Re-envisioning the Student Experience of Instruction Survey Questions from the Student Perspective

July 30, 2021

Authored by:

Stephanie McKeown, PhD

Abdel-Azim Zumrawi, PhD

Camilo Peña, PhD (candidate)

Acknowledgements

The authors would like to acknowledge the following individuals who played a substantial role in the project assisting with data collection and analysis.

Academic Assistants (students)

Anshul Dhariwal
Lisvet Parra Montas
Dyuti Raghu
Thishani Rajapakshe

Offices of the Provosts and Vice Presidents Academic

Debbie Hart (Vancouver)
Laura Prada (Okanagan)

Office of Planning and Institutional Research

Jodie Foster
Brent Harris
Kristi Hoffman
Irene McKechnie
Gavin Yap

Executive Summary

> ***"They were good questions that were able to help me re-evaluate my learning experience in this course and reflect upon it."***
>
> - Student comment provided during the SEI pilot testing phase

This report summarizes the results of an eight-step project to evaluate the proposed wording of the six University Module Items (UMI) on the Student Experience of Instruction (SEI) survey.

We used a mixed-methods approach for this project. We first conducted 24 online focus group sessions with 116 students (16 focus groups) and 40 faculty members (8 focus groups), and held 29 online think-aloud interviews with individual students. All focus group sessions and interviews were digitally recorded and transcribed for further analysis. The transcriptions were then uploaded into NVivo, a qualitative analysis tool, and participant comments were analysed to determine patterns of meaning and organized into general themes. The themes were further refined and coded to aid in the interpretation of the data. The results of the qualitative analysis were used to further refine the questions with the aim to clearly articulate the intention behind each of the questions, and how they were related to the student learning experience and feedback on instruction.

The next phase of the project involved pilot-testing the revised survey questions developed from the thematic analysis. Students were invited by email to participate in the pilot survey through an anonymous survey link, using the survey software program Qualtrics. We received 333 responses to the pilot survey. To determine how well the new items functioned across individuals and respondent groups, we conducted a quantitative analysis of the questions using Item Response Theory (IRT) and Differential Item Functioning (DIF), conducted using the software programs SAS and Winsteps. Results from the IRT models showed significant improvement in each individual item's contribution to the overall survey information compared with a similar sample drawn at random from the 2020/21 (Winter Term 2) course evaluations. Based on the results of this mixed-method approach, we make the following recommendations on the SEI UMI questions for use at UBC.

Recommendations

We recommend that the following six new core UMI questions be adopted for implementation across both campuses for Winter Term 1 2021/22 courses and onwards:

Note: for the reader's reference, the previously proposed questions from the SEoT Working Group in May 2020 are included in grey italicized font below each of the newly recommended questions.

1. Throughout the term, the instructor explained course requirements so it was clear to me what I was expected to learn.
   The instructor made it clear what I was expected to learn.

2. The instructor conducted this course in such a way that I was motivated to learn.
   The instructor engaged me in the subject matter.

3. The instructor presented the course material in a way that I could understand.
   I think that the instructor communicated the subject matter effectively.

4. Considering the type of class (e.g., large lecture, seminar, studio), the instructor provided useful feedback that helped me understand how my learning progressed during this course.
   I have received feedback that supported my learning.

5. The instructor showed genuine interest in supporting my learning throughout this course.
   I think that the instructor showed concern for student learning.

6. Overall, I learned a great deal from this instructor.
   Overall, this instructor was effective in helping me learn.

Response options for all questions above: *strongly agree, agree, neutral, disagree, and strongly disagree.*

We also recommend that three common open-ended questions be included on all SEI surveys across both campuses to collect text comments:

7. Please identify what you consider to be the strengths of this course.

8. Please provide suggestions on how this course might be improved.

9. Do you have any suggestions for what the instructor could have done differently to further support your learning?

1.0 Introduction and Background

In February 2019, a Student Evaluation of Teaching (SEoT) Working Group formed with membership across both UBC Okanagan and UBC Vancouver campuses. Working under the auspices of the UBCO Senate Learning and Research Committee and the UBCV Senate Teaching and Learning Committee, the group had the following remit:

1. Interrogate anonymized UBC SEoT data, to determine if there is evidence of potential biases.
2. Review and assess the recent literature on the effectiveness of SEoT, with particular reference to potential sources of bias in evaluations.
3. Review the existing University questions used in SEoT in light of the data and available literature, recommending changes where appropriate.
4. Propose recommendations for appropriate metrics, effective analysis and presentation of data to support SEoT as a component of teaching evaluation.
5. Consider the implications any proposed changes may have on other components of teaching evaluation.

Through work and consultations conducted over an extended period, the SEoT Working Group presented a report to both the Okanagan and Vancouver Senates in May 2020. The report included 16 recommendations about student evaluations of teaching, which were endorsed by both Senates. Included in the report were recommendations to revise the former SEoT questions and to create a common set of core University Module Items (UMI) to be asked across both campuses. They also recommended changing the focus of these surveys to reflect the student experience, and to write the questions in a manner that puts the student at the heart of the question, thereby making the questions more student-centred. Thus, the Working Group recommended changing the name of the course-end questionnaire to Student Experience of Instruction (SEI).

The Working Group also proposed changes to the wording of the Vancouver version of the survey, including a substantial change to UMI 4, “Overall, evaluation of student learning (through exams, essays, presentations, etc.) was fair.” The changes proposed for the Okanagan version of the SEoT were more significant, reducing the questions asked from nineteen to six. Please see Appendix 1 for a list of the existing SEoT questions at each campus as well as the question wording proposed by the SEoT Working Group in their May 2020 report.

In the Fall of 2020, two new committees were formed to oversee the process of implementing the Working Group’s recommendations: a Steering Committee, and an Implementation Committee. Since one of the recommendations in the original Working Group’s report was to change the name of the process

from “student evaluations of teaching” to “Student Experience of Instruction” (SEI), these new committees are called the SEI Steering and SEI Implementation Committees. The SEI Steering Committee is made up of senior leaders, faculty, and students on both campuses, and provides strategic guidance and oversight for the Implementation Committee, which is tasked with operationalizing the implementation of the recommendations at both campuses. Please see Appendix 3 for membership of these groups.

1.1 PROJECT OVERVIEW

To address the recommendation by the Working Group to revise the existing University questions, the SEI Implementation Committee developed an eight-step project plan (see Figure 1). This plan included a mixed-method approach that collected qualitative feedback from student and faculty participants through focus groups and interviews, revised the questions based on this feedback, then conducted pilot-tests of the new questions using an online survey, and finally conducted a quantitative analysis of the results to see how well the revised items functioned.

Two questions did not function as well as expected, so we collected additional qualitative data from students on their interpretation of these items and made further refinements based on their comments. A final set of six core UMI questions are recommended to the Vancouver Senate Teaching and Learning and the Okanagan Senate Learning and Research Committees for their consideration and endorsement for implementation starting in Winter Term 1 2021/22 courses.

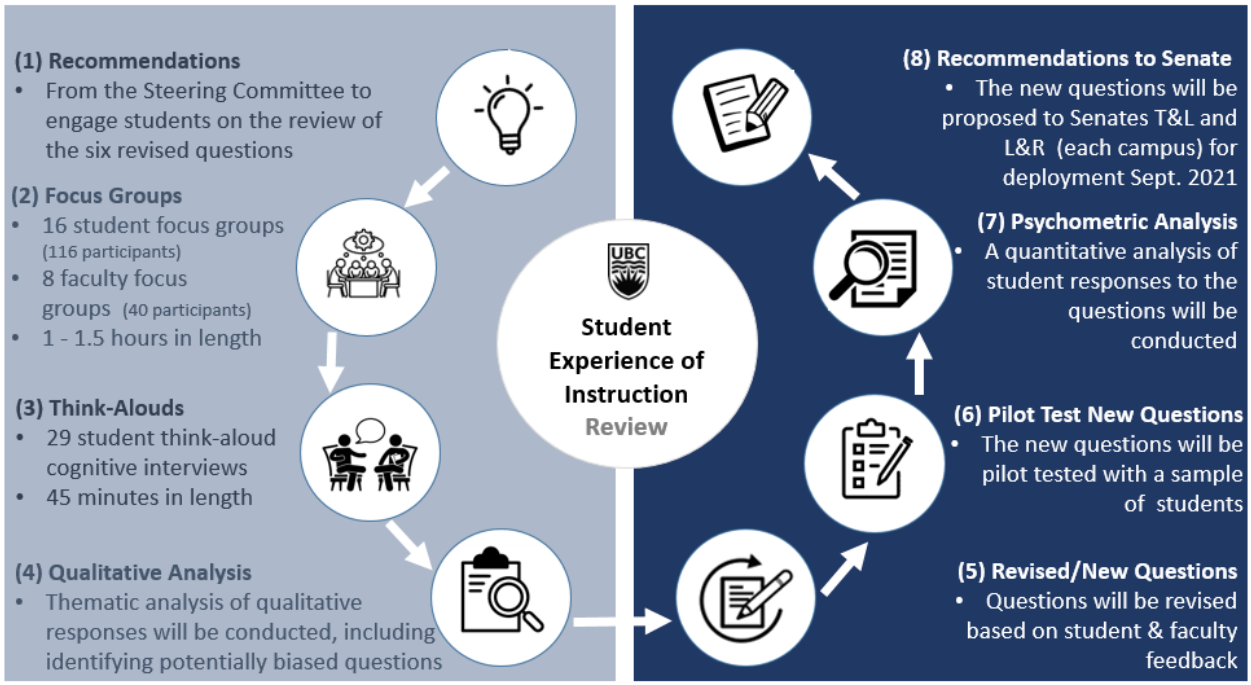

Figure 1. Eight-Step Project Plan to Evaluate the Proposed SEI Questions

2.0 Methodology

2.1 FOCUS GROUPS

We held 16 focus-group sessions with a total of 116 students across both campuses, all year levels, undergraduate and graduate, and across a diversity of programs. Each focus-group session was conducted online using Zoom and took between one hour and 1.5 hours to complete. Upon permission of the participants, each session was digitally recorded for later transcription. All students who participated in the focus group session received a $20 electronic gift card of their choice.

The goal of the focus groups was to introduce the six proposed questions and to gain an understanding of how students interpreted and would respond to the survey questions. Further, we asked them to identify any possible confusion that might occur in terms of different interpretations of the questions, and suggestions on how to improve the questions that might be understood differently by students or in different environments, such as a large- or small-class setting or class type.

We asked participants to think about their experiences of receiving and completing the former student evaluations of teaching questions, and asked them if they knew what the surveys were used for at UBC. We shared with the participants highlights from the report and recommendations made by the SEoT Working Group. We then introduced participants to the proposed six UMI questions put forward by the SEoT Working Group and asked them to provide their overall impression of the proposed changes to each of the questions.

We walked the student participants through each of the six UMI questions, asking them to discuss the following for each question:

- What is your understanding of the question?
- How would you respond? Does your response reflect the change in the question?
- Is this question confusing? Are there any words which need further defining or is there a better word to use? Do you think students could interpret this question differently from each other? Can you think of anyone who might be able to interpret this question differently from you?
- Would you interpret this question differently if you were enrolled in a small class compared with a large class?
- Would you interpret this question differently if you were enrolled in [subject] compared with [subject]?

At the end of the focus group session we asked participants to reflect on the following question: “Of all the things we’ve discussed today, what would you say are the most important issues, in terms of refining the new questions on the student experience of instruction survey?”

We also held eight one-hour focus group sessions with faculty members, of which four involved Okanagan faculty and four involved Vancouver faculty. In total, 40 faculty members participated in the sessions, coming from a range of programs, and employed in tenure-track and non-tenure track positions. Again, we asked faculty participants to provide insight on how they interpreted the proposed questions and their thoughts on how students would understand and respond to the questions. We also collected suggestions on how to reword the questions. Faculty members who participated in the focus group sessions did not receive any remuneration for their involvement.

2.2 THINK-ALOUD INTERVIEWS

In addition to the focus-group sessions, we conducted 29 one-on-one interviews with students who had not previously attended a focus group. Each interview was held online using Zoom and took between 45 minutes and one hour to complete. Upon permission of the participants, each session was digitally recorded for later transcription. Similar to the focus group sessions, all students who participated in the interview received a $20 electronic gift card of their choice.

The goal of the think-aloud interviews was to collect information from student participants on the six UMI questions by way of verbal feedback about their understanding of the questions, and how they process the questions to be able to respond to them. These types of interviews are known as think-aloud sessions, or cognitive interviews, and are very different from a focus group or a typical interview (Ryan et al., 2012; Trenor et al., 2011). Students are asked to verbalize everything they are thinking about as they read through the survey question and recall experiences and thoughts that inform how they would answer each question. The objective is for the participant to talk constantly as if they were alone in the room speaking aloud to themselves. It is a useful technique to gather information on whether students who complete the survey make sense of the question in the same manner as it was intended to be interpreted from the survey designer, or if they are struggling to understand what the question is asking.

We began each think-aloud session by introducing the purpose of the interview and describing the process of a think-aloud interview. To get students feeling comfortable with the approach, the interviewer conducted a practice round with two survey questions from the UBC Undergraduate Experience Survey, which included, “I am proud to say that I attend UBC,” and “I feel a strong sense of connection to UBC.” Providing the participant with time to practice was an important step in this process because it enabled the interviewer to provide feedback on how well the student was thinking aloud and to encourage additional talking if necessary. In the practice round, the interviewer asked the participant to read each question aloud and verbalize their thoughts about the question itself.

The following suggestions were offered to the participant to consider while thinking about the question:

- What do you think this question is asking you?
- What are you thinking about while considering your response?

- What does the question mean to you when thinking about your experiences? Do you have any examples in mind?
- Are you thinking about something other than the question?
- Is there anything about the question that is confusing? What is it?
- Is the question vague?
- Are you able to answer the question easily?
- How did you arrive at your answer?
- Do the response options capture your answers appropriately? If not, how would you want to respond?

Once the participant understood what was expected of them, the interviewer then asked the participant to "think-aloud" while reading through the proposed six SEI questions. The think-aloud interviews are considered "facilitator light," meaning that we want the participants to speak openly without too many prompted questions; however, students were prompted to give a response if they were silent for any long period of time, or if they seemed to be struggling and needed additional support from the interviewer.

Students were reminded that the aim of the interview was to evaluate the SEI questions, not the participant's performance nor their instructor's performance. We asked each participant to consider a lecture course they were currently enrolled in to use as an example when reviewing the question. We collected information about the course name and number, the year level of the course, the number of students enrolled in the course, if it was a required course for their program or an elective, if there was a teaching assistant (TA) assigned to the course, and if there was any additional information they wanted us to know about the course. With that course in mind, the participant began the SEI review using the think-aloud approach.

2.3 QUALITATIVE THEMATIC ANALYSIS

All focus-group sessions and interviews were digitally recorded and transcribed for further analysis. The transcriptions were uploaded into NVivo, a qualitative analysis tool, and participant comments were analysed to identify patterns of meaning, and organized into general themes. The themes were further refined and coded to aid in the interpretation of the data. A few members of the Implementation Committee were involved in the analysis of the qualitative data (see Appendix 3 for a list of Committee members). After individually analysing the qualitative data, the members met online to discuss the themes and any disagreements or differences they had with the interpretation of the data until agreement on the themes and interpretations was reached.

2.4 PILOT-TESTING THE REVISED QUESTIONS

The next phase of the project involved pilot-testing the revised survey questions developed from the thematic analysis. The 280 students who had indicated their interest in participating in the SEI project were contacted by email and asked if they would complete the pilot test of the revised questions through an anonymous survey link using the survey software program Qualtrics. In addition, students who had not previously participated in the project (through either focus groups or interviews) were invited by email to participate in the pilot survey by and asked to provide their feedback on the revised questions.

To collect contextual information, students were asked to provide some additional information at the start of the survey including: a course name and number that they were considering when responding to the questions; the number of students enrolled in the course; and whether it was a required course for their program or an elective. They were also asked to provide some additional information about themselves: whether they were an undergraduate or graduate student; at which campus they were enrolled; program of study; year level; and whether they were a domestic or international student. Participants were reminded at the start of the survey, and on each page of the survey, that this was a pilot project and that the focus of the survey was to review the revised questions, not the student's nor the instructor's performance.

2.5 ITEM RESPONSE THEORY AND DIFFERENTIAL ITEM FUNCTIONING

Quantitative data collected from the pilot survey were analysed using Item Response Theory (IRT) and Differential Item Functioning (DIF). IRT is an approach used for test development and can be used in a similar fashion for survey item development or refinement. Through IRT, we are able to: 1) predict individual survey responses based on a respondent's attitude or perception, and 2) to establish a relationship between an individual's item response and the set of traits underlying item performance through a function called the "item characteristic curve" (Hambleton et al., 1991). This information can help the survey developer evaluate how well the questions function across different attitudinal levels, and how well the response options work for each question.

DIF analyses examined whether students responded to the pilot survey questions differently across groups, such as focus-group participation, required vs. elective courses, class size, campus and year level. In surveys, DIF is conceptualized as occurring when survey respondents who have similar attitudes on a measured trait respond differently due to construct-irrelevant factors such as differential interpretation of terms used in the survey. If an item is flagged as having DIF it suggests that a survey question may indicate a different understanding across the student groups. When DIF is detected, further analyses examine why some items function differentially across respondents to determine whether refinement of the survey question is needed.

3.0 Findings

3.1 QUALITATIVE THEMATIC ANALYSIS

Most student and faculty participants supported re-writing the current UMI core questions from the perspective of the student. Participants from the Okanagan campus were overwhelmingly in support of reducing the number of items from 19 to six. Participants suggested that proposed questions from the SEoT Working Group were not consistently written as student-centred. They argued that simply adding "I think" to a question did not make it student-centred. In addition, participants interpreted some of the terms and phrases used in the proposed questions differently, and some participants suggested that terms could possibly lead to biased responses (e.g., the use of "concern" and the use of "communicated"). Much of the feedback from participants suggested that more clarity and specificity was required in the questions to reduce the potential ambiguity and multiple meanings that could be inferred from certain statements.

The results of the qualitative analysis were used to refine the questions with the aim to articulate the intention behind each of the questions clearly, and to relate them to the student-learning experience and feedback on instruction. Below is a list of the six UMI proposed by the SEoT Working Group in May of 2020, along with feedback from the student and faculty participants regarding each survey question. The revised wording on each question is included at the bottom of each of the sections below. These newly worded questions were used in the subsequent pilot survey to test how well students responded to them.

Q1. The instructor made it clear what I was expected to learn.

There was quite a bit of discussion on this item, and a variety of interpretations were drawn across the focus-group participants. The diverse interpretations were grounded in a lack of clarity on what it was that "the instructor made clear" in the sentence. Some participants thought it referred to clear communication of the syllabus at the start of the course, while others thought it meant that the instructor spoke clearly about the expected learning outcomes at the start of each class, and others wondered if it referred to clarity around course learning outcomes or course objectives. Some participants interpreted "what I was expected to learn" to be about tests and assignments delivered throughout the term, while others suggested it could also include broad skills learned throughout the term that might not be directly tied to the stated learning objectives for the course. Most participants suggested that clarifying the timing of what is being referred to in the question, such as throughout the term or at the start of the term, would help with interpretation. They also acknowledged that not all courses have articulated learning objectives, but all do have course requirements, so that would need

to be kept in mind when refining the question further. Some participants felt that this question was not student-centred and was still focused on the instructor rather than the student experience.

As a result of the feedback, the proposed new question wording for the pilot survey is:
Q1. Throughout the term, the instructor explained course requirements so it was clear to me what I was expected to learn.

Q2. The instructor engaged me in the subject matter.

In this question, there was lack of understanding by participants of the term “engaged” used in the sentence. Some participants thought this referred to time spent participating in class, or communicating with the instructor one-on-one during office hours, or in a group setting. Other participants suggested this could refer to being engaged in a class because of the subject matter alone, or due to the way in which the instructor taught the course. Many argued that due to the lack of clarity in understanding the term “engaged,” participants could respond differently to the question based on their own interpretation, which might not reflect the original intention of the question. In addition, some felt that they might have difficulty responding to the question because they could feel engaged with the instructor’s teaching style but not engaged with the subject matter, given that it is not of their own interest. Many suggested that the question should be reworded to ask about the way in which the course was taught, and they also suggested that we did not use the word "engaged".

As a result of the feedback, the proposed new question wording for the pilot survey is:
Q2. The instructor conducted this course in such a way that I was motivated to learn.

Q3. I think that the instructor communicated the subject matter effectively.

Overall, participants thought it was a good idea to focus on the student experience of instruction and write the questions so they are student-centred. Yet many participants said that adding “I think” to the sentence does not make it student-centred, and some students indicated that it actually made them feel as if their feedback they provided to instructors on the evaluations were less important. In addition, the term “subject matter” was interpreted as being too broad, making participants unsure about how to answer the question. Some participants interpreted “subject matter” as referring to the course content, while others suggested it could include the field of study, which would imply more than the course content. As a result, many participants suggested using the term “course material” to make it specific to the actual course. There was also further ambiguity with the word “communicated” in this question. Some participants were not sure if this was referring to communication in terms of the announcements, emails, discussions, communication about course activities in Canvas, or if it referred to the communication style of the instructor. Some worried that if students interpreted the question to be

asking about the communication style of the instructor, ratings could be possibly biased against instructors with an accent, or for instructors for whom English is not their first language.

As a result of the feedback, the proposed new question wording for the pilot survey is:
Q3. The instructor presented the course material in a way that I could understand.

Q4. I have received feedback that supported my learning.

Across all focus-group sessions, participants thought that this question should include an adjective to describe the quality or timeliness of the feedback provided. They suggested that sometimes feedback could be given, but not necessarily in a way that informed them what they needed to do to improve in the course. Others provided examples of when they had received feedback too late in the term, when they did not have time to improve or prepare sufficiently for their next assignment/exam, or even when the course was almost over. As they read the question, some participants were not sure if they would interpret "feedback" as referring to grades, written/email communications, oral feedback given during class, out-of-class questions, or written feedback (e.g., from quizzes and exams). Also, some student participants indicated that they do not actively ask for feedback, or take advantage of instructor office time to ask for feedback, so they were unsure about how to respond to this question. Many participants also discussed how class size could influence how a student might respond to this question, and that instructors teaching large classes might not be able to provide feedback to students in the same manner that they would if it were a smaller class.

As a result of the feedback, the proposed new question wording for the pilot survey is:
Q4. Considering the type of class (e.g., large lecture, seminar, studio, etc.), the instructor provided constructive and timely feedback that helped me understand how my learning progressed during this course[1].

Q5. I think that the instructor showed concern for student learning.

For many participants, the word "concern" had a negative connotation to it and could be interpreted as "worried," "apprehensive," or "fearful". As such, it was mentioned that this could be quite confusing for certain students for whom English is not their first language. Participants also thought the word "concern" could be associated with an emotional reaction and could result in biased responses based on instructor personality or gender identity. Other students thought that it was a good question and that

[1] Results from the pilot survey indicated that further refinement of this question was needed, so the final recommended question is: *Considering the type of class (e.g., large lecture, seminar, studio, etc.), the instructor provided **useful** feedback that helped me understand how my learning progressed during this course.* This is discussed further in the report.

showing concern for student learning, and how well they progressed during the course, was a positive characteristic for an instructor and in alignment with quality instruction. Nonetheless, many participants thought this question needed to be more specific and should provide examples of what specific actions they were being asked to associate with an instructor who shows concern for student learning.

As a result of the feedback, the proposed new question wording for the pilot survey is:
Q5. The instructor showed genuine interest in supporting my learning throughout this course.

Q6. Overall, this instructor was effective in helping me learn.

Most participants agreed that this was a good closing question, either to summarize what was already asked or to cover additional aspects that were not evaluated in the previous questions. There were participants who said the question was both too vague and not as specific as the other questions in the survey, or they felt that the question was too similar to other questions, making it difficult to answer as a unique question. They suggested that further refinement of this question was warranted to make it more specific and to provide clarity on the criteria being used to determine the term "effective", or they recommended that the word be excluded from the question altogether. Many respondents commented on how similar questions 5 and 6 were and recommended making more of a differentiation between the two items.

As a result of the feedback, the proposed new question wording for the pilot survey is:
Q6. I learned a great deal from this instructor.[2]

3.2 QUANTITATIVE ANALYSIS

A total of 333 completed student responses to the pilot survey were received. Tables 1.a and 1.b provide a breakdown of some student demographics and course attributes of participants in the pilot survey. There were fairly balanced representations from students who had previously participated in a focus group or interview for the SEI project, and those who did not participate, as well as across program year level, class size, and whether the course was required or an elective. A larger number of students who participated in the survey indicated they were enrolled in a program at the Okanagan campus (76% of the sample) compared with students from the Vancouver campus (24%). A large majority, 76%, of the respondents were female.[3] Not all participating students answered all six UMI questions, resulting in 13

[2] Results from the pilot survey indicated that further distinction of this question compared with UMI 5 was needed so the final recommended question is: ***Overall****, I learned a great deal from this instructor.* This is discussed further in the report.
[3] Student gender is based on administrative records, which are currently recorded as a binary variable, Male or Female.

observations with partially missing data. Most of these analyses cannot be conducted on missing data, and so for two of the three methods described further in this document, a reduced sample of 320 responses was used in the final analysis[4].

In addition to the pilot survey data, and for comparative purposes, a sample of equal size was randomly drawn from the 2020/21 Winter (Term 2) SEoT data to see how the newly revised questions compared with the existing questions.

Table 1.a Distribution of Pilot Survey Responses by Student Demographics

Focus group participant	Number of responses
Yes	156
No	177

Gender	Number of responses
Female	232
Male	73

Campus	Number of responses
Okanagan	244
Vancouver	79

Residency	Graduate	Undergraduate	Total
Domestic	19	254	273
International	17	32	49
Total	36	286	322

[4] The Winsteps implementation of the Mantel-Haenszel is slightly different than usual Mantel-Haenszel computations in that cases with missing data are stratified at an estimated measure and so it does not delete cases with missing data (Linacre, n.d.). The Winsteps method was used in this project, so all 333 cases were analysed.

Table 1.b Distribution of Pilot Survey Responses by Year Level, Class Size and Course Requirement

Year level	Number of responses
1st	44
2nd	86
3rd	92
4th	80
5th	21

Self-reported Class size	Number of responses
1 – 49	85
20 – 99	70
100 – 199	98
200+	80

Course	Number of responses
A requirement	209
An elective	124

We used IRT to analyse the questions in the pilot survey. There are several assumptions of the data that need to be met before conducting and interpreting this IRT analysis: 1) unidimensionality of the measured trait; 2) local independence of the survey items; 3) monotonicity; and 4) item invariance. Unidimensionality means that all items on the survey are measuring just one underlying construct (e.g., quality of instruction) and that one main factor should explain most of the variance in the survey responses (Hambleton et al., 1991). When items on the survey have local independence, it means that the response to one item is independent of the other questions on the survey, except for the fact that they measure the same underlying construct. Monotonicity occurs when the probability of positively endorsing an item continuously increases as an individual's attitude/perception level increases. Finally, item invariance means that the estimated item parameters do not differ across different groups (e.g. domestic vs. international students), due to misunderstanding or misinterpretation of the questions. These assumptions were met for this analysis and therefore we were able to continue with interpreting the results.

Three methods were used to determine DIF and to see if the results corresponded across the different methods: 1) Mantel-Haenszel, 2) logistic regression, and 3) the cumulative logit approach. Rather than determining sample size requirements alone, researchers suggest that a combination of sample size and the number of questions on the survey should be considered together to determine if item parameters are estimated accurately in IRT models. Şahin & Anil (2017) concluded that a sample size of 250 with 30

items is viable for a 2-parameter model. Zumbo (1999) suggested that 20 test items can be successfully used to run a DIF analysis and have enough information to be able to match individuals on ability level and form meaningful groups. Due to the small number of items on the SEI survey (only six UMIs) and small sample size (N=333), we conducted further analysis to determine if our sample size in this analysis was adequate. We drew random sample sizes of 150, 250 and 300 from the pilot data and used each sample to estimate item parameters in a 2-parameter IRT model. For the 2020/21 Winter data, we used sample sizes of 320 and 500, 1000 and 2000. The model parameter estimates were examined as the sample size increased to gauge the stability of the model and parameter estimates and to ensure that a sample of 320 suffices to estimate model parameters. Additionally, for the Mantel-Haenszel method, the computation used (from the software program, Winsteps) relied on both the Mantel-Haenszel and Rasch procedures (e.g., 1-parameter model). For these types of procedures, researchers have suggested having at least 30 responses (Linacre, 1994), with valid findings demonstrated using short tests (4 to 39 items) and small sample conditions (100-300 responses) (Paek and Wilson, 2011). Based on these additional analyses, we felt that we satisfied the sample size assumptions to continue with the IRT and DIF analyses.

Factor analysis was used to test if all six UMI questions represented a single underlying construct measuring quality of instruction from the student perspective (unidimensional assumption). The results of the factor analysis showed that all six UMI items had high factor loadings, i.e. all six UMI questions represent one underlying construct. The Scree and Variance plots in Figure 2 summarize the results of the factor analysis. The elbow in the Scree plot in Figure 2 indicates minimal contributions from subsequent factors. The first factor explained more than 75% of the variation. These findings support the unidimensionality assumption for the IRT analysis.

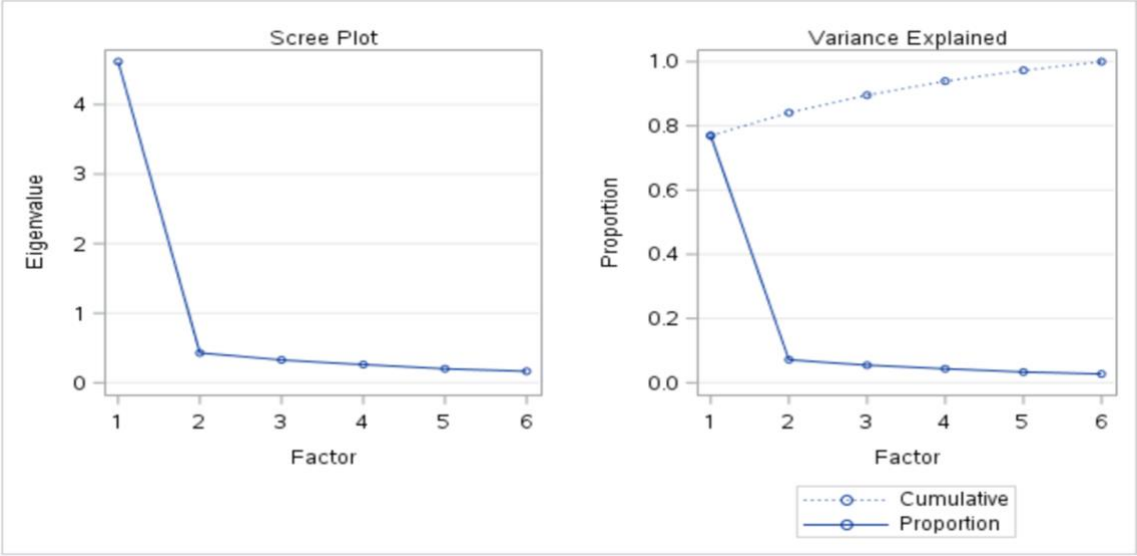

Figure 2. Scree and Variance Plots (UMI Pilot Survey)

Using DIF, we also examined whether students responded differently across groups, such as focus group participation, required vs. elective courses, class size, campus, year level, and student gender. The results of the DIF analysis will flag an item if it functions differently across participant groups, will indicate the direction of the DIF, and will also indicate if an item has uniform or non-uniform DIF. Uniform DIF occurs when DIF is the same for all attitude levels across the two groups, whereas non-uniform DIF occurs when there is an interaction between attitude levels and group membership.

The Mantel-Haenszel procedure is a commonly-used approach for detecting DIF. The Mantel-Haenszel method was run using the software program, Winsteps, which stratifies the sample by total survey scores to determine appropriate "attitudinal" groupings (Linacre, n.d.). To interpret the magnitude of DIF, we followed the criteria as defined by Zwick et al. (1999):

a) none or negligible DIF was detected if the absolute value logits were less than 0.43;
b) slight to moderate DIF was detected with absolute value logits between 0.43 to 0.64, and $p \leq 0.05$; and
c) moderate to large DIF was detected if the absolute value logits were larger than 0.64 and $p \leq 0.05$.

We used SAS statistical software to run the logistic regression model approach (Proc Logistic) and the generalized linear model procedure (Proc Genmod) for the cumulative logit method. In the logistic regression model, DIF is detected if individuals matched on attitude/perception have significantly different probabilities responding to a survey question and therefore will have differing logistic regression curves. We followed a three-model approach for the logistic regression method. The first model used a binary approach for the dependent variable (e.g., UMI survey item), where responses on the Likert scale of 4 "agree" and 5 "strongly agree" were combined and coded together as "favourable." A logistic regression model was fit to the binary data as a function of "attitude/perception" as measured by the overall survey score. The second model includes both "attitude/perception" and a variable representing the reference and focal groups of interest, such as gender. Finally, the third model included the variables in the second model and an interaction term (e.g. attitude/perception*gender).

Model 1: $\boldsymbol{Logit(P) = \beta_0 + \beta_1 \theta}$
Model 2: $\boldsymbol{Logit(P) = \beta_0 + \beta_1 \theta + \beta_2 Z}$
Model 3: $\boldsymbol{Logit(P) = \beta_0 + \beta_1 \theta + \beta_2 Z + \beta_3 \theta Z}$

Where: Logit(P) is the logit of the probability of respondent's endorsement;
$\boldsymbol{\beta_0}$, $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$ and $\boldsymbol{\beta_3}$ are model parameters;
θ denotes the value of the responder attitude/perception as measured by total score; and
Z denotes group membership (e.g. gender or focus group)

The cumulative logit-model method applies a similar three-model approach, except that the dependent variable uses the ordinal response scale values (Likert scale strongly agree "5" – strongly disagree "1") of the dependent variable (e.g., UMI survey item) and fits a cumulative logit function. For both approaches, a significant difference in fit statistics between models 1 and 2 i.e. a significant $\boldsymbol{\beta_2}$ would indicate uniform DIF, whereas a significant $\boldsymbol{\beta_3}$ in model 3 would indicate non-uniform DIF.

The results of the DIF analysis between different groups of student demographics and course attributes are summarized in Table 2 below.

Table 2: Differential Item Functioning (DIF) between different student groups and course attributes

Test Method	Grouping						
	Focus group Participation (Yes vs. No)	Course (Required vs. elective)	Class size (< 100 vs. > 100)	Class Size (1-49 vs 200+)	Campus	Year level 1st & 2nd vs. 3rd & 4th	Student Gender**
Mantel-Haenszel Procedure	None	None	**UMI 3**	**UMI 1**	None	None	**UMI 6**
Logistic Regression Models*	None	None	**UMI 1** **UMI 3**	**UMI 1**	None	None	None
Cumulative Logit Models*	None	None	None	**UMI 1**	None	**UMI 1**	**UMI 6**

*DIF significance based on p-values < 0.05; **Student gender is based on administrative records, which are currently recorded as a binary variable, Male or Female.

Results reported in Table 2 indicate that DIF was not detected, or was negligible for most of the groupings. DIF was detected for both class-size categories, year level and gender. Across all three methods, UMI question 1 showed moderate DIF between the smallest and largest class sizes (enrolments of 1-49 compared with classes with 200+ enrolments), with more positive responses given to the largest class size over the smallest (DIF, 0.67 and p-values of 0.006, 0.001 and 0.003 for the 3 methods, respectively). UMI 1 also exhibited non-uniform DIF between the lowest and highest year levels using the cumulative logit model (p=0.03), where 1st and 2nd year students provided more positive responses compared with students in their 3rd and 4th year, but did not show DIF using the other approaches. There

was slight DIF detected (DIF 0.43 and p-values of 0.03 and 0.01 for methods 1 and 2, respectively) for question UMI 3 comparing class sizes over 100 to those below 100 (again favoring the larger class sizes), and in UMI 6 (DIF, 0.46 and p-value of 0.03 for both method 1 and 3) for student gender; female students were more positive in their responses to this item. The UMI 3 and UMI 6 DIF results were not consistent across the different testing methods; therefore, these results were inconclusive. Fit statistics for DIF analysis using logistic and cumulative logit models are shown in Appendix 2. It is also worth noting that class size was self-reported by students and there was some inconsistency in the reported class size information with the same course names, which may be influencing the results of the DIF analyses.

There were fewer graduate and international student participants in the pilot survey; nonetheless, there was no differential functioning between graduate and undergraduate nor between domestic and international students. There was no DIF in all UMI questions between students who participated in the focus group discussions and those who did not participate, and no DIF based on whether the course was a requirement for their program of study or a chosen elective.

Finally, a two-parameter IRT model (graded response model, using Marginal Maximum Likelihood estimation method) was used to assess item response characteristics, item information and overall information functions, and to evaluate whether similar profiles were found between the pilot data and a comparable random sample from the 2020/21 version of the survey. A two-parameter IRT model estimates the difficulty and discrimination parameters of the survey items along the attitudinal scale of respondents. Random samples of size 150, 250 and 300 were drawn from the pilot data, and used to estimate the 2-parameter IRT model. Also for the 2020 winter data, model estimates were compared for the sample sizes of 320, 500, 1000 and 2000. The results showed that changes in parameter estimates were negligible as the sample size is increased. This indicates that the model is stable and that a sample of 320 can be used to estimate item parameters in the 2-parameter, unidimensional, IRT model.

The item difficulty parameter, or location parameter, which is perhaps a more appropriate term for this analysis, provides information on how difficult it is to achieve a 50% probability of a correct response for a specific item given the respondent's level on the underlying attitudinal scale. For example, if a student responds to UMI question 6, "I learned a great deal from this instructor," by answering with the most positive response option available, "strongly agree," this item would be located to the right or higher end on the attitudinal scale. A student who was very positive about the quality of instruction within the course would be more likely to have a 50% probability of endorsing the most positive response options for the UMI questions than a student with a more negative attitude about the quality of instruction within the course.

The item difficulty or location parameter also provides information on how the different response options (i.e., Likert scale options) function within each item. Although the UMI questions have essentially the same response options, with the exception of UMI 4 that has a "not applicable" option, the respondents

may not use the scale in the same equivalent manner across the questions. The item difficulty parameter estimates can provide information to the survey developers about the allocation of appropriate item and response-option weightings. Item difficulty parameter estimates (thresholds) were fairly consistent across response options for the six UMI questions (see Appendix 2 for IRT model parameter estimates), which indicates that the 5-point Likert scale options function similarly within each of the six new UMI questions. Reliability estimates were consistent across approaches; Cronbach's Alpha of 0.89 suggests a high survey reliability. Person and item reliability estimates were also generated for the Mantel-Haenszel procedure, ranging from 0.80 to 0.81 and from 0.84 to 0.85, respectively. The person reliability value suggests that the test discriminates the sample into enough levels while the item reliability value suggests that the sample is big enough for the analysis. The reliability estimate (Cronbach's Alpha) for the existing UMI questions from the 2020/21 sample was 0.94.

The item discrimination parameter indicates the strength of the relationship between an item and the measured construct, i.e., quality of instruction. It determines the rate at which the probability of positively endorsing an item changes given the individual attitude/perception levels (Thorpe & Favia, 2012). The higher the discrimination parameter, the steeper the slope will be on the item characteristic curve, indicating a stronger ability to detect differences in the attitude/perception of respondents compared with less steep slopes. The item discrimination parameter estimates (slopes) for the two-parameter IRT model are given in Table 3 for both the new UMI pilot survey questions and the random sample from the 2020/21 Winter (Term 2) version of the survey (the UMI questions currently in use). Typically, the larger the discrimination parameter, the steeper the slope, which implies that the item is more effective at discriminating among different attitudes along the continuum. Thus, for a given level of endorsement, an item with a discrimination parameter of 8.5 would have more than 10 times the contribution to the survey information compared to an item with a discrimination parameter of 2.5. Yet a discrimination parameter of 8.5 is quite high, which is an indication that the survey question is not working properly. Reeve and Fayers (2005) suggest the useful range of discrimination values is from 0.5 to 2.5. Following their recommendation, the only item with a discrimination parameter value in that range for the existing questions is UMI 4, and for the pilot survey all items except UMI 2 fall within that range.

Table 3: Item Discrimination Parameter Estimates

Data Source	Parameter Estimates					
	UMI 1	UMI 2	UMI 3	UMI 4	UMI 5	UMI 6
Sample from 2020/21 W2	3.62	5.38	4.15	2.02	3.28	8.67
UMI Pilot Survey	2.45	3.28	2.62	1.84	2.47	2.58

UMI question 4 has the least relative discrimination in the existing question that asks if *the evaluation of student learning was fair* (2.02) and the new UMI question asking about *timely and constructive feedback* (1.84), indicating that this item does not discriminate as much as the other items, among different attitude/perception levels. A low discrimination estimate may imply that the item is too complex for respondents to answer. Overall, the parameter estimates in the new UMI questions have been reduced from those reported for the sample from Winter 2020/21 (Term 2), and they are now more consistent across the items and fall closer within the range of useful parameter values of 0.5 to 2.5.

Figures 3 and 4 display the Item Information Curves (IIC) for each of the new UMI questions, and for the existing UMI questions from the 2020/21 sample, respectively. The IICs measure the statistical information an individual item contributes to the overall survey. The x-axis is the individual's level of endorsement; a person with an endorsement level of 2 has a more positive attitude regarding the course than someone with a level of -0.2. The y-axis indicates the magnitude of the information provided by each of the survey items. Higher information signifies higher precision (or reliability) in differentiating among respondents (Reeve & Fayers, 2005). In addition, items should be well spaced across the continuum (x-axis).

There are notable differences evident when comparing the item information curves in Figure 3 and 4. Figure 3 indicates improvement in the relative contributions of UMI questions 1, 2, 3 and 5 to the overall survey information compared with the 2020/21 sample. There was also some slight improvement in the contribution of UMI question 4. The newly worded UMI items shown in Figure 3 appear to differentiate across a broader range on the x-axis than existing UMI items shown in Figure 4. The y-axis scales differ between Figures 3 and 4 as a result of the disproportionately large UMI 6 discrimination parameter (8.67) in Figure 4. Although UMI 6 has a relatively large discrimination parameter estimate in the existing UMI question, it appears to discriminate across a very narrow range on the x-axis and displays sharp peaks on the information curve, which implies that the item is not functioning well.

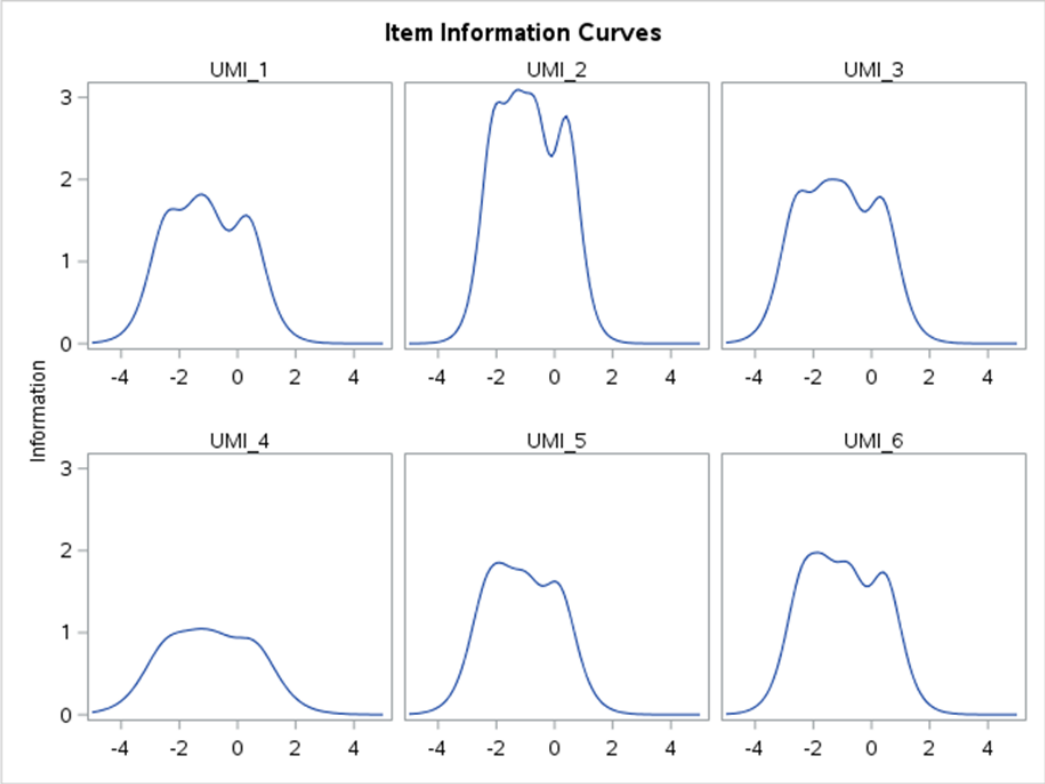

Figure 3: Item Information Curves for New UMI questions (UMI Pilot Survey)

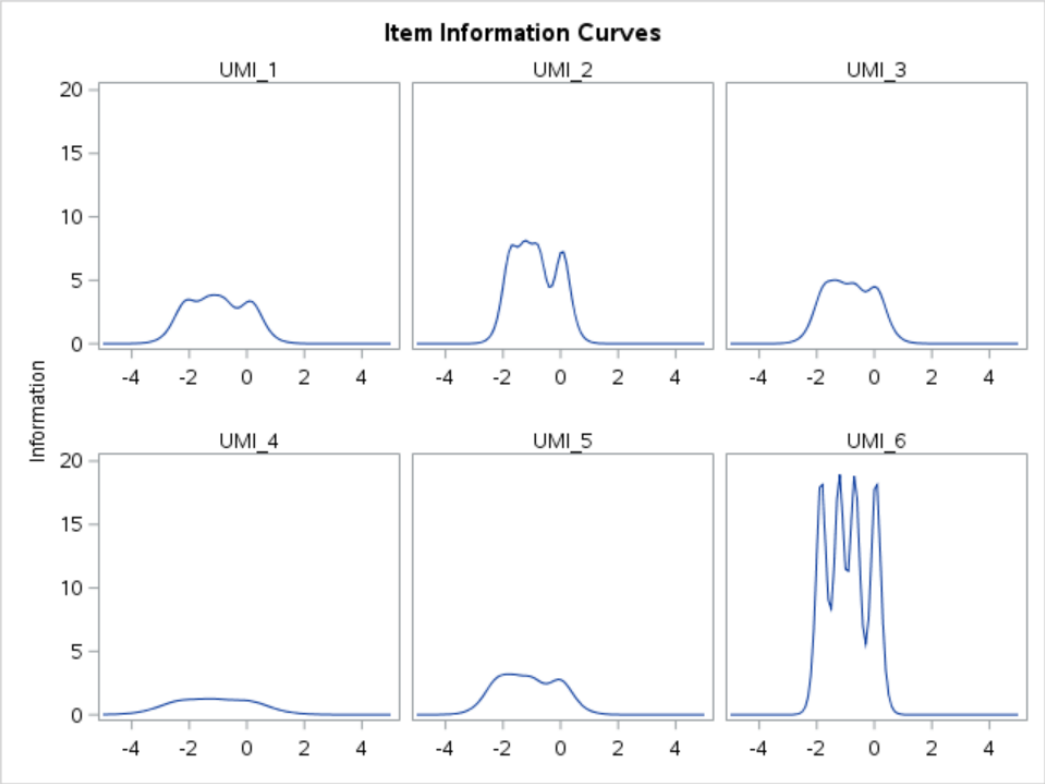

Figure 4: Item Information Curves for existing UMI questions (2020/21 W2 sample)

Looking at Figure 4, the IICs for existing UMI questions in the 2020/21 sample show that UMI 6 disproportionally contributes to the overall survey information; however, for the new set of UMI questions, the contribution of each item seems to be more consistent. Overall, the proposed changes to the UMI questions appear to have improved their relative discrimination among students with varying levels of endorsements for most items. While most of the newly worded UMI questions showed no DIF among different student groups, UMI 1 exhibited moderate DIF, and UMI 3 exhibited slight DIF between different class sizes. Slight DIF between genders was also detected for UMI 6, with female students positively endorsing that question more than male students (recall that only binary data are currently available for gender).

During the pilot survey, students were also asked to provide their feedback on the wording of the questions using an open-text field on the survey. Although most participants supported the changes to the questions, a few students indicated that UMI 4 may be asking about two different things: constructive *or* timely feedback. Some students also suggested that UMI 5 and UMI 6 still read as very similar questions, and they recommended further refinement to distinguish these questions from each other. Based on this additional feedback, and the results from the IRT and DIF analyses, questions 4 and 6 have been further refined. For UMI 4, we have removed the terms “constructive and timely” and replaced them with “useful” to simplify the question. UMI 6 has been revised to include the word “Overall” at the start of the sentence to capture more appropriately the comprehensive nature of that question and to further differentiate it from UMI 5.

4.0 Conclusion

Overall, the feedback from participants indicated support for a more student-centred questionnaire to be used for the end-of-term course evaluations. Participants from the Okanagan campus were overwhelmingly in support of the shorter core set of questions and for alignment across UBC campuses. Upon the recommendation of the SEoT Working Group, the six UMI questions were tested using a mixed-methods approach. Based on participant feedback during the focus-group sessions and the think-aloud interviews, further refinement of the proposed questions was warranted due to multiple interpretations of questions, and to the use of terms or words that could lead to potentially biased responses. The thematic analysis of the qualitative data provided information to refine the questions with the aim to reduce the potential ambiguity and multiple meanings that could be inferred from certain statements or words. Further, the qualitative data helped to articulate clearly the intention behind each of the questions and how each is related to the student learning experience and feedback on instruction, as well as being student-centred.

The results from the quantitative analyses indicated that the revised statements seem to function better than the existing questions. In the existing version, UMI question 6 provides most of the statistical information for the overall survey, but does not differentiate broadly among the respondents; sharp peaks in the item information curve indicate the item is not functioning well. The IIC results from the pilot test data provide preliminary evidence that the revised questions are much more consistent in their contribution to the overall survey, and are more widespread across the attitudinal continuum (x-axis). Although moderate DIF was detected in class size for UMI 1 and UMI 3 in the pilot survey, the class size variable was self-reported, and closer inspection of the data identified discrepancies on how the class size was reported, which could be falsely detecting DIF. In addition, the DIF results were not consistent across test methods and thus were not conclusive. The results did detect slight DIF for UMI 6, with respect to student gender, which suggests that further examination is needed to see how the newly worded questions function across demographic variables for students and instructors.

Based on these results, we recommend that the following new questions be adopted for implementation at UBC for the upcoming 2021/22 Winter Term and onwards.

Note: for the reader's reference, the previously proposed questions from the SEoT Working Group in May 2020 are included in grey italicized font below each of the newly recommended questions.

1. Throughout the term, the instructor explained course requirements so it was clear to me what I was expected to learn.
   The instructor made it clear what I was expected to learn.

2. The instructor conducted this course in such a way that I was motivated to learn.
   The instructor engaged me in the subject matter.

3. The instructor presented the course material in a way that I could understand.
   I think that the instructor communicated the subject matter effectively.

4. Considering the type of class (e.g., large lecture, seminar, studio, etc.), the instructor provided useful feedback that helped me understand how my learning progressed during this course.
   I have received feedback that supported my learning.

5. The instructor showed genuine interest in supporting my learning throughout this course.
   I think that the instructor showed concern for student learning.

6. Overall, I learned a great deal from this instructor.
   Overall, this instructor was effective in helping me learn.

Response options for all questions above: *strongly agree, agree, neutral, disagree, and strongly disagree.*

We also recommend that three common open-ended questions be included on all SEI surveys across both campuses to collect text comments:

7. Please identify what you consider to be the strengths of this course.

8. Please provide suggestions on how this course might be improved.

9. Do you have any suggestions for what the instructor could have done differently to further support your learning?

5.0 Limitations and Further Analysis

The scope of this project was restricted to reviewing the six core UMI questions proposed by the SEoT Working Group in May 2020. There are other questions that faculties, departments, or instructors may be using to collect additional information from students. Those questions were not included in this analysis.

The quantitative results of the analyses in this report were based on a small sample size (N=320) and only six UMI questions. Further analysis will be conducted on a larger data set collected during the fall deployment of the SEI to test the accuracy of item-parameter estimates and the detection of DIF for the newly worded survey questions. Additional work is required to determine how to support instructors interpreting responses to their SEI results between the new version of the UMI questions and the existing questions. Analyses regarding bias need to be conducted using both faculty and student demographic data. Currently, the demographic data required to conduct such an analysis were not available. The UBC Employment Equity Survey is used to gather demographic data from faculty and staff, but due to low response rates and non-random missing data they are not usable for these analyses. A revised Employment Equity Survey will be available to all employees starting in early Fall 2021, and a campaign to promote the completion of this new survey will begin at the same time, which should increase the number of responses and thus provide a more complete data set that will allow for a bias analysis. In addition, a student demographic project has been launched that will focus on collecting a broader range of student demographic data, similar to the questions asked in the Employment Equity Survey.

References

Columbia University Mailman School of Public Health (2019*). "*Item Response Theory." https://www.publichealth.columbia.edu/research/population-health-methods/item-response-theory.

Hambleton, R.K., Swaminathan, H, & Rogers HJ. Fundamentals of Item Response Theory. Thousand Oaks: Sage Publications; 1991.

Li, Zhusahn (2015). "A Power Formula for the Mantel-Haenszel Test for Differential Item Functioning." *Applied Psychological Measurement*, 39(5), 373–388.

Linacre, J.M. "Mantel & Mantel-Haenszel DIF statistics."(n.d.). http://www.winsteps.com/winman/mantel_and_mantel-haenszel_dif.htm

Linacre, J.M. (1994). "Sample Size and Item Calibration Stability." *Rasch Measurement Transactions*, 7(4), 328.

Paek, I. & Wilson, M. (2011). "Formulating the Rasch Differential Item Functioning Model Under the Marginal Maximum Likelihood Estimation Context and Its Comparison With Mantel– Haenszel Procedure in Short Test and Small Sample Conditions." *Educational and Psychological Measurement*, 71(6), 1023-1046.

Reeve, B., & Fayers, P.M. (2005). Applying item response theory modelling for evaluating questionnaire item and scale properties. In *In: Assessing Quality of Life in Clinical Trials: Methods and Practice, 2nd edn (ed Fayers, P.M.; Hays, R.D.), Oxford University Press, Oxford* (pp. 55-73).

Ryan, K., Gannon-Slater, N. & Culbertson, M.J. (2012). Improving Survey Methods With Cognitive Interviews in Small- and Medium-Scale Evaluations. American Journal of Evaluation, 33, 3, 414-430.

Şahin, A. & Anil, D. (2017). The Effects of Test Length and Sample Size on Item Parameters in Item Response Theory. Educational Sciences: Theory & Practice, 17, 1, 321-335.

Thorpe, G.L. & Favia, A. (2012). Data Analysis Using Item Response Theory Methodology: An Introduction to Selected Programs and Applications. Psychology Faculty Scholarship, 20. https://digitalcommons.library.umaine.edu/psy_facpub/20

Trenor, J.M., Miller, M.K., & Gipson, K.G. (2011). Utilization of a Think-Aloud Protocol to Cognitively Validate a Survey Instrument Identifying Social Capital Resources of Engineering Undergraduates. Paper presented at 2011 ASEE Annual Conference & Exposition, Vancouver, BC.

Zumbo, B. (1999). A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zwick, R., Thayer, D.T., & Lewis, C. (1999) An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. Journal of Educational Measurement, 36, 1, 1-28.

Appendix 1

Current and Proposed Changes to the University Module Items

CURRENT VANCOUVER CAMPUS CORE UMI QUESTIONS

1. The instructor made it clear what students were expected to learn.
2. The instructor helped inspire interest in learning the subject matter.
3. The instructor communicated the subject matter effectively.
4. Overall, evaluation of student learning (through exams, essays, presentations, etc.) was fair.*
5. The instructor showed concern for student learning.
6. Overall, the instructor was an effective teacher.

Response options: *strongly agree, agree, neutral, disagree, strongly disagree*
*UMI 4 includes *not applicable* as a response option

Open-ended questions differ amongst faculty and departments in Vancouver.

CURRENT OKANAGAN CAMPUS QUESTIONS

1. The instructor set high expectations for students.
2. The instructor showed enthusiasm for the subject matter.
3. The instructor encouraged student participation in class.
4. The instructor fostered my interest in the subject matter.
5. The instructor effectively communicated the course content.
6. The instructor responded effectively to students' questions.
7. The instructor provided effective feedback.
8. Given the size of the class, assignments and tests were returned within a reasonable time.
9. The instructor was available to students outside class.
10. The instructor used class time effectively.
11. The instructor demonstrated a broad knowledge of the subject.
12. Students were treated respectfully.
13. Where appropriate, the instructor integrated research into the course material.
14. The evaluation procedures were fair.
15. I would rate this instructor as very good.
16. The textbook and/or assigned readings contributed strongly to this course.
17. I found the course content challenging.
18. I consider this course an important part of my academic experience.

19. I would rate this course as very good.

Open ended Questions:

- What were the strengths of the course?
- What were the weaknesses?
- What did you most enjoy about it?

PROPOSED SEI QUESTIONS BY SEOT WORKING GROUP IN MAY 2020

Note: words in red font and italics below indicate wording changes proposed by the SEoT Working Group in May 2020 compared with the current Vancouver version of the UMI questions.

1. The instructor made it clear what *I* was expected to learn.
2. The instructor *engaged me in the* subject matter.
3. I *think* that the instructor *communicated the subject matter effectively*.
4. *I have received feedback that supported my learning.*
5. *I think that* the instructor showed concern for student learning.
6. Overall, this instructor was *effective in helping me learn*.

Appendix 2

Additional Model Statistics

Summary of DIF Analysis using Logistic Regression Models

Table A2.1 Logistic Regression Models: Class Size (1-49 vs. 200+)

		β_1		β_2		β_3		
	-2log L	**Estimate**	**p-value**	**Estimate**	**p-value**	**Estimate**	**p-value**	**DIF**
UMI 1								
Model 1	136.193	-1.2572	<0.0001					
Model 2	73.193			2.414	***0.0012***			***Uniform***
Model 3	73.192					0.0157	0.9716	
UMI 2								
Model 1	167.910	-1.2845	<0.0001					
Model 2	89.848			0.0364	0.9473			None
Model 3	89.801					-0.0767	0.8294	
UMI 3								
Model 1	151.276	-1.2272	<0.0001					
Model 2	87.250			-1.0936	0.0628			None
Model 3	87.142					0.1148	0.7430	
UMI 4								
Model 1	198.037	-1.1275	<0.0001					
Model 2	95.091			-0.7986	0.1452			None
Model 3	94.958					-0.1291	0.7172	
UMI 5								
Model 1	173.304	-1.0146	<0.0001					
Model 2	68.227			-1.2411	0.0667			None
Model 3	66.650					-0.6554	0.2561	
UMI 6								
Model 1	166.096	-1.2322	<0.0001					
Model 2	88.230			0.2190	0.6948			None
Model 3	88.107					0.1339	0.7268	

Note: Using logistic regression models to examine class size (1-49 vs. 200+), UMI question 1 exhibited uniform moderate DIF. This finding implies that students who self-reported in the survey that they were enrolled in larger classes (200+) provided more positive responses compared with students in self-reported smaller classes (1-49). Uniform DIF is the simplest type of DIF where the item exhibits differences in the location parameter but equally discriminates at all levels of the attitudinal scale.

Table A2.2 Logistic Regression Models: Class Size (<100 vs. 100+)

	-2 Log L	β_1		β_2		β_3		DIF
		Estimate	p-value	Estimate	p-value	Estimate	p-value	
UMI 1								
Model 1	136.193	-1.2572	<.0001					***Uniform***
Model 2	131.405			1.0429	***0.0344***			
Model 3	131.369					-0.0585	0.8484	
UMI 2								
Model 1	167.910	-1.2845	<.0001					None
Model 2	167.285			0.3279	0.4315			
Model 3	167.281					0.0191	0.9467	
UMI 3								
Model 1	151.276	-1.2272	<0.0001					***Uniform***
Model 2	143.600			-1.2753	***0.0085***			
Model 3	143.600					-0.00158	0.9956	
UMI 4								
Model 1	198.037	-1.1275	<.0001					None
Model 2	197.410			-0.2984	0.4298			
Model 3	197.408					0.00873	0.9710	
UMI 5								
Model 1	173.304	-1.0146	<.0001					None
Model 2	173.303			0.0103	0.9796			
Model 3	173.261					0.0454	0.8370	
UMI 6								
Model 1	166.096	-1.2322	<.0001					None
Model 2	165.932			0.1686	0.6857			
Model 3	165.474					-0.1860	0.5045	

Note: Using logistic regression models to examine class size (<100 vs. 100+), UMI question 1 exhibited significant uniform DIF as did UMI question 3, again favouring the larger class sizes. These findings imply that students who self-reported in the survey that they were enrolled in larger classes (100+) provided more positive responses compared with students in self-reported smaller classes (<100). Uniform DIF is the simplest type of DIF where the item exhibits differences in the location parameter but equally discriminates at all levels of the attitudinal scale.

Summary of DIF Analysis using Cumulative Logit Models

Table A2.3 Cumulative Logit Models: Class Size (1-49 vs. 200+)

	Full log L	β_1		β_2		β_3		DIF
		Estimate	p-value	Estimate	p-value	Estimate	p-value	
UMI 1								
Model 1	-296.844	-1.0401	<.0001					**Non-Uniform**
Model 2	-153.942			0.7422	0.0200			
Model 3	-149.161					-0.5113	***0.0027***	
UMI 2								
Model 1	-304.164	-1.1758	<.0001					None
Model 2	-154.989			0.1217	0.6980			
Model 3	-154.976					0.0248	0.8725	
UMI 3								
Model 1	-299.394	-1.0364	<.0001					None
Model 2	-154.641			-0.3007	0.3344			
Model 3	-153.747					0.2136	0.1838	
UMI 4								
Model 1	-343.483	-0.9096	<.0001					None
Model 2	-164.801			-0.4536	0.1431			
Model 3	-164.494					-0.1216	0.4339	
UMI 5								
Model 1	-291.248	-0.9807	<.0001					None
Model 2	-136.884			-0.2831	0.3999			
Model 3	-136.647					0.1127	0.4921	
UMI 6								
Model 1	-296.263	-1.1261	<.0001					None
Model 2	-153.569			0.2681	0.3909			
Model 3	-153.523					-0.0471	0.7610	

Note: Using cumulative logit models to examine class size (1-49 vs. 200+), UMI question 1 exhibited significant non-uniform DIF. This finding implies that students who self-reported in the survey that they were enrolled in larger classes (200+) provided more positive responses compared with students in self-reported smaller classes (1-49). Non-uniform DIF is more complicated than uniform DIF, where the item exhibits differences in the location parameter and differences across levels of the attitudinal scale.

Table A2.4 Cumulative Logit Models: Year Level (1st & 2nd years vs. 3rd & 4th years)

	Full Log L	β_1		β_2		β_3		DIF
		Estimate	p-value	Estimate	p-value	Estimate	p-value	
UMI 1								
Model 1	-296.844	-1.0401	<0.0001					***Non-Uniform***
Model 2	-230.868			-0.5958	0.0216			
Model 3	-228.507					0.2835	***0.0314***	
UMI 2								
Model 1	-304.164	-1.1758	<0.0001					None
Model 2	-241.712			0.0572	0.8194			
Model 3	-241.638					-0.0490	0.7013	
UMI 3								
Model 1	-299.394	-1.0364	<0.0001					None
Model 2	-236.728			-0.3228	0.2070			
Model 3	-236.630					0.0561	0.6577	
UMI 4								
Model 1	-343.483	-0.9096	<.0001					None
Model 2	-270.912			-0.3307	0.1774			
Model 3	-269.709					-0.1904	0.1229	
UMI 5								
Model 1	-291.248	-0.9807	<.0001					None
Model 2	-229.298			-0.0169	0.9483			
Model 3	-228.449					-0.1665	0.1952	
UMI 6								
Model 1	-296.263	-1.1261	<.0001					None
Model 2	-237.678			-0.2823	0.2623			
Model 3	-237.667					-0.0184	0.8821	

Note: Using cumulative logit models to examine year level (1st & 2nd years vs. 3rd & 4th years), UMI question 1 exhibited significant non-uniform DIF. This finding implies that students who self-reported in the survey that they were in the 1st and 2nd year of their program provided more positive responses compared with students in 3rd and 4th year of their program. Non-uniform DIF is more complicated than uniform DIF, where the item exhibits differences in the location parameter and differences across levels of the attitudinal scale.

Table A2.5 Cumulative Logit Models: Gender (Male vs. Female)

	Full Log L	β_1		β_2		β_3		DIF
		Estimate	p-value	Estimate	p-value	Estimate	p-value	
UMI 1								
Model 1	-296.844	-1.0401	<.0001					None
Model 2	-277.463			0.0631	0.8183			
Model 3	-276.868					0.1496	0.2745	
UMI 2								
Model 1	-304.164	-1.1758	<.0001					None
Model 2	-288.528			-0.2315	0.3895			
Model 3	-287.891					0.1487	0.2592	
UMI 3								
Model 1	-299.394	-1.0364	<.0001					None
Model 2	-286.928			-0.0474	0.8583			
Model 3	-284.817					0.2718	0.0411	
UMI 4								
Model 1	-343.483	-0.9096	<.0001					None
Model 2	-321.731			0.0385	0.8838			
Model 3	-320.904					-0.1814	0.2070	
UMI 5								
Model 1	-291.248	-0.9807	<.0001					None
Model 2	-275.617			0.4094	0.1341			
Model 3	-274.661					0.1853	0.1644	
UMI 6								
Model 1	-296.263	-1.1261	<.0001					***Uniform***
Model 2	-282.775			0.5675	***0.0311***			
Model 3	-282.324					-0.1290	0.3459	

Note: Using cumulative logit models to examine gender, based on binary administrative data available, UMI question 6 exhibited slight uniform DIF. This finding implies that female students answered more positively to this item than male students in the pilot survey. The majority of respondents were female (76%) which may have influenced the findings. Uniform DIF is the simplest type of DIF where the item exhibits differences in the location parameter but equally discriminates at all levels of the attitudinal scale.

IRT Model Parameter Estimates and Associate Statistics

Table A2.6 The IRT Procedure: Pilot UMI Questions

Item Parameter Estimates				
Item	**Parameter**	**Estimate**	**Standard Error**	**Pr > |t|**
UMI_1	Threshold 1	-2.39771	0.26767	<.0001
	Threshold 2	-1.50133	0.16862	<.0001
	Threshold 3	-1.09966	0.14039	<.0001
	Threshold 4	0.35646	0.11028	0.0006
	Slope	2.19052	0.28182	<.0001
UMI_2	Threshold 1	-2.05169	0.20537	<.0001
	Threshold 2	-1.31050	0.13837	<.0001
	Threshold 3	-0.64382	0.10446	<.0001
	Threshold 4	0.44359	0.10032	<.0001
	Slope	3.14382	0.41281	<.0001
UMI_3	Threshold 1	-2.52349	0.27780	<.0001
	Threshold 2	-1.57872	0.15881	<.0001
	Threshold 3	-0.86133	0.11487	<.0001
	Threshold 4	0.35722	0.09930	0.0002
	Slope	3.02993	0.40007	<.0001
UMI_4	Threshold 1	-2.56614	0.30786	<.0001
	Threshold 2	-1.50355	0.18223	<.0001
	Threshold 3	-0.76689	0.13341	<.0001
	Threshold 4	0.46683	0.12355	<.0001
	Slope	1.84834	0.24299	<.0001
UMI_5	Threshold 1	-2.15778	0.22601	<.0001
	Threshold 2	-1.59566	0.16526	<.0001
	Threshold 3	-0.86355	0.11711	<.0001
	Threshold 4	0.10887	0.09789	0.1330
	Slope	2.81557	0.37124	<.0001
UMI_6	Threshold 1	-2.40848	0.26671	<.0001
	Threshold 2	-1.70015	0.17895	<.0001
	Threshold 3	-0.69357	0.11413	<.0001
	Threshold 4	0.42394	0.10711	<.0001
	Slope	2.48822	0.31326	<.0001

Table A2.7 The IRT Procedure: 2020/21 Winter 2 Sample

Item Parameter Estimates				
Item	Parameter	Estimate	Standard Error	Pr > |t|
UMI_1	Threshold 1	-2.12152	0.17804	<.0001
	Threshold 2	-1.34732	0.11453	<.0001
	Threshold 3	-0.82825	0.08764	<.0001
	Threshold 4	0.14158	0.07877	0.0361
	Slope	3.61894	0.36565	<.0001
UMI_2	Threshold 1	-1.73431	0.13671	<.0001
	Threshold 2	-1.24201	0.09961	<.0001
	Threshold 3	-0.78375	0.08049	<.0001
	Threshold 4	0.06511	0.07379	0.1888
	Slope	5.38393	0.60850	<.0001
UMI_3	Threshold 1	-1.74377	0.14140	<.0001
	Threshold 2	-1.26538	0.10528	<.0001
	Threshold 3	-0.68511	0.08113	<.0001
	Threshold 4	0.06301	0.07617	0.2040
	Slope	4.14696	0.43046	<.0001
UMI_4	Threshold 1	-2.36069	0.22743	<.0001
	Threshold 2	-1.47748	0.14753	<.0001
	Threshold 3	-0.98293	0.11563	<.0001
	Threshold 4	0.08154	0.09161	0.1867
	Slope	2.01884	0.21183	<.0001
UMI_5	Threshold 1	-2.19906	0.18933	<.0001
	Threshold 2	-1.65832	0.13868	<.0001
	Threshold 3	-1.01607	0.09908	<.0001
	Threshold 4	-0.01400	0.07888	0.4295
	Slope	3.27754	0.33456	<.0001
UMI_6	Threshold 1	-1.84873	0.14098	<.0001
	Threshold 2	-1.21931	0.09336	<.0001
	Threshold 3	-0.67438	0.07437	<.0001
	Threshold 4	0.05644	0.07082	0.2128
	Slope	8.66758	1.46121	<.0001

Appendix 3

Steering and Implementation Committees Memberships and Consultations

The Steering committee and Implementation Group began work in the Fall 2020, and smaller groups also worked on specific items. Additional information can be found on the website seoi.ubc.ca.

Steering Committee, 2020-2021

Support: Debbie Hart, Senior Manager, Strategic Projects

Simon Bates	Associate Provost, Teaching and Learning, UBCV (Co-chair)
Moura Quayle	Vice Provost, Associate Vice-President Academic Affairs, UBCV, (Co-chair)
Stefania Burk	Associate Dean Academic, Faculty of Arts, UBCV
Sage Cannon	Students Union Okanagan - Faculty of Creative & Critical Studies Representative, UBCO
Julia Mitchell	Director, Communications & Marketing, Office of the Provost & Vice-President Academic, UBCV
Karen Ragoonaden	Chair, Senate Learning and Research Committee, UBCO
Rehan Sadiq	Professor and Executive Associate Dean, School of Engineering, UBCO
Naznin Virji-Babul	Assistant Professor, Physical Therapy Senior Advisor to the Provost on Women and Gender-Diverse Faculty, UBCV
Georgia Yee	Vice-President Academic and University Affairs, UBCV

Implementation Committee, 2020-2021

Support: Debbie Hart, Senior Manager, Strategic Projects

Christina Hendricks	Academic Director, CTLT, Professor of Teaching, Philosophy, UBCV (Chair)
Vanessa Auld	Professor / Head, Research Group Co-leader - Cellular Mechanisms of Development and Disease, UBCV
Breeonne Baxter	Communications Manager, VPA Communications, UBCV
Brendan D'Souza	Lecturer, Department of Biology, UBCO
Tanya Forneris	Interim Academic Lead, CTL, Associate Professor of Teaching, School of Health & Exercise Sciences, UBCO
Mark Lam	Lecturer, Department of Psychology, UBCV
Stephanie McKeown	Chief Institutional Research Officer (PAIR)

Marianne Schroeder	Sr. Associate Director, Teaching and Learning Technologies, CTLT, UBCV (Sept. 2020-Feb. 2021)
Abdel-Azim Zumrawi	Statistician, PAIR, UBCV (Feb. 2021 onwards)

Advisory group on changes to UMI

Christina Hendricks	Academic Director, CTLT, Professor of Teaching, Philosophy, UBCV
Stephanie McKeown	Chief Institutional Research Officer (PAIR)
Catherine Rawn	Professor of Teaching, Psychology, UBCV
Bruno Zumbo	Professor, Canada Research Chair in Psychometrics and Measurement, Tier 1; & Paragon UBC Professor of Psychometrics and Measurement Educational and Counselling Psychology, and Special Education, UBCV
Abdel-Azim Zumrawi	Statistician, PAIR, UBCV

Starting in the Fall of 2020, the Implementation Committee consulted with several groups, which have informed and provided feedback on the work of implementing the recommendations.

In addition to the work detailed above to test the new UMI, discussions were held with and feedback was collected from:

- UBC Vancouver:
  - Senate Teaching & Learning Committee
  - Associate Deans Academic, Students, and Faculty
  - Heads & Directors (at a Provost's Heads & Directors meeting)

- UBC Okanagan:
  - Senate Learning & Research Committee
  - Deans Council
  - Student Academic Success Committee

- Across both campuses:
  - Senior Appointments Committee
  - Open forum March 10, 2021 (over 100 faculty, staff and students joined)