Guide for Academic Instructors

Last updated: March 2021

Student [Course Perception Surveys] are valuable when they ask the right questions, report response rates and score distributions, and are balanced by a variety of other sources and methods to evaluate teaching.

Stark and Freishtat, 2014

DRAFT

Contents

DRAFT

1. Introduction

Teaching assessment, and its longstanding reliance on student course perceptions (also referred to as "student evaluations of teaching" or "course evaluations") as a primary mechanism for data collection is riddled with conflicting perspectives, but most scholars in this field agree that student voices are essential when it comes to understanding how students perceive their learning experience, particularly at institutions of higher learning. This document is designed to guide instructors in interpreting their student course perception survey results. It will be updated and informed by ongoing testing and monitoring of the SCP process, discussions with stakeholders, experiences shared by other Canadian universities, and continued review of the research literature.

2. Reviewing your results

The SCP data report summarizes the results of the SCP survey for each instructor/course. To date, this report includes the results of core measures collected for institution-level analysis and will eventually contain the results of faculty- and department-specific measures.

The report provides an overview of all seven core survey items and their associated scores (Figure A) and statistics and a detailed view of the scores and statistics associated with each survey item (see Figure B). The report will also include the results of three open-ended questions.

The following figures (A and B) illustrate the (A) overview and (b) detailed view of the core (institutional) survey questions. Some faculties will develop faculty-level and/or department level questions in addition to the core items; the report will include similar figures for each level. The data in Figure A provide an overview of the scores for each question for a particular instructor for a specific course.

The data in Figure A tell us the mean and median of the scores students assigned to each of the survey questions. The figure also lists average score (for questions relating to design or implementation) and an overall average of scores for all the questions.

Each of the core questions is categorized as relating to course "design" or "implementation."

Figure A: Overview of survey questions and associated scores

	Question	Mean ⓘ	Median	Average	Overall Average
Design	The instructor identified the intended learning outcomes for this course.	4			
Design	The course activities prepared me for the graded work.	3.7		4.1	
Design	The intended learning outcomes were assessed through my graded work.	4.6			3.9
Implementation	The instructor helped me to understand the course concepts.	3.6			
Implementation	The instructor created a supportive environment that helped me learn.	4.1		3.7	
Implementation	The instructor stimulated my interest in this course.	3.3			

Questions relating to design will only be relevant for you if you were responsible for designing the course, and vice versa for questions relating to implementation. If you designed and implemented a course, both categories of scores will be relevant.

If you see any scores that stand out as different from the rest, refer to the histogram for that specific question for more information about what might be causing the discrepancy (Figure B).
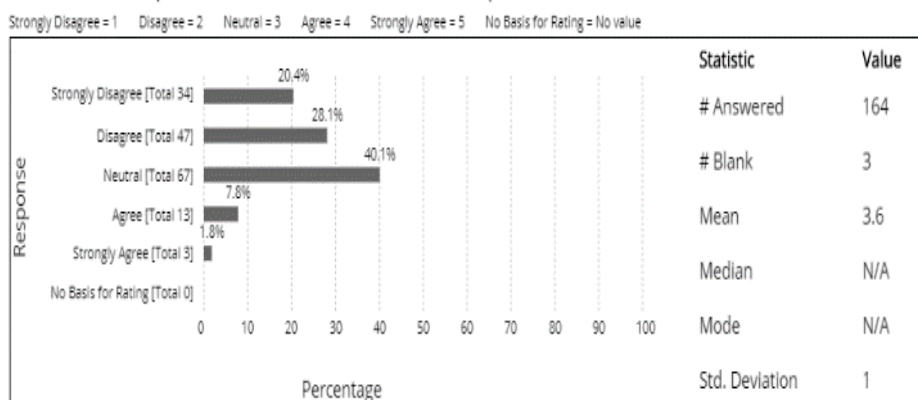
The data in Figure B tell us that almost half of the students in this class selected "neutral" in response to the item, "*The instructor helped me to understand the course concepts*," and that 40% of respondents

Figure B: Detailed view (histogram) of the scores associated with each survey

The instructor helped me to understand the course concepts.

Strongly Disagree = 1   Disagree = 2   Neutral = 3   Agree = 4   Strongly Agree = 5   No Basis for Rating = No value

Statistic	Value
# Answered	164
# Blank	3
Mean	3.6
Median	N/A
Mode	N/A
Std. Deviation	1

selected "strongly disagree" or "disagree." A total of 164 students responded to this survey item; #Blank indicates that three respondents did not answer this item. The mean shows that this item has an average score of 3.6 with a standard deviation of 1.

The following table defines and provides guidance for each of the numbers included in the SCP report.

Score	Definition	Guidance
# Answered/ # Blank	Number of students who replied to this question or left this question blank.	Impacts response rate
Average	Combined score for "design." Combined score for "implementation."	Relevant if you only designed the course or only delivered the course
Mean	Average score for each survey question	Provides information about the "typical score" for a specific question but *does not provide a complete picture*. May be affected by extreme (high/low) scores or outliers, which is particularly problematic for courses with a small number of students. Pay attention to the **distribution of ratings**. An unusually high or low score may require investigation but should not be considered in isolation.
Median	Score at the midpoint of the range of scores for each question.	Combined score for "design." Combined score for "implementation."
Mode	The number that occurs with the most frequency in this data set.	More reflective of the collective student experience than the "average," which can be impacted by outliers.
Overall average	Combined overall score; includes *both* "design" and "implementation"	Only relevant if you both designed and taught the course.
Standard deviation	Distribution of responses around the mean. Indicates the degree of consistency among responses. See page x for more information.	A smaller standard deviation (<1) typically reflects a high degree of consensus. A larger standard deviation (>1) typically reflects larger differences of opinion. Refer to comments to help understand this variation.

3. Understanding the Open-Ended Responses

Students' written comments can be a rich source of contextual data to help you understand the numerical scores on your SCP. Interpreting these comments can feel overwhelming; as our own worst critics there is a tendency to hyper-focus on and overemphasize any negative feedback, even when this feedback is outweighed by positive student feedback.

The Teaching Assessment Processes office is working with students, providing education about how to offer useful feedback to instructors and also emphasizing that inappropriate comments of any kind (e.g., sexist, racist, discriminatory, or harassment) will not be tolerated. If you receive a comment of this nature we encourage you to contact the Teaching Assessment Processes office, the Office of Equity, Diversity, Inclusion & Anti-racism, or the Chair of your department so that action can be taken (e.g., that particular student's responses to all survey questions can be removed from the results).

Consider walking through the following steps when reviewing open-ended responses:

1) Sort the comments into two categories: positive feedback and negative feedback.
   - You could include a category for ambiguous or confusing comments that you can't immediately sort.

2) Review your mean scores, considering:
   - how accurately the comments reflect your mean scores;
   - whether the comments were positive or negative o*verall;*
   - the total number of comments in each column.

3) Look for patterns.
   - Are there things students seem to consistently identify?
   - Are most comments focused on course design or course delivery?

4) Highlight any comments that reference scheduling, class length, timing and frequency or class composition, as these are items that are out of an instructor's control and should be discussed with your department Chair.

Appendix A: Response rate

If every student in the class completed the SCP, the response rate would be 100% and indicative of the true collective experience of the entire class. In reality, however, a response rate of 100% is extremely rare so we rely on an *estimate* of the true experience of the collective and **estimates always include some degree of measurement error.**

Don't assume that the average score represents the collective experience of the entire class. Exercise caution, especially when there is a low response rate. Scores associated with higher response rates are a better reflection of the collective experience.

Class size and response rate

Generally speaking, smaller classes require a higher response rate to achieve the same confidence that scores reflect the collective experience. Larger courses result in more data than smaller courses, so even with similar response rates, a smaller class is more likely to achieve a "precise" estimate of the overall experience.

There is a high degree of inaccuracy in mean scores calculated in samples with fewer than 20 respondents. In a smaller class, one unhappy student can seriously undermine the overall average. The outlier cases can have a significant impact on the overall score. **With smaller classes, it is extremely important for administrators to review the distribution of scores.**

Don't make assumptions about the collective class experience if the response rate is low. The size of the course determines what should be considered to be an adequate response rate (see Table 2). Consult Table 2 to determine how much confidence you can have in the scores based on the level of "precision" for the response rate in a particular course.

Table 2: Response rate necessary to have sense of students' collective experience based on class size. Adapted from University of Toronto data.

Interval around mean	Quality of mean estimate	Course size				
		1-25	26-50	51-100	101-200	200+
<±0.1	Very reflective of the collective experience	>90%	>80%	>80%	>60%	>50%
<±0.2	Reflective of the collective experience	>80%	>70%	>70%	>50%	>40%
<±0.5	Somewhat reflective of the collective experience	>70%	>50%	>40%	>20%	>10%
<±1.0	Generally unreflective of the collective experience	>60%	>20%	>10%	>10%	>10%
>1.0+	Not at all reflective of the collective experience	<30%	<10%	<5%	<3%	<1%

Appendix B: More about interpreting the mean score

The mean score provides information about the "typical score" for students' perceptions of the quality of instruction for a specific course but does not provide a complete picture. The five-point Likert scale used for the SCP survey is **ordinal**, not continuous: it uses a scale that arbitrarily numbers an ordered series of labels ranging from "strongly disagree" to "strongly agree." See sidebar: Ordinal vs. Continuous Scales. With an ordinal scale, the difference between a mean score of 3.9 and 4.2 is not overly meaningful.

A continuous scale measures numerical data. We can measure numerical differences in dollar amount. If Amy has $5 and Ping has $4, we can say that Ping has precisely 1 dollar more than Amy.

An ordinal scale orders nominal data (e.g., categories) to make it possible to measure it in a numerical way. For example, if Amy received an overall mean SCP score of 5 while Ping received an overall mean score of 4, we can say that Nancy obtained a higher score than Ping. To say that Amy is a "more effective teacher" by 1 point would be over-interpreting the numbers applied to the categories.

In fact, unless Amy and Ping taught the same course to the same cohort of students, any comparison of their scores is meaningless. All we can say is that a set of students rated Amy at a score of 5 and a different set of students rated Ping at a score of 4.

Consider the following fictional scenario:

Amir receives an overall score of 3.2. A closer look at the distribution in scores on the histograms for each item shows there are 4 extreme outliers (students who selected 2 on the scale for every survey item) but the rest of the scores are clustered between 4 and 5.

In this case, the mean score does not reflect most students' perception of this course.

Be very cautious in assigning significant weight to Amir's low average. This score is the result of only a handful of students' experiences in this course. It would be advisable to examine Amir's scores in other courses to get a clearer picture of student perceptions of his teaching performance.

Appendix C – More about standard deviation

Standard deviation indicates the variability of data—the degree to which SCP scores vary around the mean. A higher standard deviation means that there is high variability in the data. A lower standard deviation means that there is less variability in the data. A low standard deviation inspires more confidence that the mean represents the 'typical case.' Note: the standard deviation can also be affected by extreme outliers.

A higher standard deviation (SD) means that there is high variability (less agreement) in the data. Dan received a mean score of 4.5 on the SCP survey, with an SD of 2**.** An SD of 2 is quite large relatively speaking, which means that there was a lot of variation in students' responses. Anyone reviewing Dan's score should be cautious about interpreting it as reflective of the collective experience of all students in this class. *With a higher SD, administrators should take a closer look at the scores to see if they can identify discrepancies in students' experiences.*

A lower SD tells us that scores are close to the mean, meaning that there is less variability (more agreement) in the data. Mitra received a mean score of 4.5, with an SD of 0.5, which is quite small (see above comments). With a lower SD, we can be more confident that the mean measures the typical case.

Small standard deviation

Large standard deviation

(-2) 4.5 (+2)   (-0.5) 4.5 (+0.5)