-questions/; University of Toronto. (2013). Course evaluation question bank. Retrieved from http://www.teaching.utoronto.ca/Assets/Teaching+Digital+Assets/CTSI+1/CTSI+Digital+Assets/PDFs/qb-nov13.pdf; Werhun, C. & Rolheiser, C. (2014, June). Students' motivation for participation in course evaluations outside of the classroom: The role of the instructor in online

course evaluations. Presented at the annual Society for Teaching and Learning in Higher Education conference, Queen's University, Kingston, ON.

vi   Gravestock, P. & Gregor-Greenleaf, E. (2008). Student Course Evaluations: Research, Models and Trends. Toronto: Higher Education Quality Council of Ontario; Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. Review of Educational Research, 83, 598-642.

vii   Marks, P. (2012). Silent partners: Student course evaluations and the construction of pedagogical worlds. Canadian Journal for Studies in Discourse and Writing, 24(1), 1-32. Retrieved from http://www.cjsdw.com/index.php/cjsdw/article/view/16/8

viii   Gravestock, P. & Gregor-Greenleaf, E. (2008). Student Course Evaluations: Research, Models and Trends. Toronto: Higher Education Quality Council of Ontario.

ix   Gravestock, P. & Gregor-Greenleaf, E. (2008). Student Course Evaluations: Research, Models and Trends. Toronto: Higher Education Quality Council of Ontario.

x   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? [Special issue]. New Directions for Institutional Research, 109, 9-25.; Gravestock, P. & Gregor-Greenleaf, E. (2008). Student Course Evaluations: Research, Models and Trends. Toronto: Higher Education Quality Council of Ontario; MacNell, L., Driscoll, A., & Hunt, A.N. (2014). What's in a name: Exposing gender bias in student ratings of teaching. Innovative Higher Education. DOI: 10.1007/s10755-014-9313-4; Ory, J.C., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? 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, 27-44.; Sinclair, L. & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. Personality and Social Psychology Bulletin, 26, 1329-1342.; Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. Review of Educational Research, 83, 598-642.; 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.