An Investigation of the Interpretation and Use of Course Evaluations Data

EXPLORANCE FACULTY RESEARCH GRANT PROJECT

ALISON GIBBS, PHD, DIRECTOR (CTSI) & PROFESSOR (DEPARTMENT OF STATISTICAL SCIENCES)

CLARISSA LAU, PHD, DATA ANALYST, COURSE EVALUATIONS AND SOTL

CORA MCCLOY, PHD, FACULTY LIAISON COORDINATOR, SOTL

KYLE TURNER, MSC, FACULTY LIAISON, TEACHING AND LEARNING

KATHARINE LAUDER, BA, EVALUATIONS SYSTEMS SPECIALIST

Agenda

- Study Background and Research Questions
- Participants
- Preliminary Research Findings
  - Reading reports
  - Interpreting statistics
  - Interpreting scores
  - Using course evaluations to inform teaching
- Concluding Remarks
  - Next steps
  - Acknowledgements

Study Background

- Course evaluations offer a range of feedback, both formative and summative, to inform teaching and learning purposes.

- Even those who are knowledgeable about statistics are prone to misuses and misinterpretations of these data (Boysen, 2015; Linse, 2017; Spooren et al. 2013, Theall & Franklin, 2001).

- Therefore, we want to better understand how the design of reports, specific to statistics and data displays, are interpreted and used by a range of instructors.

Study Method

Survey (N=648)

- Teaching appointment, Division/Faculty, Years of teaching experience, Course sizes
- When and which sections of the report do instructors review?
- Self-report confidence of interpreting statistics and data visualizations
- Open-ended question to gather improvement feedback

Qualitative Interview (N=13)

- Overview of instructor's teaching experience
- Course evaluation report walk through
- Feedback on alternative data displays

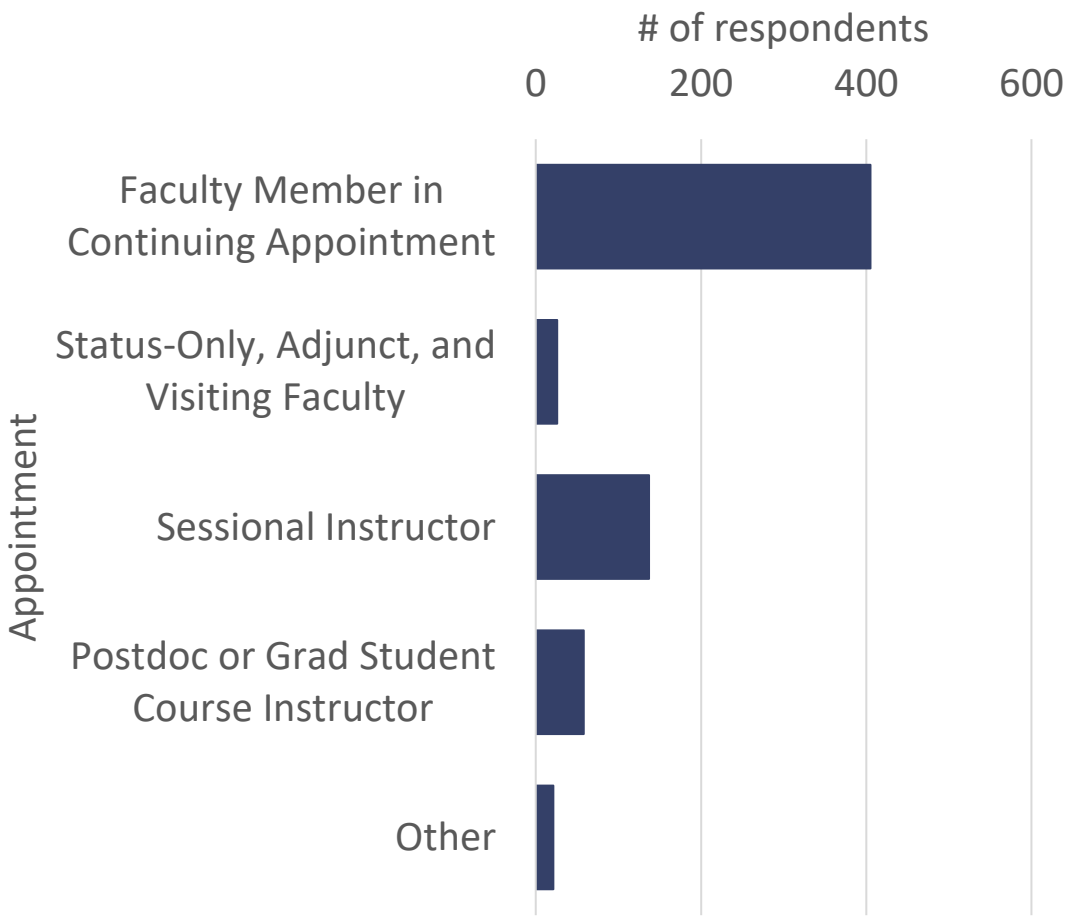
→ **Four broad questions**

Four Broad Questions

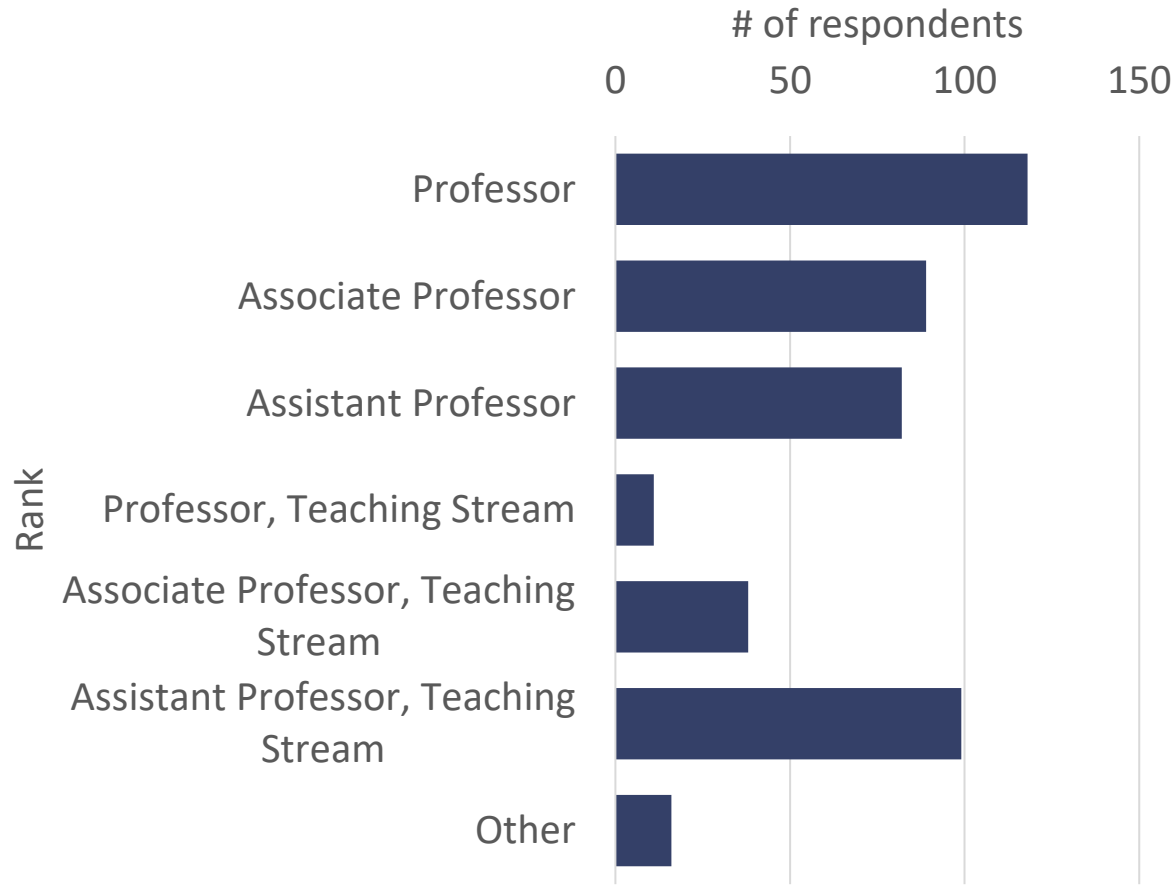
1. How do instructors read the course evaluation reports?
2. How do instructors interpret common summary statistics (e.g., mean, median, mode, standard deviation) in the course evaluation context?
3. Do instructors use the qualitative anchors and/or numerical scale to inform their course evaluation interpretation?
4. How do instructors use/interpret course evaluations to inform their teaching practices?

Survey Participants: Appointment & Rank

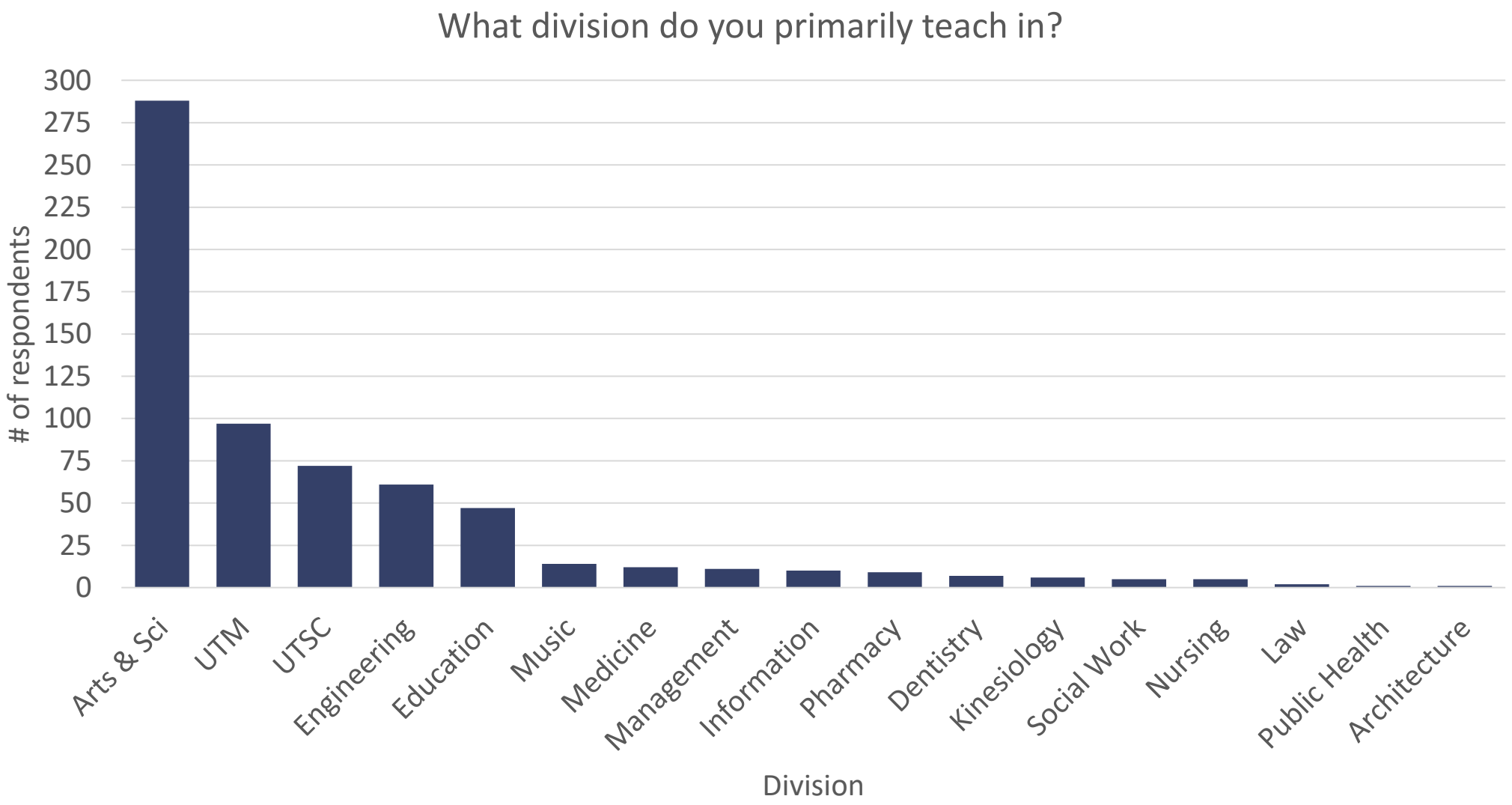
What is your teaching appointment type?



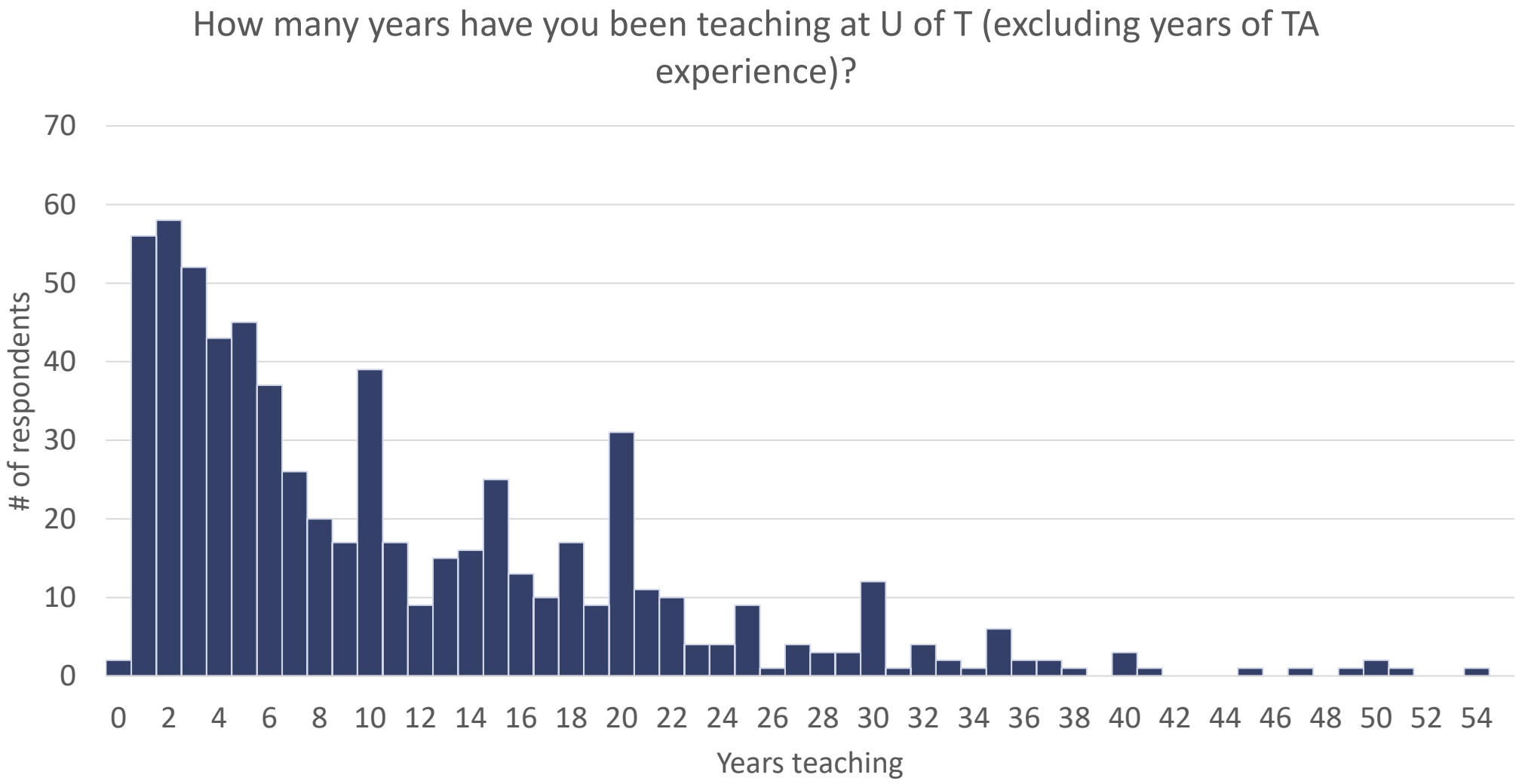
If you are a faculty member, what is your rank?



Survey Participants: Faculties/Divisions



Survey Participants: Years of Teaching

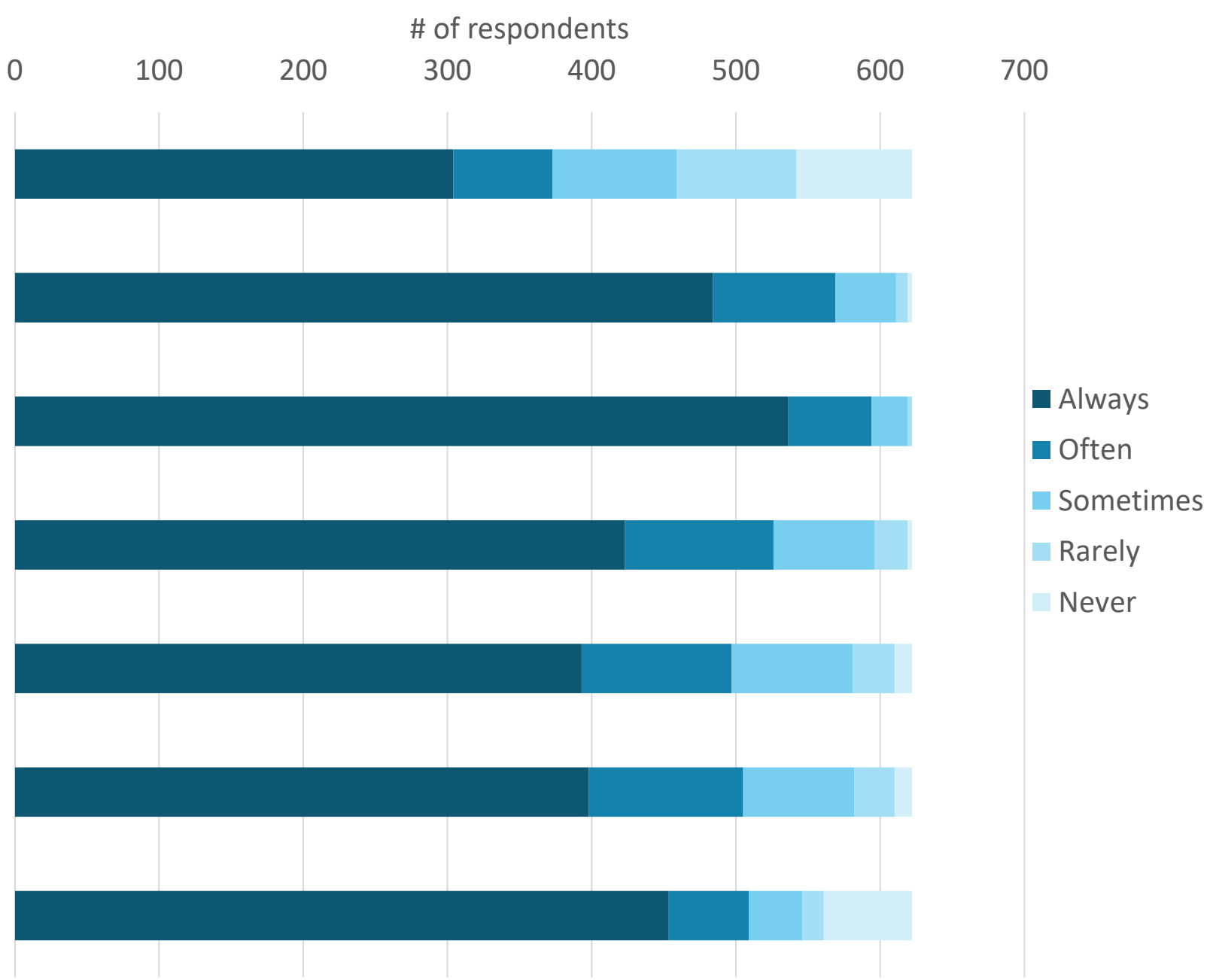


Interview Participants: Selection

- n=116 respondents (17.9%) volunteered for an interview
- n=27 were invited
- n=13 have been interviewed and analyzed so far
- Aiming for 15-20 interviews
- Selection criteria:
  - Divisional Representation (8 faculties/divisions) – purposive sampling
  - Appointment & Rank (Postdoc, Sessional, Assistant Prof, Associate Prof, Professor; Tenure & Teaching Stream)
  - Years Teaching (4 - 47)
  - Confidence with Quantitative Statistics & Visualizations (Not at all confident – Extremely confident)
  - Reviewed open-ended survey comments (if present) for feedback focused on reports

1: How do instructors read the course evaluation reports?

- Description
- Summary: Response Rate & Institutional Scores
- Qualitative Comments
- Divisional & Departmental Scores
- Distributions & Statistics
- Comparative Graphs
- Formative Instructor-Selected Items



Which section(s) of the reports do you review?

Description

Summary: Response Rate & Institutional Scores

Qualitative Comments

Divisional & Departmental Scores

Distributions & Statistics

Comparative Graphs

Formative Instructor-Selected Items

Interview Finding 1

- All interviewees start on the summary page

Raters	Students
Responded	4
Invited	6

Section 1: Course Evaluation Overview

Part A. Core Institutional Items

Scale: 1 - Not At All 2 - Somewhat 3 - Moderately 4 - Mostly 5 - A Great Deal

Question	Summary	
	Mean	Median
I found the course intellectually stimulating.	2.8	2.5
The course provided me with a deeper understanding of the subject matter.	3.3	3.5
The instructor ([NAME]) created an atmosphere that was conducive to my learning.	3.0	3.0
Course projects, assignments, tests, and/or exams improved my understanding of the course material.	3.8	4.0
Course projects, assignments, tests and/or exams provided opportunity for me to demonstrate an understanding of the course material.	3.0	3.5
Institutional Composite Mean	3.2	-

Description

Summary: Response Rate & Institutional Scores

Qualitative Comments

Divisional & Departmental Scores

Distributions & Statistics

Comparative Graphs

Formative Instructor-Selected Items

Interview Finding 1

- All interviewees start on the summary page

n=6 begin with the response rates:

> "So I always start here because **I'm curious about the response rate**, because if it's low, it's not helpful" (*KJ, Assoc Prof, Teaching*)

n=4 look at their institutional item scores first:

> "I skim. I look at my medians. (…) I know the questions already, so I skim. I'm like 'Oh, OK, this is good, they're all above 4's, that's good. Oooh, medians very high, that's nice. Quality 3.9, **OK, students basically liked the course**.'" (*CM, Assoc Prof, Teaching*)

n=8 then proceed to the comments or through the rest of the report in order.

Description

Summary: Response Rate & Institutional Scores

Qualitative Comments

Divisional & Departmental Scores

Distributions & Statistics

Comparative Graphs

Formative Instructor-Selected Items

Interview Finding 2

- All interviewees ignore the Description

- Sets of Items
  - Institutional
  - Divisional
  - Departmental
  - Formative Instructor-Selected
- Report Sections
- Statistical Terms Used in this Report
  - Mean
  - Median
  - Mode
  - Standard Deviation

Interview Finding 2

- All interviewees ignore the Description

n=4 acknowledged the value as a reference or for external readers:

> "…this is information that is maybe better for an outsider. If I were supplying these results to somebody at a different institution or who didn't know it well, **they would need this information to understand how to interpret the results**." (*CM, Assoc Prof, Teaching*)

> "…while I actually look at the content, **if I see something weirder, then I would just go back to the instruction**." (*SP, Asst Prof, Teaching*)

And many said they read it in the past:

> "So **when I first got there, it was a little bit interesting** just to say, oh, how what's how do they organize it here" (*RD, Prof*)

Interview Finding 2

- All interviewees ignore the Description

Similarly, **survey** respondents who had been teaching at U of T for fewer years tend to read this section more often:

	Do you read the Description?	
Years teaching	Always/Often	Rarely/Never
0-5 years	64%	21%
5-10 years	51%	37%
10-15 years	57%	33%
15+ years	53%	30%

But n=9 interviewees assume that the content doesn't change:

> "I don't look at the first 2 pages because **I know the content** and I'm expecting that is not changed between one year and another." (*SP, Asst Prof, Teaching*)

Interview Finding 2

- All interviewees ignore the Description

n=5 interviewees cited pre-existing statistical knowledge:

> "…but the things about math stuff—I think **I know mean, median, mode**." (*WZ, Postdoc*)

For **survey** respondents, statistical confidence did not have a clear relationship with reading this section:

	Do you read the Description?	
Quant statistics confidence	Always/Often	Rarely/Never
5 - Extremely confident	56%	34%
4 - Very confident	64%	22%
3 - Quite confident	52%	32%
2 - Somewhat confident	64%	19%
1 - Not at all confident	43%	40%

Interview Finding 3

- Despite indicating that they know the content, many interviewees were unaware of certain facts, which affected their use of the reports

The majority (n=10) of interviewees were confused about the meaning of the Institutional Composite Mean (ICM):

> "And the other thing I should say is I have no idea what that [the ICM] is. I don't know what the institutional (...). **What the h*** is that?** I have no idea." (*PM, Prof*)

> "This institutional composite mean, **I'm not sure exactly how it's calculated**, but I assume some kind of a mean for these items?" (*WZ, Postdoc*)

The result for some was to completely ignore this statistic:

> "No, because I don't remember. I believe that somebody told me this when I when these were being put together, but I've forgotten. **And so I just kind of don't pay that much attention to it**, to be honest." (*KJ, Assoc Prof, Teaching*)

Interview Finding 3

- Despite indicating that they know the content, many interviewees were unaware of certain facts, which affected their use of the reports

n=5 interviewees noted that this definition was likely in the Description, or went looking for it:

> "I'm going to be honest and say I'm not entirely sure what it means, and **I'm sure that what it says is right up above on those first few pages**." (*PK, Asst Prof, Teaching*)

Institutional Items

These eight items are consistent across the University of Toronto. They are comprised of:

- Five rating-scale items which represent institution-wide teaching and learning priorities.
  - **The institutional composite mean, a mathematical average of these first five items.**
- One rating-scale item on the overall quality of a student's learning experience.
- Two qualitative comment items.

1: How do instructors read the course evaluation reports?

- Instructors want to “get to the point” and skip information they feel they already know, yet key information is missed
- Takeaway: Don’t assume that boilerplate text will accurately guide the intended use of reports, and consider placing information directly where it’s needed

Course projects, assignments, tests and/or exams provided opportunity for me to demonstrate an understanding of the course material.	3.0	3.5
Institutional Composite Mean	3.2	-

***The Institutional Composite Mean is a mathematical average of the first five items.**

Scale: 1 - Poor 2 - Fair 3 - Good 4 - Very Good 5 - Excellent

Question	Summary

2: How do instructors interpret common summary statistics (e.g., mean, median, mode, standard deviation) in the course evaluation context?

Interview Finding 1

- Interviewees understand and are comfortable using and interpreting mean and median.

Concerns about outliers

> "As I understand it, **the one that is most subject to outliers is the mean**" (*PK, Asst Prof (Teaching), 5 – extremely confident with interpreting quant statistics and data visualizations*)

> "I mean the mean is the average if you add everything up and divide it, so it gives you some kind of basic average, I understand. The median is useful in a way because it gets rid of the highest and the lowest. **So if I'm worried about outliers, the median is going to be a more useful statistic for me**." (*CM, Assoc Prof (Teaching), 2 – somewhat confident with interpreting quant statistics and data visualizations*)

Interview Finding 1

- Interviewees understand and are comfortable using and interpreting mean and median.

General consensus from students

> "So for the mean and median, it is basically the definition of the mean and median. So I know what it is basically. So the difference between the mean, like the what's the average and then the median…**I mean the majority of the students gave basically**." (*RP, Sessional, 4 – very confident with interpreting quant statistics and data visualizations*)

Interview Finding 2

- Interviewees don't understand the use and interpretation of mode, and suggest to remove it.

n=4 share that they don't think the mode offers more information or they don't know how to define the mode.

> "The mode I don't like. The median I don't like…**I think one of them is enough…The reason is because mostly it doesn't give me that much information**…If my mean is 4.2, what (is) the mode gonna give me." (*RP, Sessional, 4 – very confident with interpreting quant statistics and data visualizations*)

> "The mode, I've never…, I don't. **I couldn't define the mode**" (*SJ, Asst Prof (Teaching), 3 – quite confident with interpreting quant statistics and data visualizations*)

Three instructors specifically mentioned the mode to be removed, and one instructor suggested to highlight the mode in the graphs.

> "**I might remove the mode in that you can read it from the graph**." (*WZ, Postdoc, 5 – extremely confident with interpreting quant statistics and data visualizations*)

Interview Finding 3

- Interviewees are aware of the concept of standard deviation, but their interpretations of this statistic are varied.

n=11 accurately define standard deviation in their own words.

> "And the standard deviation is just the **spread** of the curve." (*LE, Sessional, 5 – extremely confident with interpreting quant statistics and data visualizations*)

> "The standard deviation just tells you how **broad** it [the bell curve] is. So if you have a small standard deviation, it's very sharp." (*BC, Assoc Prof (Teaching), 3 – quite confident with interpreting quant statistics, 4 – very confident with interpreting data visualizations*)

Interview Finding 3

- Interviewees are aware of the concept of standard deviation, but their interpretations of this statistic are varied.

Instructors have different interpretations of how they use standard deviations…

- to get general consensus:

  "Again, it gives me the idea of what the consensus for the students is, **what the majority of the students think of**." (*RP, Sessional, 4 – very confident with interpreting quant statistics and data visualizations*)

  "So you kind of have a little bit of people with a little bit more of a polarizing view about your course. So there could be a subset of students who just don't like it. And then there's ones that really like it and then maybe you kind of look at it, maybe there are more extreme values on either ends, and **there's not really you know sort of a consensus**. (*TL, Asst. Prof, 3 – quite confident with interpreting quant statistics and data visualizations*)

Interview Finding 3

- Interviewees are aware of the concept of standard deviation, but their interpretations of this statistic are varied.

Instructors have different interpretations of how they use standard deviations…

- to identify statistical significance:

  "I guess when I look at…, when I think about standard deviations, I'm trying to understand whether or not something is statistically significant. I would be sort of trying to distinguish…**I would have 2 questions and I would try to figure out if they are statistically distinguishable**." (*PM, Prof, 5 – extremely confident with interpreting quant statistics and data visualizations*)

Emergent Interview Findings

- Interviewees mention bimodal or normal distribution

n=2 discuss how course evaluation ratings don't generate normally distributed graphs

> "But then again, **this is not a normal distribution**. So I tend to not look at the standard deviation cause it's…we hope it will be a skewed distribution. I don't think nobody wants to have a normal distribution around 3, right? Because then their standard deviation is the least of their problems." (*LE, Sessional, 5 – extremely confident with interpreting quant statistics and data visualizations*)

> "So standard deviation is in those same units as those things, but it doesn't really mean anything to me…**if I knew something about the distribution, then I would have some idea that like a certain percentage of people are within a certain amount of distribution**. But that's not true in all distributions. So I don't know, this number doesn't really tell me a lot." (*WZ, Postdoc, 5 – extremely confident with interpreting quant statistics and data visualizations*)

2: How do instructors interpret quantitative statistics in the CE reports?

- Most instructors primarily use mean and some use the median.
- However, there are instructors who don't understand mode or standard deviation and don't use the information to inform their interpretations.
- Takeaway:
  - In addition to definitions, could include visual examples to illustrate the statistical concept.
  - Also, can consider placing definitions directly where the information is presented.

3: Do instructors use the qualitative anchors and/or numerical scale to inform their course evaluation interpretation?

Interview Finding 1

- Most interviewees only use the numerical scale

n=7 prefer the numerical scale because it's what they believe students use to rate them and is easier to quantify.

> "I look at the numerical scale. It is from 1-5. **And that's basically how I think that I've been rated**." (*RP, Sessional, 4 – very confident with interpreting quant statistics and data visualizations*)

> "It is the numerical scale. Because again, I have a problem with all these language-based anchors. **I don't know the difference. I don't exactly know what they mean**, so to me, mostly, a great deal, it's just hard to quantify what makes that different. I think it's an arbitrary tick box for students. And so the numbers I think are just easier to interpret. 5 is higher than 4." (*KJ, Assoc. Prof (Teaching), 3 – quite confident with interpreting quant statistics and data visualizations*)

Interview Finding 2

- Few interviewees only review the qualitative anchors (Not at all, Somewhat, A great deal)

3: DO INSTRUCTORS USE THE QUALITATIVE ANCHORS AND/OR NUMERICAL SCALE TO INFORM THEIR COURSE EVALUATION INTERPRETATION?

n=2 prefer the qualitative anchors as they believe students are responding to the words.

> "Oh, yeah, **I ignore the numerical scale**. I look at the language…That's what the students are responding to. They're not reading the question and saying, "ooh, five," they're reading the question and they're saying "mostly." So the **students are responding to the words**, and I read the words." (*PM, Prof, 5 – extremely confident with interpreting quant statistics and data visualizations*)

> "I'm trying to only use the phrases…**Those are the things that are more relevant, like that is what the student is telling me**." (*WZ, Postdoc, 5 – extremely confident with interpreting quant statistics and data visualizations*)

Interview Finding 3

- A few interviewees use both numerical scale and qualitative anchors

n=3 use both the numerical scale and qualitative anchors to inform their interpretation.

> "(I look at) the descriptors. I think **they complement each other**, so I use both because the value, I mean, if you look at the mean it says 4.3, then you're like, OK, well most students are responding pretty positively to this item with "the course providing with a deeper understanding of the subject matter." But I still like to know if there are students that kind of respond on the lower end or if students respond mostly 5." (*TL, Asst. Prof, 3 – somewhat confident with interpreting quant statistics and data visualizations*)

Emergent Findings

- Interviewees raised the topic of student interpretation

n=2 were curious to know how students understood the scales (numerical and qualitative anchors).

> "So there is a question of whether **the students depict that when they answer, what they actually read the scale**, or what they actually know that there is a change of the scale. But I think it's marginal because the numerical scale it's just invariant." (*SP, Assoc. Prof (Teaching), 5 – extremely confidence with interpreting quant statistics and data visualizations*)

3: DO INSTRUCTORS USE THE QUALITATIVE ANCHORS AND/OR NUMERICAL SCALE TO INFORM THEIR COURSE EVALUATION INTERPRETATION?

Emergent Interview Findings

- An interviewee mentioned avoiding extreme ratings

3: DO INSTRUCTORS USE THE QUALITATIVE ANCHORS AND/OR NUMERICAL SCALE TO INFORM THEIR COURSE EVALUATION INTERPRETATION?

One instructor mentioned that they often avoid selecting the highest value, so 3 or 4 would be the general "ceiling".

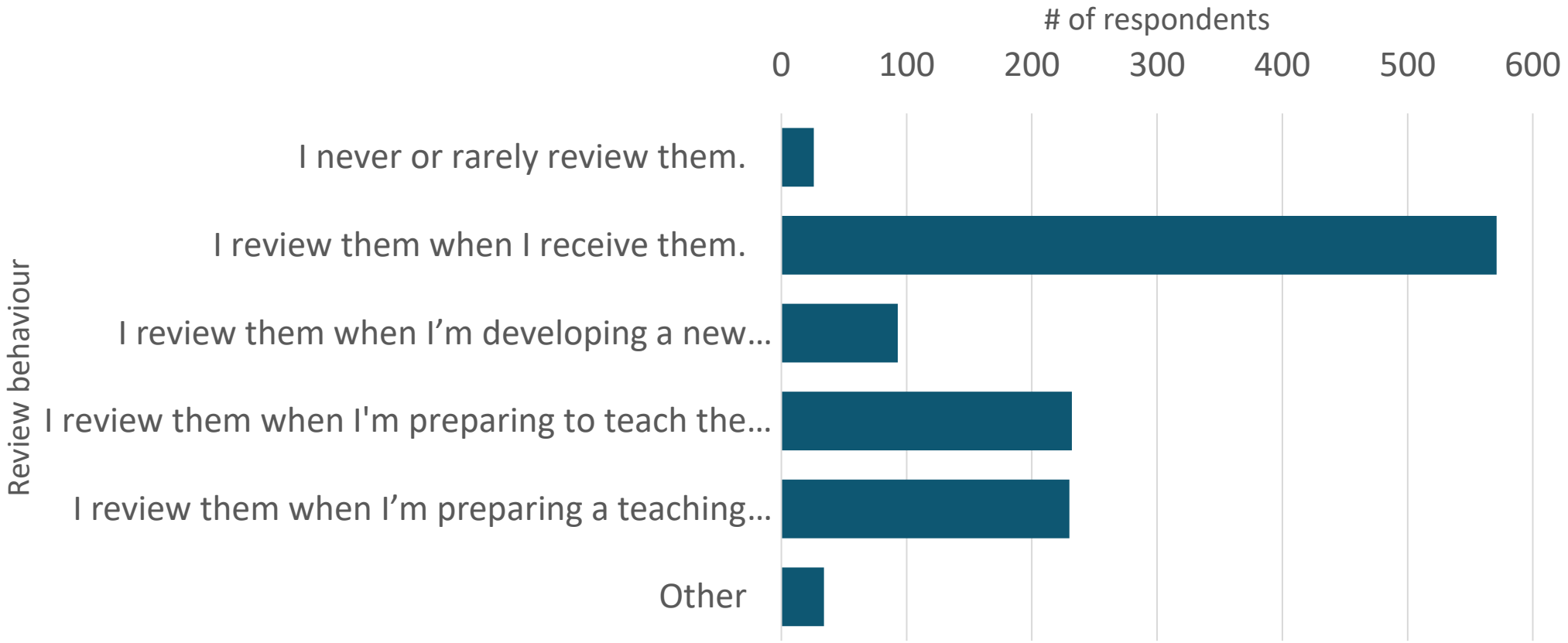
> "Yeah and I know from myself when I, I mean we are always asked to do surveys, right? And there's always some Likert scale on it and I'm always hesitant to give the highest value. Like, because I say OK for me **to give the highest value I have to be really wowed and it has to be really special**. So say like if I had a 1, 2, 3, 4, 5, probably like 3-4 would be my defaults and 5 would have to be special or 1 would have to be really crappy." (*BC, Assoc. Prof (Teaching), 3 – quite confident with interpreting quant statistics, 4 – very confident with interpreting data visualizations*)

3: Do instructors use the numerical scale and/or qualitative anchors in the CE reports?

- Instructors typically rely on the numerical scale.
- Takeaway:
  - Explore how students differentiate the numerical scale and/or qualitative anchors
  - Develop clearer guidelines to support instructors for using both the numerical scale and qualitative anchors to inform their interpretations.

4: How do instructors use/interpret course evaluations to inform their teaching practices?

Survey: When do you review your course evaluation reports?



Interview Finding 1

- The majority of interviewees open their CE Reports immediately & express varied emotions

n=7 open them when the email notice arrives:

"I'm looking forward to get them...) **I anticipate** the day that I can actually go and look at them" (*RP, Sessional*)

"I **look at them right away** within, you know, the same morning...and I look through them electronically and then I print them" (*SJ, Asst Prof, Teaching)*

"I'm always **very anxious** about it, to be honest" (*SP, Assoc. Prof*)

"Well, there's the **bracing myself**, and there's the deciding when am I actually going to click on that link?" (*PM, Prof*)

n=3 delay opening:

"I do not look at them until I'm finished teaching...because if they're bad**, I find them devastating.** And I can't take it…like it affects my confidence...And you know what? They're usually good" (*PK, Asst Prof, Teaching)*

"I guess the first thing I would do was say like, do I have time to think about and analyze what happened in this course? And I decided no. And **I moved on** to do some other things" (*WZ, Postdoc*)

Interview Finding 2

- Interviewees shared that CEs assess students' course experience, learning & enjoyment

The majority felt CEs offer insights on what the reports can assess about their teaching and their course:

> "I think they (…) provide a pretty good (…) **picture of how students respond** to my, both quality of my teaching and my (…) willingness to teach well…So maybe sometimes things are not perfect, but they appreciate my effort" (*SP, Asst Prof, Teaching*)

> "It kind of tells me that this about the **student experience**. It's really to see, OK, what do the students think about the course? And is there anything where they are clearly having some issues that I can change?" *(BC, Assoc Prof, Teaching)*

> "I think mostly they assess **how much students enjoy my teaching**...what they measure more than anything else is something kind of affective" (*MC*, *Assoc Prof, Teaching*)

> "**Student satisfaction**, meaning with certain items" *(LE, Sessional)*

Interview Finding 3

- Many interviewees chose to use the Formative Instructor-Selected items & reported value in this approach

n=7 instructors routinely use this option

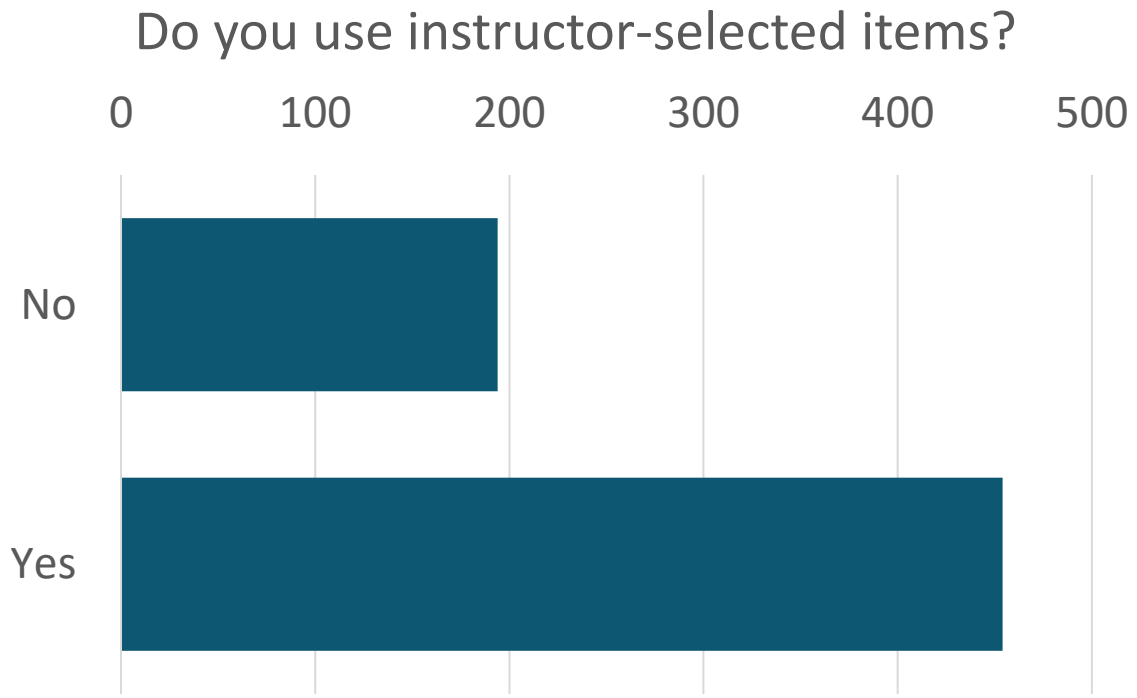
"Normally I select based on things that **I'm curious to learn more about**. I don't select ego stroke things...I select where I did something slightly differently that semester, or I select things where I'm curious if there might be a weakness" *(PM, Prof)*

"I think that if I have been given an opportunity to provide, to ask my own questions, if I don't use it, it is a wasted opportunity" *(RP, Sessional)*

"I typically use them if they're certain components of the course that I specifically implemented that are not captured with the divisional or the institutional items…So I wanna know that are those tutorials and labs sessions worthwhile? Like, am I just creating more burden for the students, do they like it? Do they **actually think it works? Or should I just like take it out and do something a little bit more useful?"** *(TL, Asst Prof)*

Formative Instructor-Selected items

Our Survey: 70% of respondents use Formative Instructor-Selected items



Interview Finding 3

- Some interviewees have abandoned use of Formative Instructor-Selected items

n=5 instructors had selected these items in previous years

"I haven't done it in years just because I felt, you know what**, I'm not getting any more out of it** than what I read in the comments" *(BC, Assoc Prof, Teaching)*

"I used to be consistent and always be really thoughtful in choosing those items and putting them on every single report and a few years ago I stopped because response rates were so low and they weren't telling me anything useful. **I just kind of gave up. I'm like, nope, this is not worth my time anymore**" *(CM, Assoc Prof, Teaching)*

"Well, I only think I did it once or twice a couple of years ago. I think it's a good idea. I just every year I get it…I've gone into it and like I can't. I don't even know why I should be selecting. You know, it requires me to step back a bit and think about what I wanna know at a time when I'm scrambling to finish" *(PK, Asst Prof, Teaching)*

4: How do instructors use/interpret course evaluations to inform their teaching practices?

- Most instructors are keen to promptly open and engage with their CE reports
- CE reports offer important student insights on instructors and their courses
- Formative Instructor-Selected items offer a more customized way to assess different course activities and teaching approaches
- Open-ended comments are very useful for course planning
- Takeaway:
  - Need to more closely examine the Instructor-Selected Item bank (e.g., number of items, uncertainty on how to select)

Concluding Remarks

Next Steps

- Continue interviews and analysis

- Draft future projects based on findings:
  - Student involvement in future CE interviews:
    - understanding their navigation of the course evaluation instrument
    - investigating their interpretation of qualitative anchors and numerical scale
    - examining their understanding of “quality of the course” item

  - Interviewing administrators and/or instructors with dual roles

Project Acknowledgements

- Gregory Hum, PhD, previous Manager of Evaluation and Assessment team
- Michele Hutyra, 3rd year undergraduate student, double major in Anthropology and Sociology
- Thank you to co-authors not in attendance
- Thank you Explorance for this research opportunity!

Presentation References

Boysen, G. A. (2015). Preventing the overinterpretation of small mean differences in student evaluations of teaching: An evaluation of warning effectiveness. *Scholarship of Teaching and Learning in Psychology, 1*(4), 269-282.

Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation, 54*, 94-106.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598-642.

Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research, 2001*(109), 45-56.

Questions?

Please contact us at ctsi.eval.assess@utoronto.ca if you have any questions about the project.