

---

## **AC 2011-631: PRACTICAL INTERPRETATION OF STUDENT EVALUATIONS FOR STARTING PROFESSORS**

### **Walter W Schilling, Milwaukee School of Engineering**

Walter Schilling is an assistant professor in the Software Engineering program at the Milwaukee School of Engineering in Milwaukee, WI. He received his BSEE from Ohio Northern University and his MSES and PhD from the University of Toledo. He worked in the automotive industry as an embedded software engineer for several years prior to returning for doctoral work. He has spent time at NASA Glenn Research Center in Cleveland, OH and consulted for multiple embedded systems companies in the Midwest. In addition to one US Patent, Schilling has numerous publications in refereed international conferences and other journals. He received the Ohio Space Grant Consortium Doctoral Fellowship, and has received awards from the IEEE Southeastern Michigan and IEEE Toledo Sections. He is a member of IEEE, IEEE Computer Society, and ASEE. At MSOE, he coordinates courses in Software Quality Assurance, Software Verification, Software Engineering Practices, Real Time Systems, and Operating Systems, as well as teaching Embedded Systems Software and other software and computer engineering courses.

### **John K. Estell, Ohio Northern University**

John K. Estell is a Professor of Computer Engineering and Computer Science at Ohio Northern University. From 2001 to 2010 he served as Chair of the Electrical & Computer Engineering and Computer Science Department at Ohio Northern. He received his doctorate from the University of Illinois at Urbana-Champaign. His areas of research include simplifying the outcomes assessment process, first-year engineering instruction, and the pedagogical aspects of writing computer games. Dr. Estell is a Senior Member of IEEE, and a member of ACM, ASEE, Tau Beta Pi, Eta Kappa Nu, and Upsilon Pi Epsilon.

### **Frederick Clayton Berry, Milwaukee School of Engineering**

# Practical Interpretation of Student Evaluations for Starting Professors

## 1 Abstract

In nearly all university environments, students evaluate their instructors. However, how these evaluations are used varies greatly from institution to institution and from position to position. Compounding this fact is the different expectations of students at different campuses.

As students and graduate students, our goal for a class was nearly always to obtain an A grade, no matter how challenging the course. However, with student evaluations, consistently obtaining perfect evaluations is nearly impossible, even for the most veteran instructor. This article provides practical ideas for interpreting student evaluations. It includes a summary of how evaluations are used at teaching oriented universities at both a departmental and college level. It then provides a set of tips for interpreting student evaluations, and provides techniques for setting improvement goals.

## 2 Introduction

As a new faculty member, one of the scariest aspects of teaching may be receiving the first set of student evaluations. Just as a student wonders what their grade will be for a course, faculty members wonder what their evaluations will show. However, unlike a student, who has a syllabus to go by and many grades recorded throughout the course of a quarter, student evaluations typically only occur after the course is completed. And, unlike student grades, course evaluations are made anonymously, making it more difficult to follow through and assess the status of one's performance.

At teaching oriented institutions, high quality teaching performance is often the number one aspect used for faculty tenure and promotion. This creates a problem for new faculty members. As has been stated by Brent et al<sup>1</sup>, faculty members transitioning to a teaching role receive very little preparation in their graduate study. Doctoral students are given guidance on how to perform research, but not how to effectively convey material to undergraduate students. As Boyce<sup>2</sup> has stated, it often takes four to five years for a faculty member to be an effective teacher. This may be too long for a person to "learn" on the job. Thus, it is important that this timeframe be shortened. One method of doing this is through careful study and analysis of student evaluations coupled with appropriate professional development and mentoring.

### **3 Student Expectations**

Student expectations vary greatly from campus to campus, and their respective teaching assessments may also vary. At large institutions, it is often common for students to have little personal interaction with faculty members. Questions of course content are handled by teaching assistants either during office hours or in recitation sessions. A faculty member who has only been exposed to this environment would then be very out of place at a school which does not have teaching assistants and has small class sizes.

Student expectations also vary between classes. Freshmen coming out of high school are exposed to many different levels of expectations. At one extreme are students who are over achievers and have been very strongly driven throughout their academic career. These students have either taken an AP course or possibly taken community college courses as part of their high school program. These students need guidance only in how to channel their energy to be successful. At the other extreme, however, are first generation college students. These students may need stronger levels of mentorship to be successful. Yet both of these sets of students are present in the same freshman course, and both of these students will be assessing the performance of course instructors. They each will have a level of expectation set for the faculty member. At the other extreme are seniors. While each one started as a freshman, they now have four (or more) years at the institution, and they have developed a level of expectation for each faculty member based on the performance and expectations of other faculty members. Obviously the needs of the seniors are going to be different than the needs of freshman. However, in most cases, the institutional instructor evaluation that is given to freshman will be identical to the one given to seniors.

Student expectations also vary across programs. Each academic department or program at an institution develops its own personality. A program may have high expectations of students to solve any and all problems on their own, only coming in for help with the most advanced of issues. Another program, however, may have a different attitude, hoping and expecting that students will come in sooner so that simple problems can be dealt with and the main educational objective of the assignment can be achieved rather than hours of work being spent resolving an unrelated technical glitch. These are radically different philosophies, and a professor who teaches a service course to students in a different program may have very poor results based upon the different expectations.

### **4 Institutional Usage of Evaluations**

While each institution has different policies and programs, at teaching institutions, instructor evaluations tend to be very important artifacts for faculty retention and advancement. This, of course, is derived from the institutions main objective, which is ensuring that quality instruction is always occurring. That being said, there is tremendously wide variation amongst institutions (and even within institutions) in their usage of evaluations.

The important thing to understand about student evaluations is that the effectiveness of an instructor cannot be translated into a single number which relates teaching performance with student ability. Research has consistently shown that student evaluations can be subject to small degrees of bias. Fischer<sup>3</sup> reports that biases in student evaluations may be present based on student grades, with professors who give higher grades receiving higher evaluations, based on instructor expressiveness, based on instructor's attractiveness, and other factors. Gender has also been shown to bias student evaluations<sup>4,5</sup>. These issues, while out of the control of the faculty member, are important to know when analyzing student evaluations.

With these issues present within student evaluations, some general guidelines for the usage of evaluations have been made. Fich<sup>6</sup> provides some recommendations for the usage of student evaluations. Amongst other things, she highly recommends that comparisons only be made across similar courses. Subjective items should also be avoided, such as the preverbal questions "what was the overall effectiveness of the instructor?" One should be cautious in analyzing results which include new courses or courses taught for the first time by a faculty member. Lastly, Fich indicates that the results of student evaluations have low precision, and thus any interpretations based on small sets of evaluations will also have low precision.

In assessing one's performance, the general makeup of the student body must be considered. It is well known that students taking a course outside of their major tend to rank the faculty member lower than a course which is directly required within one's major.

At the lowest level, evaluations can be voluntarily used by a faculty member and their mentor. A faculty mentor can take a look at an instructor's evaluations and suggest areas for improvement and growth. This, however, is most likely informal in nature, as instructor evaluations fall under a vast degree of confidentiality issues. However, for a new faculty member, this may be one of the most important uses for evaluations.

Beyond basic mentoring, the first area where faculty evaluations are used is in faculty retention. At a teaching oriented school, a professor who constantly receives poor instructor evaluations most likely will not be employed for extended periods. This is especially important for faculty members who are not in a tenure track position.

The first person who reviews instructor evaluations is typically the department chair. In reviewing instructor evaluations, the department chair is typically looking for two things. At some point, there is a minimum quantitative score that is desired across all axes. On a five point scale, this minimum might be set to a value of 3.0. That does not mean that a score of below 3 would be entirely inappropriate, but rather, a score below 3.0 might require reflection by the faculty member.

However, a department chair may also be looking for scores that are excessively high. For example, on a five point scale, scores of 4.5 and above might not be indicative of effectiveness but of popularity. A professor who pushes students to the edge of their comfort zones may receive lower evaluations from weaker students, thus lowering their overall scores versus a professor who does not push students as much. Thus, depending on the philosophy of the chairman (or potentially dean), it may not be entirely appropriate to have evaluations which are

“too high”. Scores which are consistently high may need to be justified by a faculty member. Methods of justification might include an analysis of student performance on Concept Inventories or an analysis of student performance on standardized tests. Conversely, a chair who has access to results on concept inventories or other standardized tests may question a faculty member who has high student evaluations but does not generate strong performance in the areas of their courses.

The department chair may also use instructor evaluations in making course assignments. In a curriculum, there may be certain key courses which are vital to achieve student success in the overall program. This is especially true of freshman courses, for freshman courses build a foundation that will be applied over the following four years. It is well noted that part time instructors and other adjunct instructors can have a negative influence on student retention<sup>7,8</sup>. Typically this comes from students perceptions of poor professor availability and other factors. Department chairs concerned about retention may use faculty evaluations to ensure that the right instructors are being assigned freshman courses. Thus, given two equally qualified faculty members, the “better teacher” may be placed in the early course. The “better teacher” may be defined by student evaluations coupled with other supplemental assessments.

Instructor evaluations are certainly used as a portion of the tenure process, which varies across institutions. Evaluations may also be used as part of pre-tenure reviews, serving as a gateway to determine if a professor should continue seeking tenure or should not. As with everything else involved in the tenure process, it is important to ask questions as to the weight placed upon course evaluations, as well as the level of analysis expected.

## **5 Tips**

This article has focused on providing an overview of student evaluations, namely the differences in student expectations and an overview of how evaluations are used in an institutional setting. This section, however, begins to shift the focus, providing a list of ideas that new faculty members can use to improve their performance goals.

### ***5.1 Obtaining Baseline Assessment***

The most important set of student evaluations is the first set of evaluations, for this provides a baseline of an instructor’s initial performance. Effective teaching takes practice, and usually the first set of evaluations is not fully reflective of the complete potential of a faculty member. The key point of the first evaluation is to establish which areas have exemplary performance (relative to the other areas), which areas have acceptable performance, and which areas need significant improvement.

There are many ways which one can determine which areas need the most improvement, both quantitative and qualitatively. As most evaluations use some form of a Likert scale, it is very easy to determine the percentage of students in each class which either agree or strongly agree with each question on the survey. From this, one then establishes a baseline of what one considers effective teaching. For example, the goal may be to have 90% of students either agreeing or strongly agreeing with the quality measured. A simple sort can then be performed on the results. The areas which are above the target do not need any significant work, while those

that are significantly below the target need to be improved, and those which are farthest from the target need to be improved the most.

This method works very well for analyzing a single class of data, but may not be the most effective measure across multiple classes. For multiple classes, the individual class assessments need to be combined. One simple method is to take the average of the assessments in each class and generate an overall effectiveness score which reflects the average of the three courses. For example, an instructor with an 80%, 92%, and 90% approval rating in three courses would achieve an overall effectiveness rating of 87%. This method works well if the classes are similar in size. However, this method can result in skewed data. For example, if the numbers above had come from class sizes of 5, 25, and 50 respectively, it is very possible that a single outlier student biased the overall summary. To deal with this, a better method might be employed.

To start, we want to convert the Likert scale responses to a numerical value. A “Strongly Agree” response is provided a value of 5, an “Agree” is provided a value of 4, a neutral response is provided a value of 3, a “disagree” is provided a value of 2, and a “Strongly Disagree” is provided a value of 1. For each class, we can then calculate a numerical average in the range between 1 and 5 by adding the individual scores together. To aggregate this data across multiple section, we simply, employ a weighted average and sum these results across all classes.

Using such a method, a professor may set a goal of no result below a 3.75. As with the previous method, it is easy to perform a sort on the results and focus on the key areas which are obtaining the lowest scores. It is also possible with this method to calculate an overall assessment score as the aggregate of all sub scores. This gives an aggregate baseline as well as an idea of those areas which most need improvement.

The entire purpose for setting performance goals is to determine where you desire your performance to be at. While there are differing approaches, as a new faculty member, one of the strongest sources for performance goals should be your faculty mentor. Your mentor has been through the assessment process and has ideas about what are considered “acceptable” levels for quality teaching as well as “unacceptable” levels for teaching quality. Thus, your mentor is an excellent source for such information.

While the initial baseline is set based off of the first set of evaluations, it is important to reestablish one’s baseline. Reestablishing one’s baseline will occur whenever one changes institutions, departments, or transitions from an adjunct to a full time professor.

## **5.2 Trending**

Once the first set of student evaluations have been obtained, the process of trending begins. With trending, the goal is to see small but continuous improvement in the areas that were identified as needing improvement when performance goals were set based on the initial baseline assessment.

Acceptable trending tracks data across multiple dimensions. As a whole, we want to see what is happening to the aggregate score. It is our hope that it never goes down after the baseline is

established. The goal is that it slowly but continuously improves until a plateau is reached. An overall effectiveness score which does this is indicative of a professor who is improving.

Simply trending the aggregate score, while providing useful information, does not necessarily allow one to track against the performance goals outlines previously. As one tries to improve in a given area, one of three things may happen. The score may improve. That is indicative of success. The area may stay the same. That is indicative that the changes the professor made in that area were not entirely effective or were not appropriate for the problem. The area may also get worse. That may indicate that the solution developed was not effective. As trends develop, the trends can be assessed with ones faculty mentor, and it may be that different areas of improvement are identified. The results of trending may also influence the need for professional development or further training from qualified educational instructors.

Another thing that is useful from trending is to compare aggregate scores from semester to semester or quarter to quarter. It may be that trending data indicates that students rate your performance harsher in one quarter than another and this is done consistently. For example, spring quarter may yield the lowest results not due to teaching style or effectiveness but because students are simply ready for summer.

An area where trending is very useful is to compare one section of a course to another section of a course. Assuming that an instructor teaches the same course in two different years, trending data can truly be used to show how the instructor has improved. A course which is taught for the first time typically has problem areas that do not go smoothly. By comparing the first and second offering of a course, one can see if their teaching has improved or not. However, the real comparison is between the first and third or first and fourth offering of a course.

In analyzing trending data, it is important to keep in mind things which are out of control of the professor. Figure 1 shows sample trending data from a course offered three times. At first glance, it would seem as if there are significant problems with the professor's effectiveness during the 2010 offering. However, this view may be changed when one is informed that the first offering of the course was at 13:00 in the afternoon, the second offering was at 9:00 and the third offering was at 8:00.

Question	2008	2009	2010
The Professor seemed to be in command of the subject material	4.45	4.25	4.54
The professor clearly explained the course requirements	4.27	4.33	4.38
The professor used the class time effectively	4.36	4.25	4.17
The professor made the subject matter interesting	4.09	3.58	3.31
The professor showed enthusiasm for the course material	4.73	4.33	4.15
The professor was accessible	4.64	4.33	4.00
The professor demonstrated an interest in helping students to learn	4.73	4.25	4.38
The professor treated students fairly and respectfully	5.00	4.42	4.46
Assignments and exams were representative of the material covered in class	4.00	4.08	3.92
Assignments and exams were graded fairly	3.82	4.50	4.08
Assignments and exams were returned in a timely manner	4.36	4.25	3.92
The assignments were helpful in increasing my understanding of the course material	4.55	4.17	3.46
	4.42	4.23	4.07

**Figure 1 Sample Trending Data from Student Evaluations for a single course over three offerings. Areas to watch, based on performance objectives, are marked in yellow and areas of significant concern are noted in Red.**

### 5.3 Avoid Dwelling on Angry outliers

When assessing course evaluations, avoid dwelling on angry outlier comments. It is possible that there are some students who simply do not like your teaching style, the course, your expectations, or other aspects of your instructing. These students may be very vocal, at least in their written comments, but may not be able to provide substantive comments. For these students, completing the student evaluation proves the student some purpose, but most likely provides little value to assessing one's effectiveness. In fact, some of these comments can be downright spiteful. So long as these comments are from a very small number of students, and the comments are balanced by positive comments from other students, it is perfectly acceptable to ignore them.<sup>9</sup>

### 5.4 Taxonomize Comments

Thusfar, we have discussed quantitative data that can be obtained from student evaluations. While the quantitative data is important, more substantial insight can often be obtained from the written comments. Written comments allow students to explain why they gave a ranking to a given professor rather than just providing the ranking. For example, in the data shown in Figure 1, one student commented that the professor did "very well trying to make the material interesting, but really, it was 8:00 and we were half asleep."

Student written comments, while insightful, need to be categorized in order to effectively track them across courses and quarters. One way of doing this is developing a taxonomy. A taxonomy allows a professor to assign a given student comment to one or more categories within the taxonomy. The professor can then track this information from course to course and quarter to quarter. An example of such a taxonomy is shown in Figure 2. Note that while the student comments are extensive, the main item that can be gleaned from the comments is that the professor needs to work the labs from the student's perspective prior to lab and attempt to anticipate the students problems. Teaching more VHDL in class was also commonly noted, and the length of tests was also noted as an issue.



Now that this taxonomy has been developed, trends in written comments can be tracked across quarters, noting whether there is an improvement or degradation in performance. It should be noted that, while the taxonomy can be helpful, the quantitative data gleaned from the taxonomy may not be as accurate as quantitative data recorded in previous sections. Because written comments are free form, each student may only comment on their most significant issues. Thus, while in the example two of the students commented on test length, it may have been something that all students felt but did not feel the need to comment.

### 5.5 Calibrated Peer Reviews

Another tool which can be used to aid in the interpretation of student comments is the Calibrated Peer Review™ (CPR). CPR is an online-tool with four structured workspaces that perform in tandem to create a series of activities that reflect modern pedagogical strategies for using writing in the learning process. The learning materials that were developed for guiding students through an engineering design experience serves as the data for assessment.<sup>10</sup>

Table I gives the results of student evaluations of the course. This table shows that the student course satisfaction did improve Post CPR. The CPR process provided the faculty member with detailed suggestions for improvement and was able to apply those ideas to improving their performance.

**Table 1 Results from Eight Items on Student Evaluations over a Six-Year Period**

Questions Asked of Students	Pre CPR 97-98	Pre CPR 98-99	Pre CPR 99-00	Post CPR 04-05	Post CPR 05-06	Post CPR 06-07	Pre CPR Average	Post CPR Average	Percent Change
<b>4. The professor was well prepared for class.</b>	<b>3.74</b>	<b>3.50</b>	<b>3.32</b>	<b>3.75</b>	<b>3.68</b>	<b>4.36</b>	<b>3.52</b>	<b>3.93</b>	<b>11.65%</b>
<b>5. The professor used teaching methods which helped me learn.</b>	<b>2.93</b>	<b>2.80</b>	<b>3.14</b>	<b>3.25</b>	<b>3.32</b>	<b>3.67</b>	<b>2.96</b>	<b>3.41</b>	<b>15.45%</b>
6. The professor was available for help outside the classroom.	3.59	3.62	3.57	3.57	3.70	3.79	3.59	3.69	2.60%
7. The professor seemed genuinely interested in teaching this subject.	3.52	4.18	3.59	3.82	3.79	4.30	3.76	3.97	5.49%
<b>8. Please rate the professor's overall performance in this class.</b>	<b>3.22</b>	<b>3.08</b>	<b>3.39</b>	<b>3.39</b>	<b>3.57</b>	<b>3.97</b>	<b>3.23</b>	<b>3.64</b>	<b>12.80%</b>
Number of Students	23.00	60.00	36.00	28.00	37.00	33.00			

Comment	Test Length Too Long	Bad scheduling	Less Powerpoint usage	Teach more VHDL outside of class	Work the labs from the students perspective prior to lab.	Liked real world application of the problems.
Count	2	1	3	4	5	2
It seemed to me like every quiz or test got longer with the steps and things that needed to be done. This then took away from time that could have been used to either teach at a slower rate or go over things a little more thoroughly. Also I think that the material covered at the beginning of the quarter could have been covered a little quicker, due to the fact most if not all of us just took discrete the quarter before, and by doing so it would have made more time for other material. Other than that, I saw no problem to the way things were taught. For this being the first time you taught this class, I think you did an excellent job.	1					
This was the professors first time teaching the course and I believe he did an excellent job.						
Overall the class was good. I think in the future maybe less powerpoint slides, especially when it's early in the morning. I feel I learn better when I have to be writing notes. Also lab seemed to be heavy on VHDL while in class we really didn't do much with it. I realize this wasn't the instructor's fault but I really didn't like how this class was 2 hours T/R at 8:00, I think I would have been able to hold my attention span with 4 classes instead of 2, or even have the 2 classes later in the day, either way it was hard to stay focused that early for that long, especially when some of the material became harder. I think overall the professor did a good job for this being his first class and I would probably take one of his classes again in the future.		1	1	1		
The course was fun to learn and relatively well taught. This was the first time teaching a course for the professor, and I think he did a well job. There were some problems every now and then with different concepts, but he was able to clear these up when we asked questions. Overall I enjoyed the course and thought the professor did a fine job.						
Perhaps teach more VHDL outside of the lab.				1		
Don't use power point, make students take notes off the board. They'll concentrate more on the material, instead of just being able to print off the notes.			1			
The class in and of itself is a good course. You obviously know your material and are very good at answering one-on-one questions, etc. There are two points I think that could be improved on in future courses. First, let your Power Point speak for you. The class will be a lot less monotonous for both you and your students if you show the same slides but generalize and show trends and important points instead of reading off every zero and one on the slide. Second, and perhaps this only applies to a course such as ECCS 261 where students are being introduced to a lot of new software, but try to anticipate the little quirks and bugs that aren't intuitive in the lab. Frequently I found myself with near-perfect comprehension of the concepts but I'd get caught up on some arcane feature of the software. If you get the chance in the future, running labs through the eyes of a student might help.			1		1	
For the first attempt at teaching this course, I would say it didn't go that poorly. Some things to improve upon: preparedness for the topic and being able to explain concisely, teach more VHDL in class, and the labs. I felt the labs were not very well written and often were very confusing in terms of how certain procedures should be done. Also, the lack of actually understanding VHDL made some of the labs unbearable. These things are sure to get better with more experience. One major thing I liked about the class was the real world application. This was one of the few times I could actually see where ideas were applicable, and that increased my interest in learning the subject.					1	1
I think the course could use some improvements but for a first time teaching it, it wasn't bad. The lectures themselves were usually put together well and provided good information at a reasonable pace. Sometimes a question would come up and the majority of the class time was used trying to answer the question, but that's happened in other classes also. The tests and quizzes were a little longer than the professor thought they would take but he let us use the time we needed. The labs are where the most improvement could be made. Often the instructions were not specific enough and were written as though the student had a fundamental knowledge of VHDL. More specific directions would have been appreciated. Also, sometimes the labs didn't work and no one knew how to fix them so a lot of lab time was wasted trying to figure out one problem. Overall, the class was a good one for it being taught for the first time and most of the problems will probably be taken care of in the future.	1					1
The professor will greatly improve and succeed with more experience.						
Some of the lab organization was a little disarrayed, a lot of that had to do with it being the first time as an instructor and the first run of the lab as well, as the second section typically went more smoothly than the first. I found almost all topics covered especially real world examples to be extremely interesting and thought provoking. Professor, you did a great job of really getting me interested in the material and excited about the prospect of doing work like this in the field, something most of my previous classes have not yet done very well! Thank you for all the work that you put in, I'm sure it wasn't easy!					1	1
The homework was at times off course with the class. It was sometimes too far ahead, and problems arose because of this. But if it is properly aligned with the class it would help the understanding. Also, the vhdl could have been explained more thoroughly in class instead of in the labs. The labs seemed distant compared to what we were doing in class, and did not really help us understand the class work. Other than this, the class was good for the first year.				1	1	

Figure 2 Analysis of student comments. Comments are listed on the left and the taxonomy of comments are indicated in the right columns.

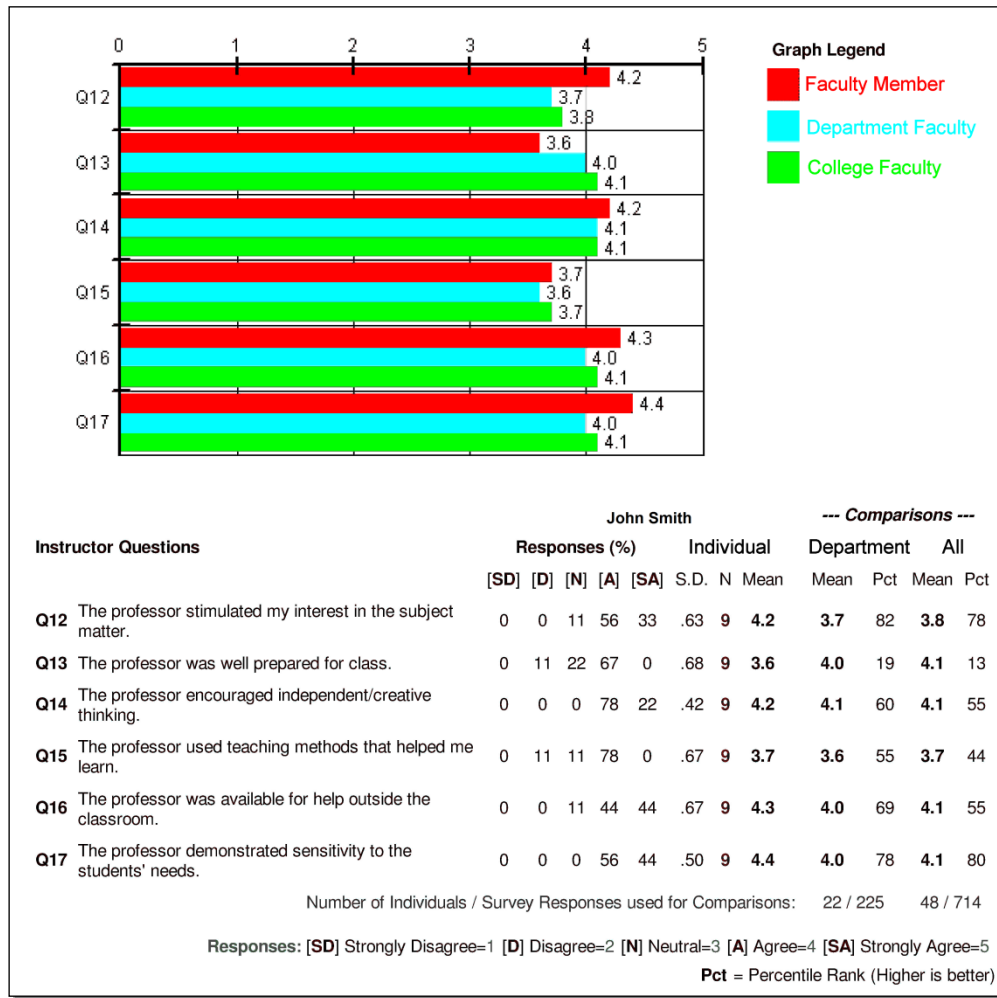


Figure 3 Sample report benchmarking faculty performance with department and college faculty.

### 5.6 Benchmark against other faculty members

Performance benchmarking is vitally important in order to be able to assess how one’s performance compares with other faculty members. Benchmarking summaries typically provide a summary of a professor’s performance relative to an anonymous set of institutional peers. For example, an online system may provide a faculty member with aggregate data for others teaching in their department, college, and at the institution itself across each dimension assessed on the performance evaluation. A sample benchmark summary is shown in Figure 3.

Having benchmarking data helps tremendously in interpreting one’s results. For example, if there is an area where, in general, the department has difficulty connecting with students, the aggregate average may be lower or higher.

Benchmarking data also helps tremendously in setting performance goals. A new faculty member who is already performing on par with their peers may not feel that they have a need to improve. That feeling, however, may be misplaced, as the aggregate data includes many different professors, both tenured and untenured, who have joined the institution under different situations. It may be that just performing “averagely” is not good enough for a new faculty member to be successful.

Conversely, it is also possible that benchmarking data may reflect that a target set is generally unachievable given the student body composition. For example, if the student body feels that assignments and exams were returned in a timely fashion rates a neutral response, and the average return rate is 2 class periods, it may not be possible for a professor to hit a target of 4.5 in this area given the student expectations.

Benchmarks provide yet another area of discussion for a faculty member and a faculty mentor. The faculty mentor, having experience at differing levels and also having access to the benchmarking data, can suggest adjustments to ones performance goals based on the mentor’s results, the faculty member’s results, and the benchmarking data.

Benchmarking also provides another critical resource for the new faculty member. By having faculty benchmarks available, especially for the current quarter, it is possible for the faculty member to detect trends in the overall evaluation of faculty members. For example, a faculty member may note that their performance decreases from fall quarter to winter quarter to spring quarter. This may be established through multiple years. By the same token, with benchmarking data, the faculty member can detect if this is a general trend present in all student evaluations throughout the year.

### **5.7 *Independent Surveys***

While the official, standardized survey is the survey endorsed by the university, this does not preclude the development of your own surveys of students. One technique that can be beneficial is to have the students write an end of term essay soliciting what they liked and disliked most about the course as well as soliciting feedback for improvement in the course. This may allow the students to provide more indepth feedback than the quantitative feedback from a standardized assessment. This written feedback may be useful in interpreting the numerical data as well as providing a mechanism for making quick improvements in a course.

Independent surveys also may provide relevant feedback that can be used to assess if changes in your teaching style have been effective. For example, a professor who has always used power point for his/her lectures may decide to switch to using the whiteboard in order to be more interactive with students. This may improve student attention, but also may have the disadvantage of not having their notes available electronically. An independent survey may help the instructor to determine if this was an effective change to make. The standard university survey may be incapable of assessing the effectiveness of this change.

An independent survey is also very important for a new or significantly revamped course. A new course or significantly revamped course offers many opportunities for bias, and ideally will

be discounted in overall administrative assessment. However, an independent survey may help to provide extra points which can be analyzed in a narrative on a given course.

## **6 Managing Performance Evaluations**

The management of performance metrics is not a one step activity, but rather, it is a recurring process which must be executed repeatedly to ensure success. Developing the discipline to follow through on the process ensures that one obtains the maximum benefit from the student evaluations.

The process can generally be broken into two pieces. The first is the immediate triage and assessment of the previous quarter's results. The second portion is the long term assessment of effectiveness.

The immediate analysis of results involves a cursory, preliminary analysis of the data. This should occur as soon as possible once the results are released by the surveying entity.<sup>1</sup> Ideally, the results are available to the faculty member between terms. In this stage, the faculty member calculates numeric averages and compares the results with their performance targets, noting any significant differences. The faculty member also does a preliminary read through of all written comments. If time permits, the comments can be classified using the taxonomy and sorted accordingly. The goal of this immediate analysis is to detect major problems that occurred in order that adjustments can be made for the following quarter.<sup>2</sup>

Quick analysis of results also provides a professor with an ability to determine if other in class surveys are going to be used during the following quarter. If an area is noted as a problem area based on the previous quarters surveys, it is possible that the professor can perform targeted monitoring of that area through other less formal approaches, such as minute papers.

The second part of the process involves a more detailed analysis of student evaluations. This process typically would occur over the summer or at another period in time when a more significant analysis can be conducted. This activity involves completing the taxonomy of written comments and analyzing them for patterns, completing quarter by quarter comparative analysis, and completed course comparisons. An outcome of this effort will be an updated professional development plan showing which areas of teaching may need further improvement, an assessment of how well the previous goals were met, and a new set of goals for the following year. This detailed analysis can then be a topic of discussion with your faculty mentor.

While the second part of the process is more in depth, this process may also be viewed as work completed towards tenure and review. Tenure dossiers, especially at smaller teaching oriented

---

<sup>1</sup> Most institutions withhold the release of student evaluations until after grades have been submitted to protect the grading process. However, once grades have been turned in, it is in the faculty members best interest to have the results of their evaluations available immediately thereafter.

<sup>2</sup> Because of the desire for rapid turnout to integrate results into the following quarter, it is imperative that the institution ensure that student evaluations are made available to students in a timely fashion. Just as students need continuous feedback to improve in a course, a faculty member can be more effective if feedback through student evaluations occurs in a timely fashion.

schools, require a narrative on teaching and teaching effectiveness.<sup>11</sup> By doing this analysis sooner and drafting the document in pieces, the overall tenure preparation may be reduced.

In summary, there are many tools that faculty can use to improve student learning and their teaching effectiveness. It is important to find the appropriate tool for the class and the faculty member at the given institution.

## 7 Bibliography

<sup>1</sup>Preparing New Faculty Members to Be Successful: A No Brainer and Yet a Radical Concept. ASEE Annual Conference 2006.

<sup>2</sup>R. Boice, *Advice for New Faculty Members*, Needham Heights, MA: Allyn & Bacon, 2000.

<sup>3</sup>Judith D. Fischer. Implications of Recent Research on Student Evaluations of Teaching, *The Montana Professor* 17.1, Fall 2006

<sup>4</sup>Ellyn Kaschak, Sex Bias in Student Evaluations of College Professors, *Psychology of Women Quarterly*, 2(3), 1978, pp. 235-242.

<sup>5</sup>Lisa Sinclair and Ziva Kunda, Motivated Stereotyping of Women: She's Fine if She Praised Me but Incompetent if She Criticized Me, *Personality and Social Psychology Bulletin*, 25(11), 2000, pp. 1329-1342.

<sup>6</sup>Fich, Faith E. Are Student Evaluations of Teaching Fair? *Computing Research News*, Vol. 15/No. 3, pp. 2, 10.

<sup>7</sup>Retention and Contingent Faculty use in a State System of Public Higher Education

<sup>8</sup>Jaeger, Audrey J.; Hinz, Derik The Effects of Part-Time Faculty on First Semester Freshmen Retention: A Predictive Model Using Logistic Regression

<sup>9</sup>Steven Volk. Reading Student Evaluations. Center for Teaching Innovation and Excellence online publication, 2010.

<sup>10</sup>Carlson, P.A.; Berry, F.C.; , "Calibrated peer review<sup>TM</sup> and assessing learning outcomes," *Frontiers in Education*, 2003. FIE 2003. 33rd Annual , vol.2, no., pp. F3E\_1- F3E\_6, 5-8 Nov. 2003 doi: 10.1109/FIE.2003.1264740

<sup>11</sup>Micah Hale, Findlay Edwards, and Norman Dennis. Preparing Your Tenure Dossier from Day One. Proc. of 2006 ASEE Annual Conference, Chicago, IL, 2006.