Recommendations from the Teaching Assessment Implementation Committee to the JCAA on the USAT

Committee Members

Elizabeth Hanson (QUFA – Co-Chair)
John Pierce (University Co-Chair)
Patricia Collins (QUFA Member)
Denise Stockley (University Member)
Terry Soleas (Grad Student)
Julia Göllner – AMS (Undergrad Student)
Leslie Jermyn (QUFA Staff)
Jackie Cleary (University Staff)

May 22, 2019

Table of Contents

Recommendations from the Teaching Assessment Implementation Committee to the JCAA on the USAT

Executive Summary

This document presents the recommendations of the Teaching Assessment Implementation Committee (TAIC) to the JCAA, following an extensive review of the USAT, a careful consideration of current practices in the evaluation of teaching and learning, and a contextualization of these matters in the context of Appendix E and Article 29 of the 2015-2019 Collective Agreement. The recommendations offer a change in the focus and design of the current USAT and a change in the mode of delivery from paper-based to electronic format. It should be noted that since its first development in 1994 as QUEST and the later change to USAT in the early 2000s, the questions and format of the survey have not changed. A review in 2007 included a number of recommendations for change, but these were not adopted. Thus, the recommendations in this document are part of the first serious review of the teaching survey in over a decade.

The transformation in overall approach taken by the committee appears in its first recommendation that the name of the USAT be changed to the Queen's Survey of Student Experience of Teaching (QSSET). The shift from the "Assessment of Teaching" in the former survey to a measure of the "Student Experience of Teaching" reframes the survey into an attempt to measure the student's participation in the course, the experience each student has of the instructor, the role of the course materials in shaping student academic experience, and the contexts of the timing of a course, the rooms, and the technological supports which all shape and potentially impact student learning. In this reconceptualization of the survey, the TAIC has done much more than simply revise the questions used. Instead, TAIC proposes a teaching and learning survey that is driven by clearly stated purposes, ranges across distinct areas of pedagogical experience with different sections on the Student, the Instructor, the Course, and Infrastructure, and is contextualized by documents outlining the best use of the survey by students, instructors, and Heads, Deans and RTP committees.

Background

Appendix E to the 2015-2019 Queen's QUFA Collective Agreement mandated a review of the USAT and an exploration of methods for the evaluation of teaching in courses not currently using USAT. Having completed this review, the TAIC proposes both a new questionnaire, described in the specific recommendations below that may be used across all types of teaching at Queen's, and principles for its use in conformity with Article 29 of the Collective Agreement. The proposed survey has been the subject of a Pilot in the Fall term of 2019, conducted across faculties in graduate and undergraduate, in-class, blended, and online courses. The recommendations below reflect the results of the Pilot. In developing the survey and the principles for its use, the TAIC also relied on the work of its predecessor committee, the Teaching Assessment Committee which had been established per Appendix E in 2016. The TAC

extensively reviewed literature about teaching assessment and available survey models. Its recommendations included the survey design now captured in Recommendation 3 below.

If implemented, the recommendations set out below would represent a paradigm shift in the evaluation of teaching at Queen's. At the same time, TAIC believes, they would bring our practices into closer conformity with Article 29 of the Collective Agreement by which the TAIC was guided in developing its recommendations. Specifically, the TAIC remained mindful of the distinction between the assessment and evaluation of teaching for personnel processes, which the article as a whole addresses, and which is to be performed by Heads, Deans and RTP Committees, and the survey of students described in 29.3, which serves as but one source of evidence to be weighed in that assessment along with other material that might be included in the teaching dossier outlined in 29.2 In other words, TAIC recognized that Article 29 does not intend for surveys of students to serve as proxies for the evaluation of teaching, noting, for instance, that 29.1.3 requires that the "assessment and evaluation of teaching" consider qualities such as "familiarity with recent developments in the field" to which students were not qualified to speak. Moreover, TAIC was aware that much of the scholarly literature surveyed by the TAC concluded that surveys of students did not furnish reliable direct evidence of teaching effectiveness. This point was also affirmed in William Kaplan's arbitration award with respect to Ryerson University (June 28, 2018). (Note: the Ryerson award was released after TAIC had taken most of its decision; we were, however, heartened to see that we had anticipated and addressed many of the concerns it articulated.) Nevertheless, the TAIC considered that students were important sources of information about the presentation of material, instructor availability to students and other aspects of their experience of instruction, as long as the responses they provided were appropriately contextualized.

The proposed survey is designed to provide assessors of teaching (Heads, Deans, RTP committees and any others charged with that task) with information about a range of matters that affect students' experience of teaching so that they may weigh student responses about teaching effectiveness appropriately. Specifically, the survey asks assessors to contextualize student responses about instructor effectiveness in relation to other information covered in the survey about the students, the course, and the material conditions of teaching. In order to ensure this contextualizing of survey responses, the data will need to be presented in a different manner than 29.3.5 currently specifies, as a distribution of responses rather than as means. The comparisons that means invited were always dubious, as Arbitrator Kaplan noted in the Ryerson award; their use in the current survey would render its design meaningless. While the contextualization of responses the proposed survey facilitates may make the task of evaluating teaching more painstaking, we are convinced that the results will also be fairer, and in this respect will support the desires of Heads, Deans, RTP committees and others. It is also worth noting that while the TAIC crafted its recommendations with the stated purposes of Article 29 in mind, the recommended survey design will afford information that may be useful for other institutional purposes. In 2017 QUFA and the University agreed to allow the Office of Institutional Research and Planning to analyze anonymized USAT results in order to better understand the correlations between student experience and resource allocation decisions. Should the Parties decide to revive this project, the proposed survey design will prove helpful.

2

Per Appendix E (2) the TAIC defined the two main purposes of the new assessment tool as follows:

1. To help instructors improve their teaching and
2. To help Heads, Deans, and others in the assessment and evaluation of teaching by Queen's faculty for annual/biennial performance review, contract renewal, tenure and promotion.

Recommendations

Recommendation 1: That the name of the USAT be changed to Queen's Survey of Student Experience of Teaching (QSSET)

The change in name reflects the nature and scope the survey itself. The redesigned survey is designed to capture the students' "experience" of teaching, and to distinguish their comments on their experience from the "assessment and evaluation" of teaching as performed by Heads, Deans and RTP committees.

Recommendation 2: That QSSET, per Appendix E 2 (a) and (b) be defined by two purposes:

1. To help instructors improve their teaching, and
2. To help Heads, Deans, RTP committees and others evaluate the job performance of Queen's faculty for annual/biennial performance review, contract renewal, tenure and promotion.

Recommendation 3: That the QSSET consist of four sections with the following purposes for each:

I.   Student Section

Purposes:
   i. To prompt students to reflect on their contributions to their experience of the course.
   ii. To permit the evaluation of survey responses in light of the students' contributions to their experience of the course.

II.   Instructor Section

Purpose:

    i. To elicit students' feedback about the Instructor's teaching, as established under the "Overall Purpose of the TAIC Questionnaire."

III.   Course Section

    Purpose:

    i. To elicit students' feedback on course elements which may or may not be performed by, or under the control of, the Instructor.

IV.   Course Infrastructure

    Purpose:

    i. To permit the evaluation of survey responses in light of students' experience of the material conditions (room, technology and schedule) of course delivery.

    Note: The Course infrastructure section may be used by the University to assess the conditions of teaching and learning spaces and may inform decisions with respect to renovation or creation of new teaching and learning spaces on campus. It may also inform decisions about technological support for instruction.

Recommendation 4: That there will be a comments section devoted to each of the four subsections of the survey.

The TAIC affirms that the comments sections of the current USAT often include valuable information for instructors to use in improving their teaching. The Committee has decided to include a comments box at the end of each of the four sections of the questionnaire and a section for final comments. The comments shall be reported per Article 29.3.7.

Recommendation 5: That the Instructor section be the only section of the survey that may be used directly by Heads, Deans, RTP committees, and others in assessing the instructor's teaching effectiveness, save where the Member requests that responses to questions in the Course section also be considered.

Because only the Instructor section addresses matters that the Instructor fully controls in all cases, only the responses to questions in that section afford evidence about student experience of the Instructor's teaching effectiveness and may be used as such. The additional sections may be used by the instructors to frame or qualify scores presented in the Instructor section and, where appropriate, by Heads, Deans, RTP committees, and others as context for understanding the responses to questions about teaching

effectiveness. However, some questions in the Course section may pertain directly to the Instructor's pedagogical work. Where the evaluator knows that the Instructor has performed the marking and/or developed course materials, for example, because in all courses at a certain level in that Unit marking and course design are performed by the Instructor, response to those questions may also be considered in evaluating the Instructor. Instructors may also clarify for evaluators where they have been responsible for marking and/or course design and the assembling of materials.

Recommendation 6: That the questions included in Appendix A will constitute the core QSSET.

The questions in Appendix A are those that were tested in the QSSET Pilot in November 2018, save for two questions which the Pilot disclosed to be problematic. One, concerning respect for social diversity, could not be revised to address the problem; the second, concerning accessibility outside the classroom, has been revised.

Recommendation 7: That the survey will include the statement indicating that comments that are disrespectful, demeaning, discriminatory, or harassing in nature are unacceptable.

There is anecdotal evidence that such comments are present in the current USAT. The TAIC affirms the University's efforts to create an environment that supports inclusivity and diversity. The TAIC recommends that a policy and process be developed for the exclusion of surveys which include demeaning comments, however, because such a process will depend on the capacities of the software through which QSSET is administered, the TAIC cannot recommend a specific process for reviewing and judging these comments.

Recommendation 8: That the survey will be conducted online and, in the case of on-campus courses, administered during class time.

While the survey must be administered electronically in the case of online courses, there are compelling reasons for administering all surveys that way. These reasons include the ease of presenting the data that the surveys yield, as well as the relationships among responses to different questions. Research has shown, however, that online administration of teaching surveys can lead to a lower response rate. To mitigate this effect, the TAIC recommends that in-class time (at least 15 minutes) be allowed for students to take the survey. For the Pilot, the survey was administered using Qualtrics and there were no difficulties for participants in accessing the survey. The on-campus participants showed a reasonably robust response rate, though absent data from the Registrar's office about the

USAT response rate in comparable courses, it is not possible to precisely demonstrate the effect of online administration. See Appendix B.

Recommendation 9: That a gender question will also be included as developed by the Teaching Assessment Committee and agreed to by the JCAA.

The TAC, meeting during the 2016-17 academic year, had recommended a change to the portion of USAT which asks a student's "sex." That question, as it was phrased, was in direct contravention of the Senate "Policy on Collecting Information about Sex and Gender at Queen's University" (approved September 2016). Appendix C contains the proposed wording for this additional question.

Recommendation 10: That a set of administrative documents accompany the survey.

In its discussions TAIC members repeatedly noted the fact that students, instructors, and Heads, Deans, RTP committees, and others are not always entirely informed about the purpose and use of the student survey. The TAIC recommends adoption of:

1. Guide for Students Administering the QSSET (Appendix D)
2. Interpretive Guide for Heads, Deans, RTP Committee, and others using QSSET in assessing teaching effectiveness (Appendix E)
3. Guide for Instructors on the purposes, administration, and use of QSSET (Appendix F)

Recommendation 11: That the data from QSSET be reported only as distribution of responses.

The use of means in aggregating the data from student surveys invites fundamentally flawed comparisons between courses of different kinds, at different levels of instruction, and delivered under different conditions. Moreover, such comparisons would negate the contextualization that the design of QSSET intends. Finally, it is well established that results from course evaluations often do not follow a Normal distribution. For this reason the use of means in assessing a Member's performance is problematic because it masks the heterogeneity of the true distributions of evaluations. Therefore, the aggregated data from QSSET should be reported only in terms of the distribution of responses to each question, shown as a number and percentage of total responses.

Recommendation 12: That the privacy and security of the data be monitored.

The results of any instrument used for assessing an instructor's job performance and impacting promotion and tenure must be handled with the highest attention to privacy

and security. It is important that the processes used to obtain the relevant data, the software, and the data obtained are maintained within strict security protocols.

Recommendation 13: That a standing committee be created to monitor, maintain, and revise the QSSET and its processes as needed.

The TAIC has observed that the USAT has run over several decades without monitoring of the effectiveness of the questionnaire or the mechanisms whereby it is delivered. The TAIC therefore recommends that a standing committee be created to ensure that the QSSET remains effective and relevant and that the processes used to deliver it are efficient and responsive to emerging needs.

Recommendation 14: That a question bank be created with optional questions that may be selected by individual instructors, Faculties and Schools.

A bank of additional questions is necessary to tailor the QSSET to reflect the different pedagogies of Faculties and Schools within Queen's. While the current USAT includes a bank of additional questions, there is no mechanism for vetting new questions or monitoring their effectiveness. Therefore, TAIC recommends that its successor committee invite the development of additional questions and develop a protocol for their vetting and their use. While TAIC supports the development of an additional question bank, TAIC strongly recommends that there be a strict limit on the number of additional questions added to any survey. TAIC advises that adding more than two questions may reduce the response rates, thereby limiting the effectiveness of the survey and the utility of the results.

Appendix A

Proposed QSSET Survey Questions

Please indicate your level of agreement with the following statements. Responses can range from 1 (Strongly Disagree) - 7 (Strongly Agree) and you have the option to respond with not applicable as well.

Student

1. This course fit my interests.
2. I was usually prepared for class.
3. My academic background prepared me for work at the level of this course

Instructor

1. The instructor clearly communicated the expectations for learning in this course.
2. The instructor encouraged students' engagement in the course.
3. The instructor presented the course material effectively.
4. The instructor made themselves available to students.

Course

1. The course materials and resources contributed to my learning.
2. The feedback I received in this course provided guidance on how to improve my learning and performance.
3. The workload in this course was comparable to other courses I have taken at this level.

Course Infrastructure

1. The physical and/or virtual environment promoted my learning.
2. The technology was adequate to the purposes of this course.
3. The course meeting time and/or online module timing was beneficial to my learning.

Appendix B

QSSET Pilot –Data Analysis

Summary

Student evaluations of teaching are an established process not only at Queen's, but at most universities. They form an important part of the process to evaluate faculty members' performance and to improve teaching. Thus, it is critical that student evaluation tools are valid, reliable and context appropriate. Psychometric analyses of the QSSET questionnaire indicate that it provides relevant, reliable information about students' experiences in their courses. It is clear from the student data that students preferred the QSSET to the USAT. However, instructor reaction was mixed. All instructors applauded the move to an online tool and appreciated the questionnaire preamble and its attempt to guide students towards appropriate, constructive responses. However, there were concerns about the quality of the generated report. These concerns are easily addressed through software and formatting changes. It should be noted the USAT has been in place for many years, allowing faculty members and Unit Heads years to develop processes to maximize the utility of the provided information. If a switch to the QSSET is made, there will likely be an implementation dip as personnel learn to use the QSSET data well. However, the online nature of the QSSET means that changes are easy to implement, allowing the QSSET to be a much more responsive tool than the USAT. Thus, while an implementation dip is expected, it is also expected that over time both faculties and the QSSET will evolve to maximize the utility of the instrument and resulting reports.

Introduction

The QSSET was designed to allow students to provide meaningful feedback about their experiences in a course. This report presents important findings from the pilot project. These findings include the psychometric properties of the questionnaire, student and faculty reactions to the questionnaire, and evidence about the feasibility of implementing the QSSET in an effective manner.

The data informing this report comes from 1968 completed, usable QSSET surveys, follow-up surveys to students about their QSSET experience, follow-up surveys to faculty members about their QSSET experience, interviews with faculty members, and the observations of the personnel who implemented the pilot project. This is a rich data set that includes a wealth of both quantitative and qualitative data. The report is divided into the following sections: implementing the QSSET, psychometric properties, student reactions, faculty reactions, and reporting.

Implementing the QSSET

Faculty were initially invited to participate in the QSSET by completing an online questionnaire. Sixty-nine faculty members declared an interest in participating in the pilot project. The final project included 49 course sections. The two most common reasons for faculty not participating in the project were not meeting eligibility requirements (faculty needed to be tenured or continuing adjuncts) or the course was being evaluated in the winter term instead of the fall term. There were three faculty members who opted to stay with the USAT instead of trialing the QSSET, one course section was missed because of a spreadsheet error, and another was missed because the person implementing the pilot arrived late to the class.

The classes participating in the QSSET pilot project included undergraduate, graduate, online, blended, and face-to face courses (see Table 1). A broad range of faculties and departments were included, with 21 different departments or faculties represented in the sample.

Table 1. Course types

Course Type	N
Undergraduate	39
Graduate	10
Online	3
Blended	8
Face-to-face	38

To maximize response rates and mimic the USAT implementation, a member of the QSSET pilot project team went to each class. As with the USAT, the instructor left the room, and the QSSET team member briefly explained the QSSET pilot project to the students and gave the questionnaire URL. Questionnaire URLs were individualized for each course section, and each survey had a header identifying the course and section. For online courses, the survey URL was given to the instructor, who then passed it on to the students via email. There were also three cases where a QSSET team member could not be present in the classroom and so the URL was given to the instructor to give to students in class. Questionnaires were open for completion for 24 hours for face to face classes and 7 days for online classes. There was no evidence of security issues (e.g., students completing multiple questionnaires, instructors accessing and completing the questionnaire) but thought will need to be given as to how to improve security to ensure confidence in the process.

Students were required to bring their own device to access the internet and complete the questionnaire. Not all instructors were aware of this, so not all instructors informed their students they needed to bring a device. Despite this, among all participating classes, only one student did not have a device. The questionnaire and survey software were trialed on a variety of platforms and the result was that only a single student was not able to access the questionnaire due to a hardware or software issue. This student was using a 12-year old Blackberry phone and while the phone was able to access the questionnaire, some of the characters appeared on screen as black

10

squares. All other participants were able to access the questionnaire regardless of which phone, tablet, or laptop they used.

The major difficulty with the implementation of the QSSET pilot project was the fact that only two people were available to come to classes to implement the questionnaire. This created numerous scheduling conflicts that were difficult (and in a few cases - impossible) to resolve. Other potential impediments (e.g., students not having devices, poor internet access, lack of interest from faculty members in participating in the project) did not prove to be problematic.

Psychometric properties

As with any questionnaire, it is important to understand the psychometric properties of the instrument. This information is critical to knowing if the questionnaire is yielding reliable, valid information, and to finding items which may be problematic.

Approximately 95% of students completed the entire questionnaire if they started it. The items that were most often left blank or selected as "not applicable" by the students were "The instructor is available to students outside the class" (14.2%). We believe that for some courses (i.e., blended learning and online courses) there may be some confusion as to what "outside the class" means. Further, it is also possible that students who did not seek the instructor's time outside of class left this question blank or responded with "not applicable", regardless of the instructor's availability. The other item that was commonly left blank or "not applicable was "The instructor was respectful of social diversity" (11.1%). It may be that some students did not feel the course context was one in which the instructor's respect for social diversity would be evident.

One potential concern is that students are not responding to items sincerely. While it is impossible to know to what extent this is the case, one possible indicator is that 5.6 percent of respondents (N = 110) gave the same rating across all items. As a rough estimate, this would indicate that approximately 95% of respondents considered each item separately, and 5% gave the same general, holistic rating of the instructor to all items. T-tests revealed that for all items, there was a statistically significant difference in mean scores between when the item response was by students who gave the same rating across all items and when the item response was by students who completed each item individually. In all cases the mean score was higher when completed by students who gave the same response across all items. Thus, it appears that students are most likely to give the same response across all items when they have very positive feelings about an instructor.

Another indicator of the sincerity of response is the amount of time taken to complete the questionnaire. The median survey completion time was 194 seconds, with 90% of students completing it in under 7.5 minutes. With 14 items, a median time of 194 seconds is approximately 13 seconds per item. Given that item stems were typically about 6 words long, this is a reasonable length of time to respond to the item. About 10% of respondents completed the questionnaire in less than 90 seconds, meaning less than 5 seconds per item. This is a

11

reasonable estimate of the lower bound of the time in which the questionnaire items could be read, interpreted, and responded to. Thus, it appears that approximately 90% of students are completing the questionnaire sincerely and thoughtfully.

In addition to the 14 quantitative items, there were also five items where respondents had the opportunity to make qualitative comments. Of the 1968 respondents, 223 commented after section 1, 237 after section 2, 198 after section 3, and 69 after section 4. Final comments were made by 1199 respondents. Some courses had almost every student making comments, whereas others had very few. No obvious patterns were observed in which courses garnered more comments.

No formal analysis has been conducted of the comments at this point, but an initial inspection reveals that most comments are positive. This aligns with the responses on the quantitative items, which are also positive. Negative comments tend to be specific to an element of the course or the instructor (e.g. the marking is unfair, the instructor is boring, the course is not well structured). I have yet to read a comment that is profane or disrespectful. This does not mean such comments are not within the data set, as there are more data to analyze.

Mean scores of the items

Responses were coded on a 1 to 7 scale with 1 = Strongly disagree and 7 = Strongly agree. The midpoint of 4 was phrased "Neither agree nor disagree" on the survey. Responses of "Not Applicable" were not included in the analyses. All items had the entire scale from 1 to 7 used by students and every item had "Strongly agree" and "Agree" as the two most selected responses. Descriptive statistics for each item are presented in Table 2. Mean scores are high, indicating high levels of agreement with the items.

Table 2. Descriptive statistics for quantitative items.

Item	Mean	SD
This course fits my interests	5.42	1.62
I was usually prepared for class	5.40	1.43
My academic background prepared me for work at the level of this course	5.56	1.48
The instructor clearly communicated the expectations	5.62	1.50
The instructor encouraged students' engagement in the course	5.92	1.39
The instructor presented the course material effectively	5.44	1.68
The instructor was respectful of social diversity	6.32	1.02
The instructor was available to students outside	5.92	1.27
The course materials and resources contributed to my learning	5.69	1.36
The feedback I received in this course was helpful	5.15	1.67
The workload in this course was comparable to other courses I have	5.60	1.42
The physical and/or virtual environment promoted my learning	5.57	1.39
The technology was adequate for the purposes of the course	5.82	1.19
The course meeting time and/or online module timing was beneficial	5.58	1.42

Subscales and reliability

The survey was divided into four subscales that were intended to measure the student's contribution, the instructor, the course, and the learning environment. To gauge how well each section represented a cohesive subscale, Cronbach's alpha was calculated for each section. Generally, values below 0.70 are considered unacceptable, whereas values above 0.90 are considered excellent. Values between 0.70 and 0.80 are considered low and values between 0.80 and 0.90 are considered medium. For the three items that comprised the student contribution subscale Cronbach's alpha was calculated to be $\alpha = 0.72$, with all three items contributing positively to the reliability. The instructor subscale (5 items) had a value of $\alpha = 0.88$. Four of the five items contributed positively to the reliability, but "The instructor was respectful of social diversity" item had a slightly negative effect on the reliability of the subscale. The course subscale (3 items) had a value of $\alpha = 0.74$, with all three items contributing positively to the reliability of the subscale. Finally, the learning environment subscale (3 items) also had a value of $\alpha = 0.78$, with all three items contributing positively to the reliability. Overall, these values indicate acceptable levels of reliability for each subscale. Subscale reliability tends to be low when the number of items is low, so improving the reliability coefficients would likely require adding items. In the case of the instructor subscale, reliability would be improved by removing the "The instructor was respectful of social diversity" item.

Factor Analysis

An exploratory factor analysis was conducted to determine the dimensionality of the survey. Results from the KMO measure of sampling adequacy (KMO = 0.94) indicated that factor analysis would likely yield interpretable results. The factor analysis results (using principal axis factoring) revealed that a two-factor solution would best model the data parsimoniously. This was determined using both the Kaiser criterion and a scree plot. Forcing a four-factor solution onto the data yielded worse fit figures and poor conceptual clarity. The factor loadings suggest that the first (and strongest) factor model is the positivity of the student experience in the course. All items in the questionnaire loaded most strongly onto this factor. The second factor is the student contribution to the course experience and only the first three items (the student sub-scale) load onto this factor (See Table 3).

Table 3. Factor loadings of survey items.

Factor Loadings		
	Factor	
	1 (Positive experience)	2 (Student Contribution)
This course fit my interests	.657	.401
I was usually prepared for class	.486	.428
My background prepared me for work at the level of this course	.544	.289
The instructor clearly communicated the expectations	.798	-.149

The instructor encouraged students' engagement in the course	.768	-.237
The instructor presented the course material effectively	.829	-.141
The instructor was respectful of social diversity	.598	-.231
The instructor was available to students outside	.684	-.221
The course materials and resources contributed to my learning	.750	.191
The feedback I received in this course	.676	.009
The workload in this course was comparable to other courses I have	.614	.037
The physical and/or virtual environment promoted my learning	.748	-.018
The technology was adequate to the purposes of this course	.642	-.094
The course meeting time and/or online module timing was beneficial	.650	-.029

The factor analyses do not show any obviously problematic items. The results suggest that students are able to separate their own contribution to the course from other factors such as the instructor and learning environment, but that all factors external to the student are rated holistically. Thus, if a student gives a high rating to one aspect of the course, they are likely to give high ratings to all aspects of the course. The item with the highest loading "The instructor presented the course material effectively" is a general teaching effectiveness item, and students' experience with respect to teaching effectiveness is likely what the first factor (and hence the questionnaire) is measuring.

Differences across groups
As the QSSET pilot project included different types of courses, it is possible to examine how the responses differed across demographic groups. As an example, with the exception of the item "The workload in this course was similar to other courses I have taken", graduate courses had statistically significant higher means on all items than undergraduate courses. Likewise, with the exception of the item "The physical and/or virtual environment promoted my learning" there were statistically significant differences between face-to-face, blended, and online courses. For all three student contribution subscale items the face-to-face courses had lower means than either blended or online courses. For all of the instructor subscale items, the blended courses had the lowest means. For the other two subscales which course delivery models had the highest or lowest means was mixed. Given the small sample size of blended and online courses, no conclusions about the impact of course delivery model on the results can be made.

In summary, the QSSET questionnaire has acceptable reliability (especially for such a compact instrument), and it appears to be measuring students' experience of teaching effectiveness. There are no obviously problematic items, although serious consideration should be given to removing

14

the "The instructor was respectful of social diversity" item. Doing so would improve the psychometric properties of the instrument, while also making it shorter. It also appears the vast majority (~ 90%) of students completed the questionnaire thoughtfully and sincerely.

Student reactions

Observations and informal data collection in the classrooms where the QSSET pilot was implemented indicated that students found the questionnaire easy to complete and preferred the QSSET to the USAT. These initial data collection efforts were followed up with a survey sent to students to ask them about their experience writing the QSSET. The results from this survey indicate that approximately 95% of the students understood what the QSSET items were asking and that of the students who expressed a preference, 90% preferred the QSSET to the USAT.

Other important results to come from the student follow-up survey were that most students (59%) did not read the preamble. 71% of respondents agreed the preamble section clarified the purpose of student evaluations of teaching at Queen's, but less than half (47%) agreed the preamble made them think about the types of comments to avoid. For all subscales, over 90% of students agreed the items were relevant to course evaluation. Technical difficulties were reported by 2% of the students, and of those with an opinion, 92% agreed the QSSET allowed them to convey their experience of teaching.

Qualitative comments from the students were also collected. For the preamble section the two main themes were that the preamble was too long and that guidance or clarity was needed as to what constituted "inappropriate comments." For the student contribution section there were few comments, although two students noted that for mandatory courses, items such as "This course fits my interests" were not relevant. There were two dominant themes for the instructor subscale were comments about the "The instructor was respectful of social diversity" item where some students felt the item was not relevant to the course, or unrelated to good teaching. The second theme was that the item "The instructor presented the course material effectively" was too broad and needed to be broken down into smaller components of effective instruction. There were also a few students who wished for an item that related to fairness of grading.

Comments about the course subscale were varied and dominant themes did not emerge. However, there were two students who commented that their course is graded 100% on final exam scores and so feedback during the course was not present. Another two students commented that course materials and resources are under instructor control and therefore that item should be part of the instructor subscale. Conversely, there was a dominant theme within the comments about the course infrastructure subscale. That theme was that the items were vague and difficult to understand. It should be noted that few students (3%) commented on this subscale, but of those that did, they found the item wordings vague.

Overall, the student reaction to the QSSET was extremely positive. They found the questionnaire easy to understand and the items useful. Technology did not pose a barrier to the students, and the vast majority of the students preferred the QSSET to the USAT.

Instructor reactions

Like the students, the instructors were given a follow-up survey inquiring about their experiences with the QSSET. This survey was completed before they received their QSSET results to ensure that responses were not coloured by the results. A total of 21 usable surveys were completed. The quantitative items indicated that instructors felt very good about the design of the survey and the items. Between 80 and 95 percent of instructors agreed that items were appropriate and connected to the purpose of the subscale. There were two exceptions to this trend. The first was a question that asked if the items in the instructor section of the QSSET were appropriate. Forty-three percent of instructors agreed the items were appropriate but 52% were neutral on the question. The remaining five percent disagreed. The other exception was an item inquiring about the design of the QSSET. Fifty-eight percent of instructors agree the design was appropriate, while 21% disagree and another 21% were neutral. Overall, the quantitative items indicated that instructors were satisfied the QSSET items were appropriate and connected the purpose of the instrument.

Qualitative responses were obtained through emails, items on the feedback questionnaire, and interviews. Many instructors were passionate and forceful in their written or spoken comments, and it was obvious that student evaluations of teaching was a topic that was of great professional and personal importance. There was some difficulty in thematically coding the data as many responses were unique to the context of the instructor. This included comments about individual buildings or classrooms, unusual course structures (e.g., field courses), and very large class sizes. While the comments themselves were diverse and individualized they reflected a common wish for the context in which the course had taken place to be considered when the QSSET data were analyzed and interpreted. Instructors did not want to feel disadvantaged for teaching an 8:30 course, or a mandatory course, or a large lecture hall course etc. Interviews with two Department Heads indicated they were sensitive to the course contexts when interpreting student evaluation of teaching data, but this is an area where instructors may need reassurance.

While the instructor reaction to the QSSET pilot project was varied, a few common themes did emerge. The first was that instructors were very happy to have student evaluations of teaching move to an online delivery model. They expressed several reasons for this that included: a general feeling that students preferred online surveys, less use of paper, a belief that results could be returned more quickly to them, and a belief that because students are more accustomed to typing than writing by hand, they would provide more comments. There is some evidence to support this last reason as approximately 2/3 of QSSET respondents included at least one written comment on their QSSET. We do not have frequency of comments on the USAT to serve as a comparison, but comments from instructor interviews indicated they found the comments on the QSSET to be greater in number and in length than comments received on the USAT.

A second common theme was that instructors did not use the scores on the quantitative items to inform their improvement efforts. Instead, they relied heavily on the written comments. The Department Heads interviewed talked about the importance of the scores for performance review

16

and other administrative duties but also admitted to relying on comments to inform the development of their own teaching. If one of the purposes of the QSSET is to improve teaching, it is critical that students are encouraged to include thoughtful written comments.

A third theme was an appreciation for the purpose of the preamble. Many instructors shared that poor USAT scores or disparaging comments had caused them emotional stress and hoped to avoid this in the future. Instructors were willing to read critical comments as long as the intent was to improve the course and not demean the instructor. This theme relates to the aforementioned instructor anxiety about fair use of student evaluations of teaching.

Reporting

Consideration will have to be given as to how QSSET results are communicated to instructors. For the pilot project, the default reporting tool that is provided with Qualtrics was used. This tool has some ability for customization, but this ability is limited and will not be sufficient for our needs. The default report used gave the following information for each item:

1. A bar chart showing the number of students who selected each response option

2. A short statistical summary of the item including mean score, standard deviation, maximum, and minimum value

3. A table that repeated the information in the bar chart and added percentage of respondents who selected each response option

In addition, all written comments were included in the report. At present, we have no way of censoring inappropriate comments. Doing so manually, would likely be cost prohibitive. Software solutions exist but are less than perfect. It may be that the best course of action is for instructors to see all comments and allow them to determine which (if any) comments are inappropriate. This gives instructors agency over the determination of the appropriate/inappropriate boundary but also potentially exposes them to emotionally difficult comments.

There were several identified problems with the reporting, all of which have solutions. The ease of implementing those solutions will depend on the final software package used to implement the QSSET. These issues and solutions are provided in Table 4. Reporting is critical to the success of the QSSET for several reasons. Firstly, the quality of the report is the main determinant of instructors' opinions of the QSSET. Secondly, the report is what instructors use to inform their improvement plans and what Unit Heads use to evaluate teaching quality. Finally, the suitability of the reporting software will be the primary determinant in the workload associated with generating the QSSET reports.

Table 4. Reporting issues and solutions

Issue	Solution
Quantitative item responses were coded so that "strongly agree" was assigned a value of 1, and "strongly disagree" was assigned a value of 7. This was confusing to instructors because they expected higher mean scores to reflect high levels of agreement with the item	Recode the scoring so that "strongly agree" = 7 and "strongly disagree" = 1.
"Not applicable" was coded as 8. This distorted the mean scores of the items.	Do not include "Not applicable" in the statistical calculations. Frequency can still be reported.
The length of the report was excessive. Each item took a minimum of one page.	Reformat the report so that charts and tables are smaller. Locate the charts and tables in a manner that maximizes the number of items that can fit on one page. Ideally, it would be good to have one subscale per page.
Many instructors did not understand the statistical summary	Remove extraneous information from the statistical summary. We could report only the mean item score and number of responses without impacting the report's utility.

Appendix C

Joint Committee on Teaching Assessment
Report and Recommendations to JCAA re: Gender Self-Identification
April 2017

Background

[Reason for the report, relevant research, and steps taken in developing the report]

While reviewing the USATs, the Teaching Assessment Committee was made aware of a relatively new Senate policy which has an indirect impact on the question on the current form which asks students to identify their sex as "Male" or "Female." In September 2016, Senate approved and adopted the "Policy on Collecting Information about Sex and Gender at Queen's University." The policy provides definitions for and a clear distinction between the use of the terms "sex" and "gender" on questionnaires within the university. The policy is and is intended to inform in-take forms, record documents, and surveys in use at Queen's so that these will be inclusive and free from bias related to sex and gender.

Section five of the policy states,

> "it is the responsibility of the individuals creating and/or administering in-take forms, records documents, and surveys to make every effort to ensure that the language used in these documents is free from bias or barriers related to sex and gender identification."[1]

The Teaching Assessment Committee has reviewed the policy, discussed it with the Equity Office, and is now making a recommendation to the JCAA about an important change that must be made to the current USAT.

It should be noted that this report and recommendation was developed in consultation with Dr. Erin Clow, Equity Advisor in the Equity and Human Rights Office.

Analysis and Discussion

The "Policy on Collecting Information about Sex and Gender at Queen's University" stipulates that a rationale be offered in cases where information about sex or gender is requested on any official university document. In this case, the Teaching Assessment Committee recommends that a question about gender be included on the USAT questionnaire in light of information presented in several studies which show that a gender bias often emerges in teaching evaluations.[2] The

---

[1] Queen's University. *Policy on Collecting Information about Sex and Gender at Queen's University*. (Kingston: Queen's University, 2016).

[2] See, for instance, L.R. Kogan, R. Schoenfeld-Tacher, & P.W. Hellyer (2010). Student evaluations of teaching: Perceptions of faculty based on gender, position, and rank. *Teaching in Higher Education*, 15(6),

inclusion of a question on gender will allow the faculty member, Head, and Dean to take any potential bias into account when reviewing the USAT scores. Further it will allow the University to review the overall results of the USAT to determine the potential presence of the bias in the assessment process generally and to consider ways to remedy this bias.

Recommendations

The Joint Committee on Teaching Assessment recommends that

1. the USAT be revised as follows to be consistent with the Senate "Policy on Collecting Information about Sex and Gender at Queen's University." The below chart presents the current wording and the proposed revision:

Current USAT wording:	Proposed wording:
- Are you - Female - Male?	How do you self-identify in terms of gender? - Man - Woman - I do not identify within the gender binary - I prefer not to disclose information concerning my gender

2. This change be implemented starting as soon as possible.

3. An analysis of the impact and validity of the question be undertaken starting at the end of the first academic year of implementation. Data is needed to determine the validity of the question and to find out if gender perceptions have an impact on evaluations at Queen's.

Committee Members

R. Easteal, J. Fraser, D. Hopkins-Rosseel, D. Klinger, T. Lively, J. Pierce, E. Soleas, D. Stockley

Co-Chairs: D. McKeown, K. Renders

Secretary: J. Cleary

623-636, and A. Boring (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics* (http://www.sciencedirect.com/science/article/pii/S0047272716301591)

Appendix D

Guide for Students Administering the Queen's Survey of Student Experience of Teaching (QSSET)

Thank you for helping to administer the Queen's Survey of Student Experience of Teaching in your course. The following instructions will guide you through the steps in administering the survey.

Please make the following announcements to the class.

1. **On Time Allowed to do the Survey**
   a. Fifteen minutes are allotted for completing the survey.

2. **On the Purposes of the Survey**

   The online survey solicits student feedback about instructors for two purposes:
   a. to help Instructors improve their teaching and
   b. to help assess and evaluate teaching by Queen's faculty for performance review, contract renewal, tenure and promotion.

3. **On Who Sees the Survey**
   a. The instructor will see both the numerical scores from the survey and written comments. Heads and Deans will see only the numerical scores.

4. **On Comments made by Students in the Survey**
   a. Note the introduction to the survey which indicates the following:
   b. (Please read aloud) The QSSET is your opportunity to provide feedback about pedagogy, course content, course materials, and the student learning experience generally. Written comments are especially valuable. However, the QSSET is **not** an opportunity to make comments that are disrespectful, demeaning, discriminatory or harassing in nature. **Surveys containing such comments will be discarded and will form no part of the evaluation process.**

5. **On Where to Access the Survey**

   Students registered in the course will receive an email with a link to the survey.

[Final Note: additional details, perhaps including a FAQ section, may later be included to tell students how to manage any technical details arising during the administration of the survey.]

Appendix E

Using QSSET in the Evaluation of Teaching: For Heads, Deans, and RTP Committees

The purpose of this document is to assist Heads, Deans, RTP committees and any others who are the evaluators of teaching per Article 29 of the Collective Agreement, in using QSSET responses as part of the evidence of teaching effectiveness they must consider in merit, renewal, tenure, promotion and continuing appointment decisions.

QSSET and the Evaluation of Teaching

Article 29.3.1 provides that a survey approved by QUFA and the University, now QSSET, will be used in the assessment and evaluation of teaching. However, it is important for evaluators to recognize that this survey is not in itself an assessment and/or evaluation of teaching but one source of evidence which Heads, Deans, members of RTP committees and others will consider in the course of assessing and evaluating teaching.[i] The assessment of teaching as it is described in Article 29 requires the consideration of matters that extend well beyond the scope of QSSET. Article 29.1.2 of the QUFA-Queen's Collective Agreement provides: "For assessment and evaluation purposes, teaching includes all presentation whether through lectures, seminar and tutorials, individual and group discussion or supervision of individual students work in degree-credit programs." 29.1.3 adds that "Assessment and evaluation of teaching shall be based on the effectiveness of the instructors, as indicated by command over subject matter, familiarity with recent developments in the field, preparedness, presentation, accessibility to students and influence on the intellectual and scholarly developments of students." Matters such as supervision, command over subject matter, and familiarity with recent developments in the field cannot be compassed by a survey of students as they do not have the knowledge or expertise to provide this information. QSSET has been designed in recognition of the value of information about students' experience of teaching but also the limitations in students' ability to assess teaching which means that the surveys cannot serve as a proxy for evaluation.

QSSET Design

QSSET has also been designed to acknowledge that students' experience of teaching is affected by factors beyond the Instructor's control. These include the student's own preparation for and engagement with the course, marking which may not have been performed by the Instructor, or course materials which may not have been prepared by them, the adequacy of the classroom, and/or technological support. The purpose of the questions under "Student," "Course," and "Infrastructure" are to provide context to help those assessing and evaluating teaching to determine how reliably the scores on the "Instructor" questions reflect the Instructor's actual

teaching. **Only scores on questions under the heading "Instructor" are to be used directly in the evaluation and assessment of teaching. The exception is where the evaluator knows that the Instructor also performed all evaluations of student work and/or was responsible for the design and presentation of the course materials. In such cases, appropriate questions under "Course" should be considered as well.** Evaluators may seek clarification about responsibility for course elements from Instructors should it be necessary.

Using QSSET

Consider the questions under "Student." They ask students to reflect on their own relation to the course. While the reflections these questions prompt may temper students' responses on the Instructor section, the students' responses also provide information to the evaluator about what the Instructor was up against, or alternatively what advantages the Instructor may have enjoyed. For instance, if students do not indicate strongly that the "course fits their interest" this may serve as a reminder to the evaluator that the course is a tough, required course—or alternatively that it is an elective but because of resource constraints there are few options for students. In such circumstances students might be expected to rate an Instructor as less effective in the presentation of a material than in a course where the students are already enthusiastic about the subject matter. Alternatively, if in these circumstances students rate an Instructor as highly effective in presenting the course material, their experience may indicate the considerable pedagogical intelligence the Instructor brought to the course. If many students indicate that they were not prepared for class, their rating of the Instructor's effectiveness in presenting the material may be less reliable because the Instructor, reasonably was assuming preparation. Students' dissatisfaction with marking, or frustrations with the IT support, or dislike of the course materials may cause them to experience an Instructor as less effective when these factors lie beyond the control of the Instructor. The QSSET is designed both to foreground the fact that the students are testifying to their experience, not performing an evaluation, and to provide evaluators with information about the determinants of that experience. Evaluators, who will be familiar with the curriculum, the resources, and the student culture of their Units, will need to review the responses from QSSET for the relationships the survey questions indicate between the students' experience of teaching and the conditions in which it is conducted.

The attention QSSET demands to the correlation between the students' rating of teaching effectiveness and the circumstances in which the teaching was conducted also demands that the data it yields be presented in terms of a distribution of responses rather than through means and standard deviations as was used with USAT. Such presentation was always dubious, inviting comparisons between teaching conducted under vastly differing conditions. In contrast, the presentation of QSSET results allows evaluators to discern patterns and relationships. For instance, persistent bi-modal responses for a particular course may indicate that the Instructor is teaching controversial material—off-putting to some students but exciting to others. Or it may indicate that the Instructor's teaching is highly effective and stimulating for well-prepared students, but loses less well-prepared ones, a hypothesis supported by the responses about preparation for work at course level in the "Student" section. Where a mean would simply

indicate mediocre teaching, the QSSET might show the evaluator either a strong Instructor struggling with a problem in the Unit's curricular design, or alternatively an Instructor who needs to work on explaining basic premises or concepts. The evaluator's decision about which of these interpretations is best will depend on knowledge about teaching in the Unit. It may also be influenced by the Instructor's own interpretation of the results which the Instructor has supplied in a teaching dossier.

Finally, it is important to note that scholarship regarding student evaluations of teaching indicates that responses can be biased with respect to factors not relevant to teaching quality. With respect to gender bias in student evaluations of teaching, research findings are complex and often contradictory, but the general conclusion is that when biases exist, it is female instructors who are disadvantaged.[ii] Further, it appears that students have different expectations of male and female instructors based upon gender stereotypes.[iii] For example, female instructors are generally rated higher on questions pertaining to interpersonal skills, however, when they are perceived to be weak in this area, they are rated more harshly than male counterparts. Gender biases have been shown to exist both in the quantitative survey items and the comments. It should be noted that while gender bias is the most studied, other forms of bias based on race, attractiveness, age, and accent have been shown to exist. The problem of bias is intractable because the bias lies in the students, rather than in the survey tool itself. Evaluators of teaching need to be mindful of potential bias when considering QSSET results.

Appendix F

Using QSSET – For Instructors

The purpose of this document is to assist instructors in interpreting QSSET results and presenting those interpretations to Heads, Deans, RTP committees or any others who will be making decisions about the instructor using QSSET results.

QSSET and the Evaluation of Teaching

Article 29.3.1 provides that a survey approved by QUFA and the University, now QSSET, will be used in the assessment and evaluation of teaching. However, it is important for instructors to recognize that this survey is not in itself an assessment and/or evaluation of teaching but one source of evidence which Heads, Deans, members of RTP committees and others will consider in the course of assessing and evaluating teaching.[iv] The assessment of teaching as it is described in Article 29 requires the consideration of matters that extend well beyond the scope of QSSET, or any survey of students. Article 29.1.2 of the QUFA-Queen's Collective Agreement provides: "For assessment and evaluation purposes, teaching includes all presentation whether through lectures, seminar and tutorials, individual and group discussion or supervision of individual students work in degree-credit programs." 29.1.3 adds that "Assessment and evaluation of teaching shall be based on the effectiveness of the instructors, as indicated by command over subject matter, familiarity with recent developments in the field, preparedness, presentation, accessibility to students and influence on the intellectual and scholarly developments of students." However, as a one-on-one form of instruction supervision cannot be assessed through surveying. Moreover, students do not have the expertise to comment on matters such as command over subject matter and familiarity with recent developments in the field. QSSET has been designed in recognition of the value of information about students' experience but also the limitations in students' ability to perform and full and valid assessment of teaching, which mean that the surveys cannot serve as a proxy for evaluation. The QSSET design furnishes opportunities for instructors to interpret student responses in relation to multiple determinants of students' experience of teaching. The limited scope of student surveys places an onus on instructors to furnish supplementary information or material so that Heads, Deans, RTP committees and other evaluators can make best use of QSSET and evaluate aspects of teaching that a survey of students cannot compass.

QSSET Design

QSSET acknowledges that students' experience of teaching is affected by factors beyond the instructor's control. These include the student's own preparation for and engagement with the course, marking which may not have been performed by the instructor, any course materials not prepared by the instructor, the adequacy of the classroom, and/or technological support. The

questions under "Student," "Course," and "Infrastructure," provide context to help those assessing and evaluating teaching to determine how well the scores on the "instructor" questions reflect the instructor's actual teaching. **Only scores on questions under the heading "instructor" are to be used directly in the evaluation and assessment of teaching. The exceptions are where the assessor knows that the instructor also performed all evaluations of student work and/or was responsible for the design of the course and presentation of the course materials. In such cases, appropriate questions under "Course" should be considered as well.** However, instructors must be aware that Heads, Deans, RTP committees and any other evaluators may not be aware of the courses which the instructor has designed and/or prepared the materials for. They may also not be aware of the extent to which the instructor performed assessment in the course. Article 28.2.4 of the Collective Agreement provides that for the purpose of Annual and Biennial reports "it is the Member's responsibility to provide…sufficient detail of activities and their outcomes to enable the Unit Head to assess the Member's performance." Moreover, for most personnel decisions the onus is on the Member to demonstrate that standards have been met. For these reasons, it is in the instructor's interest both to ensure that assessors have adequate information to evaluate QSSET results appropriately, and to supplement QSSET with additional material as necessary.

Using QSSET

QSSET is designed to present correlations between student perceptions of instructor effectiveness and other factors that influence their experience of the course. Consider the questions under "Student," which ask students to reflect on their own relation to the course. While the reflections these questions prompt may temper students' responses on the "instructor" section, the students' responses also provide information to the assessor about what the instructor was up against, or alternatively what advantages the instructor may have enjoyed. For instance, if students do not indicate strongly that the "course fits their interest," and the course is a tough, required course—or alternatively that it is an elective but because of resource constraints there are few options for students—the instructor may wish to remind the assessor of that fact in explaining less than enthusiastic responses to the "instructor" questions. Alternatively, if in these circumstances students rate an instructor as highly effective, the instructor may wish to underscore this fact in the context of their particular decisions on how to present the course material. Students' dissatisfaction with marking, or frustrations with the IT support, or dislike of the course materials, may cause them to experience an instructor as less effective when these factors lie beyond the control of the instructor, correlations that the instructor can also recognize and point out.

Because instructors receive the individual responses to the survey while Heads, Deans, RTP committees and other evaluators will see only aggregated data, instructors can play a role in framing their results for assessors. They can demonstrate correlations that may not be visible to assessors. For instance, if the same students who indicate that they were not prepared for class rate the instructor's effectiveness low while the better prepared students indicate greater satisfaction, this correlation suggests that lower ratings may be due to factors other than poor

26

instruction. Moreover, if instructors feel that the written comments illuminate survey results, they can help assessors of teaching by passing them on in a teaching dossier. QSSET is designed so that students can furnish written comments at the end of each section as well as at the end, but those comments are seen only by the instructor per Article 29.3.7 unless the instructor chooses to share them.

The attention QSSET demands to the correlation between the students' rating of teaching effectiveness and the circumstances in which the teaching was conducted also demands that the data it yields be presented in terms of a distribution of responses rather than through means and standard deviations as was done with USAT. Here too there are opportunities for instructors to provide valuable interpretation. Persistent bi-modal responses for a particular course may indicate that the instructor is teaching controversial material—off-putting to some students but exciting to others. Or it may indicate that the instructor's teaching is highly effective and stimulating for well-prepared students, but loses less well-prepared ones, a hypothesis that could be further supported by responses about preparation for work at course level in the "Student" section. instructors have a role to play in assisting the evaluator in interpreting such results by providing context about the course, their approach, and challenges they may face.

QUFA Members should also be mindful that if they wish to be assessed on the full range of teaching related matters that assessors and evaluators of teaching are supposed to consider per that Article 29.1.5 then they must provide assessors with appropriate information, and the way to do this is through a teaching dossier. Article 29.2 describes the purpose and possible contents of a teaching dossier which may be submitted for all personnel processes.

Bias and Inappropriate Comments

Finally, it is important to note that scholarship regarding student evaluations of teaching indicates that responses can be biased with respect to factors not relevant to teaching quality. With respect to gender bias, research findings are complex and often contradictory, but the general conclusion is that when biases exist, it is female professors who are disadvantaged.[v] Further, it appears that students have different expectations of male and female instructors based upon gender stereotypes.[vi] For example, female instructors are generally rated higher on questions pertaining to interpersonal skills, however, when they are perceived to be weak in this area, they are rated more harshly than male counterparts. Gender biases have been shown to exist both in the quantitative survey items and the comments. The problem of bias is intractable because the bias lies in the students rather than in the survey tool itself. It should be noted that while gender bias is the most studied, other forms of bias based on race, attractiveness, age, accent, and other factors have been shown to exist. Evaluators of teaching need to be mindful of potential bias when considering QSSET results; instructors who feel that such bias may have played a role in QSSET results, may wish to remind evaluators of this possibility.

[i] It is the Instructor's responsibility to provide materials that support a full assessment. In all RTP processes save Renewal of Tenure Track appointments the burden of demonstrating that the required standard has been met is on the Member. In the case of Annual/Biennial reviews, 28.2.4 requires the Member to provide "sufficient detail of activities and their outcomes to enable the Unit Head to assess the Member's performance" and where the Member fails to do that the Unit Head is to base their assessment and evaluation on the "information reasonably available" to them.

[ii] MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291-303. Young, S., Rush, L., & Shaw, D. (2009). Evaluating Gender Bias in Ratings of University Instructors' Teaching Effectiveness. *International Journal for the Scholarship of Teaching and Learning*, *3*(2), 1-14.

[iii] Mitchell, K. M., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, *51*(3), 648-652.

[iv] It is the instructor's responsibility to provide materials that support a full assessment. In all RTP processes save Renewal of Tenure Track appointments the burden of demonstrating that the required standard has been met is on the Member. In the case of Annual/Biennial reviews, 28.2.4 requires the Member to provide "sufficient detail of activities and their outcomes to enable the Unit Head to assess the Member's performance" and where the Member fails to do that the Unit Head is to base their assessment and evaluation on the "information reasonably available" to them.

[v] MacNell, Driscoll, & Hunt, 2015; Young, Rush, & Shaw, 2009.

[vi] Mitchell & Martin, 2018.