

section 13.10.6 in the current (2008-2011) CBA

FAQ on the Use of Student Evaluations in the Promotion and Tenure Process

Section 13.6.2.6 of the CBA requires a statistical summary and frequency distribution of all required student evaluations of teaching.

- *What is meant by a statistical summary?* A statistical summary is a summary of the mean scores by question from the student evaluation of teaching.
- *Is there one specific format for presenting the data?* No! It is up to the candidate to decide on the best format for presenting the data.
- *Can you give some examples of statistical summaries?* Tables 1 and 2 are both examples of statistical summaries. Table 1 presents statistical summaries by course. Table 2 presents the data in chronological order.
- *Which method is better?* There is no single correct answer. If someone initially received a letter or letters from a P & T Committee that her scores were low she might choose to present the summary in chronological order to show improvement. If you teach a required GE course and your scores are lower in this class then in classes (elective or required) for undergraduate majors or graduate courses you may wish to present the summary by course.

Table 1
Statistical Summary of Student Evaluations by Course

Course	ES 101	ES 356	ES 425	ES 745
Number of Sections	n=18	n=6	n=6	n=8
Instructor Available for Consultation	4.2	4.5	3.98	4.75
Student Responsibilities Well Defined	4.35	4.65	3.54	4.25
Class Time was Well Spent	4.7	3.95	4.15	4.12
Materials Contributed to My Learning	3.5	3.98	3.89	4.23
I was Challenged in this Course	4.65	4.02	4.25	4.89
Coming In, I was Motivated to Learn	2.36	3.52	4	4.15

Table 2
Statistical Summary of Student Evaluations by Year

Year	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010
Number of Sections	n=8	n=7	n=8	n=7	n=8
Instructor Available for Consultation	3.25	4.5	4.35	4.21	4.6
Student Responsibilities Well Defined	2.98	3.25	3.98	4.5	4.4
Class Time was Well Spent	2.89	2.98	3.65	3.97	4.5
Materials Contributed to My Learning	3.5	3.25	4.8	4.6	4.75
I was Challenged in this Course	4.25	3.98	3.65	4.15	4.35
Coming In, I was Motivated to Learn	3.2	3.6	3.15	3.89	4.1

- *Can you present your scores both ways? Yes!*
- *Do you have to present your scores both ways? No!*
- *Are there other ways of presenting your scores? Yes! The two methods illustrated above are just examples.*
- *OK, I understand the statistical summary, but what does the CBA mean by frequency distribution? A frequency distribution gives the percentage of students who responded to a question by indicating that they strongly agree, agree, are neutral, disagree or strongly disagree. Below is an example of a frequency distribution. Frequency distributions can be done for each class taught; i.e., if you have taught four classes (ES 101, ES 356, ES 425 and ES 745) you would have four frequency distributions. Likewise frequency distributions could be done chronologically i.e., a frequency distribution for each year. Alternatively you could have just one frequency distribution for all of your classes combined. Again, it is up to the candidate to decide on how to present this information.*

Table 3
Frequency Distribution & Mean for ES 101

	SA	A	N	D	SD	Mean
Number of Sections = 18	5	4	3	2	1	
Instructor Available for Consultation	65%	20%	0%	0%	15%	4.2
Student Responsibilities Well Defined	55%	20%	15%	10%	0%	4.2
Class Time was Well Spent	20%	80%	0%	0%	0%	4.2
I Learned a Lot from the Instructor	80%	0%	0%	0%	20%	4.2
Materials Contributed to My Learning	74%	0%	0%	26%	0%	4.22
I was Challenged in this Course	30%	60%	10%	0%	0%	4.2
Coming In, I was Motivated to Learn	40%	40%	20%	0%	0%	4.2

- *Does the frequency distribution have to show the mean?* No, but this example shows you how you can combine the statistical summary i.e., the means with the frequency distribution in one table. If you have taught four courses during your probationary period and you chose to present the data in this fashion you would have four tables one for each course (a table for ES 101, ES 356, ES 425 and ES 745).
- *Why does the CBA require a frequency distribution?* The administration wanted Bargaining Unit Faculty to present the average score for each question. However, most statisticians would argue that the average of the scores on a scale of 1 to 5 is not a valid statistical measure if the purpose is to compare one person's score with another person's score. The frequency distribution provides a lot of information that is not captured by looking at the mean. Notice in Table 3 that all of the means are virtually the same but the frequency distributions are quite different. In some questions most students are in the middle with few at each extreme. In other questions student are polarized creating a bi-modal distribution. This is why the CBA states that low averages are not necessarily indicative of bad teaching and high numbers are not necessarily indicative of good teaching.
- *Why doesn't it make sense to compare average scores?* There are four types of measurement scales. The first is a nominal scale in which data are simply categorized by assigning them a number. For example women are 1 and men are 2. Taking the average of this number is meaningless. The second type of scale is an ordinal scale. Ordinal scales allow for rankings but differences

between numbers are not important. A Likert scale measuring satisfaction on a scale of 1 to 5 is an ordinal scale. When a student says they strongly agree we assign a 5, when a student says they agree we assign a 4 and when a student is neutral we assign a 3. The difference between a 5 and 4 is not necessarily the same as the difference between a 4 and a 3. The numbers merely reflect and ordering but not an intensity of preference. Just like it would not make sense to take the average of males and females it does not make sense to take the average of strongly agree, agree, neutral, disagree and strongly disagree. In order for an average to be meaningful data must be measurable on an interval scale or a ratio scale. An interval scale means that data is ordered, has a constant scale but no natural zero. An example is temperature. The difference between 30° and 20° is the same as the difference between 20° and 10°. But 20° is not twice as hot as 10°. Another example of interval data is years. The difference between the year 2000 and 1950 is 50 years and that is the same as the difference between 1950 and 1900. But it does not make sense to say that the year 2000 is two times the year 1000. In order to make such a comparison, data must be measured on a ratio scale. A ratio scale is ordered with a constant scale and has a natural zero. Someone who is 40 years old is twice as old as someone who is 20 years old. However it is simply invalid to conclude that someone who has a 4 on his or her teaching evaluation is twice as good a teacher as someone who has a 2 on his or her evaluation. This measurement problem is similar to knowing the order in which people finish a race but not knowing their exact times. The person who comes in second may be only a fraction of a second behind the person who finishes first, but the third place finisher may be five minutes behind the second place finisher. The rankings are 1, 2, and 3, but just knowing the rankings does not allow us to make a judgment about whether the race was a close race.

- *So why do we even collect numbers on teaching evaluations if we know they are not meaningful?* In negotiating our first contract, AAUP-WSU took the position that we did not want any numbers at all because they were meaningless and had been abused in the past. The administration wanted numbers because they provide an easy way to evaluate Bargaining Unit Faculty. The CBA represents a compromise. That is why the CBA says that “student evaluations are important indicators of teaching effectiveness, but numerical scores from these evaluations alone neither confirm nor deny an individual’s effectiveness. It is for this reason that the CBA requires that chairs take into account other factors besides numerical scores in evaluating a Bargaining Unit Member’s teaching effectiveness.
- *What are some other measures of teaching effectiveness?* One important measure of teaching effectiveness is peer evaluation of teaching. Peer evaluation of teaching is required for all untenured Bargaining Unit Faculty and the Chair in his or her annual evaluation must include a statement reflecting peer evaluation of an individual’s teaching effectiveness.

- *What are peer evaluations?* Peer evaluations are evaluations performed by other Bargaining Unit Faculty. The specifics of what constitutes a peer evaluation are contained in department bylaws (in the case of the College of Nursing and Health and the Lake Campus the specifics are contained in college bylaws).
- *What happens if the mechanism for performing peer evaluations in department bylaws is not carried out in the manner specified in bylaws?* If your colleagues do not perform a peer evaluation this cannot be held against you in the promotion and tenure process. However, the annual evaluation by your chair is supposed to take into account peer evaluation of teaching. Obviously if your peers do not perform these evaluations, the chair cannot take them into account when your annual evaluation. This is a violation of the CBA and you should talk to your chair about this problem, and if it is not resolved you should file a grievance. If you file a grievance right away, as opposed to waiting until you are about to be considered for tenure, the University and the AAUP-WSU can resolve the grievance by making sure that peer evaluations are performed in accordance with the mechanism specified in department bylaws.
- *Is an evaluation of teaching by the Center for Teaching and Learning a peer evaluation?* No, but candidates for promotion and tenure may include such evidence in their files to help document teaching effectiveness.