Student Feedback Survey

GUIDELINES FOR INTERPRETATION OF THE RESULTS

These guidelines were prepared to assist instructors in utilizing and interpreting the results of student feedback survey. They were not meant to limit instructors from using their own ideas and applications.

I. ADMINISTRATION'S VIEWPOINT

The following is an excerpt from Provost John Morton's November 1995 memorandum to the Faculty Senate concerning the evaluation of teaching.

"**1. There are no magic thresholds that distinguish a 'good' teacher from a 'bad' teacher.**
I do not have, could not have, and will not have simple rules that say that a particular score, say 3.0 out of 4.0 on a student evaluation question, is a minimum criterion, or that a course completion ratio of 70% is the lowest acceptable value. There are too many variables related to student motivation, nature of the class, time of day, etc. to allow such a simplistic approach to evaluation.

This does not mean that I ignore the actual evaluation results. Rather, if an individual consistently shows up at either end of the range of results (high or low) and does so consistently over time, then discussion or explanation of why should be provided. Some of the reasons may well be outside the faculty member's control; some may well be improved by providing mentoring or new strategies; and some may be successes that can be passed on to others. What causes me concern in my assessment of faculty is a consistent showing at the low end of the measures without any attempt to understand, explain, or improve".

"**2. There should be multiple measures of different facets of teaching performance.**
One of the problems we have in our present system is that we treat the measurement tools, such as student or peer evaluation, as the focus of our discussion or debate. I believe that is wrong. Instead, we should be focusing on what we believe to be the key components of being a good teacher at Kapi'olani and then using multiple measures to examine performance. This means that for some factors of teaching, such as being accessible to students, student evaluations results on items related to accessibility would be an important (but not sole) component of the measure, while for other factors such as knowledge of subject matter, peer evaluations, academic credentials, professional development activities, etc. would be components of the measure but student evaluations would not be. The important ideas is that the discussion should focus on what factors we believe to constitute a good teacher and how can we provide information on performance related to those factors".

"**3. We should not be discussing individual items on the student evaluation instrument.**
In the same way that we need multiple measures from multiple sources to examine critical factors of teaching, we need to look at the individual student evaluation questions in terms of their summary contribution to a factor. So instead of discussing the results of question #2, we should be looking at the results of a composite score of perhaps three or four questions that contribute to a measure of one of the factors. Instead of there being a discussion of 20 questions on the evaluation form, there should only be reported results of the four or five factors that the entire instrument is purported to assess."

II. THE EVALUATION FORM

The evaluation form is made up of 17 items on the front side of the page and 5 open-ended questions on the back side. Two of the front-page items are negatively phrased (items 2 and 14). These two items have their exact equivalents which are positively phrased, number 13 and 3. The reason for this arrangement and its use are discussed below.

Instructors can add an open-ended question in the space under number "6" provided on the back of the evaluation form.

III. INTERPRETING THE RESULTS

Only results of the 17 items on the front page have been computed.

1. Weights

Answers to each of the first 17 items, including the negatively-phrased ones, have been weighted as follows:

Answer	***Weight***
Strongly Disagree	1
Disagree	2
Agree	3
Strongly Agree	4
No opinion / Don't know	--

2. Total Responses

The number of responses to each item are shown on the printout. Please keep the following points in mind:

a. Students who filled out the course code incorrectly may not be properly grouped. Instructors are urged to double-check the count of student responses they have received.

b. Some marks that are not dark enough may not have been picked up in the electronic scanning process, even though they may appear dark.

3. Total Scores

The computer printout shows a line preceded by the word "Comp" for *Composite*. This line provides the statistics for the combined 15 non-negative items, excluding the negatively-phrased items 2 and 14.

4. Factor Scores

Based on statistical factor analysis of the 1996-1997 results, the 15 non-negative items have been combined into 7 factors. These seven factors along with the items making up each (in parentheses) and their results are presented in the bottom of the results pages.

Statistical Indexes

1. The Mean

The mean provides the "typical" response to each item and to the composite. For the case of only one student responding, the mean was left blank.

2. The Standard Deviation

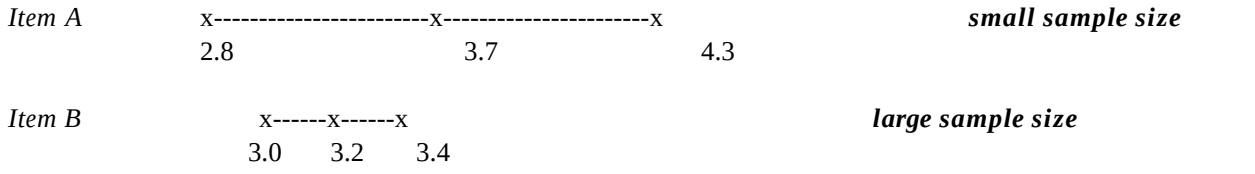
The standard deviation indicates how much variation exists in the student responses. So, a mean of 3.0 associated with a small standard deviation (e.g. 0.3) indicates that most students tended to choose "Agree". A mean of 3.0 with a large standard deviation (e.g. 1.3) on the other hand, indicates that most students chose "Disagree" or "Strongly Agree", with the average of these ratings being "Agree". The implications from these two results are different despite the equality of the two means. Large standard deviations suggest that students' opinions differ widely.

For the case of only one student responding, the standard deviation was left blank.

3. The Confidence Interval

Another way to assess the meaningfulness of results is to look at class size. Clearly, a mean score of 3.7 obtained from a class of 4 students does not have the same impact as the same mean from a class of 40 students. The larger the class the higher the confidence that can be associated with the data. This is the principle on which the confidence interval is built. Simply stated, the confidence interval for the mean is the interval within which the "true" mean of the responses falls with a certain probability. The **larger** the confidence interval the **smaller** the degree of confidence. By making judgments on the confidence interval rather than on that finite point representing the mean, conclusions that overlook sample size can be avoided.

For example, a mean of 3.7 on item A may be perceived as being higher than a mean of 3.2 on item B. However, if the former was obtained from a small sample, its confidence interval may be as large as 2.8 to 4.3; whereas the confidence interval for the mean of 3.2, based on a larger sample, may be only 3.0 to 3.4, as illustrated in the diagram below. The correct interpretation in comparing these two item means is that they are not significantly different.



For the case of only one student responding, the confidence interval was left blank.

Calculation of the Confidence Interval

The 95-percent confidence interval has been provided in this report. It is based on the Student t-statistic, with two tails and n-1 degrees of freedom, where n is the number of responses to the item. For the same class, the value of the t-statistic may vary from item to item, depending on how many students have responded to the item. In some small

sample-size cases, the confidence interval limits were beyond 0 and/or 4.00. These limits were artificially capped to 0 and 4.00, since responses to each item cannot exceed these two limits.

4. Consistency of Student Responses

The two pairs of oppositely-phrased items mentioned above, number 2 and 13, and 3 and 14, were used to identify those students who respond consistently to the evaluation form based on the *Consistency Scale*.

The Consistency Scale

Responses to the two oppositely-phrased item pairs, number 2 and 13, and 3 and 14 were used to classify students as follows:

For an item pair, responses were considered inconsistent if they were:

a. "Agree" and "Agree."
b. "Agree" and "Strongly Agree."
c. "Disagree" and "Disagree."
d. "Disagree" and "Strongly Disagree."

Students were not classified as inconsistent if they left one or both of the opposite items unanswered (or chose "No opinion/Don't know") or if they responded as follows to an opposite item pair:

a. "Agree" and "Disagree."
b. "Agree" and "Strongly Disagree."
c. "Disagree" and "Agree."
d. "Disagree" and "Strongly Agree."

Inconsistent responses to either one of the two item pairs were sufficient to classify a student as inconsistent.

Two reports are provided for your review on two separate sheets: one that includes all students and one that includes consistent students only.

IV. USE OF THE DATA

It is recommended that instructors use and compare their results for the same course over time. This type of comparison is in general more useful and accurate than that in which different instructors are compared. Also, instructors are free to use data based on all students, data based on consistent students only, or both.

V. OPEN-ENDED QUESTIONS

Results of open-ended questions often enhance the rigid statistical analysis. Instructors may choose to report specific comments or develop a method of summarizing the results of the open-ended responses. The summary could be a simple paragraph describing the overall trend in the comments or a more detailed content analysis that groups and reports quantity or quality of responses.

Revised January 1998