Technical Paper No. 1

"A Report on Student Evaluation of Faculty Teaching Performance at Sangamon State University."

by Professor Jerry A. Colliver
MEMORANDUM

TO: The University Community and its Observers
FROM: John H. Keiser, Vice President for Academic Affairs
DATE: September 26, 1972
SUBJECT: EVALUATION REPORT

Sangamon State University, from time to time, produces reports on some of its activities which are of unusual value. In these cases it is only appropriate that the work be given more than the routine status of a memorandum or a committee report.

It is the intention of this office and the Division of Academic Affairs, therefore, to make Professor Jerry Colliver's "A Report on Student Evaluation of Faculty Teaching Performance at Sangamon State University" the first of a series of technical papers on matters of importance to the Division of Academic Affairs and to the university itself. While this sponsorship does not necessarily imply total agreement with the details of these papers, it does indicate an endorsement of their quality and usefulness.
A REPORT ON STUDENT EVALUATION OF FACULTY TEACHING PERFORMANCE AT SANGAMON STATE UNIVERSITY

Jerry A. Collier

INTRODUCTION

An extensive survey of student evaluation of faculty teaching performance was conducted during the winter and spring quarters of the 1971-72 academic year at Sangamon State University. The results of the survey provided one source of information which was used in making salary and promotion decisions on a merit basis.

The present paper begins with a brief discussion of the history and rationale of the construction of the evaluation instrument and the development of the evaluation process. The results of research on the reliability and validity of the instrument as well as the correlations of the evaluations with other variables are then presented. Also, the implications of the findings are discussed. Finally, a complete summary of the paper is presented.
History and Rationale of the Construction of the Evaluation Instrument

The evaluation forms used during the winter and spring quarters originated from a parent questionnaire which was developed by a consultant during the winter quarter. The consultant had been associated with the University since June, 1970 and was familiar with the philosophy and goals of the University through numerous conversations with administrators and faculty and through extensive reading of the University's literature.

In December, 1971, a number of evaluation forms which have been used at other universities and colleges was made available to faculty and students. These two groups were invited to carefully examine the forms and to comment regarding the strong and weak points of each. They were also asked to make any recommendations regarding the construction of an evaluation instrument to be used at the University. About 35 faculty members and students responded to the request. In early January the material was sent to the consultant and in early February, the consultant visited the University bringing with him a 22-item questionnaire for student evaluation of faculty which had been constructed in light of the faculty and student response. The form required students to rate faculty along a 7-point scale in regard to the content of each item. A copy of the form is included in appendix A.

On February 14, 1972, a copy of the consultant's form was sent to all faculty members, and students were informed that copies of the form were available. Faculty and students were asked to examine the form and to make their reactions known to the Evaluation Committee. On the basis of these comments, minor revisions were made on the original form resulting in a 20-item form which required rating along a 5-point scale. A copy of the latter form is presented in appendix B.
The 20-item form was administered to all classes in the University at the end of the winter quarter with the intention of using the results from the questionnaires as a source of information to be used in making decisions regarding salary and promotion. The responses to virtually all questionnaires in all classes were read by two graduate students and myself. A total of over 2000 questionnaires were read. The student reaction to the 20-item form was generally negative; they felt that the form was too long and that many of the items were ambiguous. Fortunately, one of the revisions which was made of the original form by the Evaluation Committee was to include a space following each item on the form in which the student was asked to comment in writing to that particular item. This made it possible to determine which items were ambiguous and possibly misinterpreted.

It was observed in reading through the comments on the questionnaires that many students had misinterpreted certain items and/or that they had chosen to reinterpret certain items in light of their particular philosophy. It was also observed that on certain items half of the students in a given class rated the faculty member either very high or very low while the remaining students indicated that the item was not applicable to the course. For example, item number 17 on the winter quarter form asked "Were lab sessions valuable?" It was observed in many classes that about half the class responded "not applicable" while the other half responded with both high and low ratings. When questioned, instructors of these courses reported that there was no laboratory experience in the course. The written comments revealed that certain students had felt that a trip to a museum, a movie, or a class meeting at the instructor's home constituted a "lab session." Consequently, it was recommended that such items considered to be ambiguous be eliminated from the evaluation form.

It was also recommended that items be removed which appeared to dis-
4. crminate against faculty members who have a particular teaching style or teaching philosophy or who teach certain kinds of subject matter. For example, item number 7 in the winter form asked, "Do you feel that you have an understanding of what the course was about?" A number of faculty members criticized this item. They maintained that the item might be appropriate for a course in which a specific content or subject matter is covered, e.g. a course in "World War II" or "Matrix Algebra". However, they felt the item was inappropriate for an experiential or personal growth course in which students are challenged to think about their own personal value systems, the assumptions they use in making decisions, the manner in which they relate to other people, etc. The critics of this item argued that students in content oriented courses would respond more positively to the item than would students in experiential courses. This contention was generally supported by an examination of the written comments to this item. Thus, on the basis of conversations with faculty members concerning certain items and on the basis of the accompanying written comments, it was recommended that items which appeared to be discriminatory be removed.

After careful consideration, it was recommended that 16 items on the 20-item winter quarter evaluation form were either ambiguous or discriminatory and should be removed. However, since this would have reduced the form to only four items, it was recommended to retain four of the items which were felt to be less questionable from the pool of 16 questionable items. It was felt that possibly these four additional items had been knowingly re-interpreted rather than misinterpreted by the students. Since a comment space was provided with each item, students had the opportunity to write in their own personal interpretation of the item and then respond accordingly.
Consequently, a recommendation was made to eliminate the comment space that accompanied each item and to provide a general comment section at the end of the form. The Evaluation Committee unanimously approved all of these recommendations. This resulted in an 8-item evaluation form with an additional page for general comments which was administered in the spring quarter. A copy of the form is presented in appendix C. Due to administrative pressures, the form was distributed at the middle of the quarter rather than at the end of the quarter.

The responses to virtually all the questionnaires administered in the spring were read by the two graduate students and myself. Again it was felt that the four questionable items from the winter quarter were being misinterpreted and a recommendation was made to eliminate these items from the determination of the final evaluations. A special faculty committee composed of a member from the Council on Academic Affairs, the University administration, and the Evaluation Committee reviewed and approved the recommendation. The special committee further decided to eliminate a fifth item which had been considered discriminatory by many faculty members.

Consequently, the final evaluation was based on three items which appeared in both the winter and spring quarter evaluation forms. The three items are presented below:

(1) Do you think this teacher is competent in the content or matter offered in this course?

<table>
<thead>
<tr>
<th>exceptionally competent</th>
<th>satisfactory</th>
<th>incompetent</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>NA</td>
</tr>
</tbody>
</table>

(2) Overall, do you consider this person a good teacher?

<table>
<thead>
<tr>
<th>excellent</th>
<th>good</th>
<th>poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
(3) After careful consideration, do you think that this teacher should be:

_____ retained

_____ retained, but encouraged to improve in the following areas:

_____ not retained.

A summarization of the data obtained from the three items was prepared for each faculty member. It included:

(1) tables showing the relative frequency of different ratings,

(2) the mean (arithmetic average) and standard deviation (a measure of the variability of the ratings) for the competency and teaching items combined for each class,

(3) the grand mean\(^1\) (average of ratings across all classes) and the grand standard deviation for each faculty member.

An example of the data summarization information for a given faculty member is presented in table 1 on page 7.

**Salary and Promotion Committee**

This information for each faculty member was presented to the Salary and Promotion Committee which was composed of three representatives elected from the faculty-at-large, three academic deans, and the University librarian. The University librarian voted on library faculty only. The chairman of the Evaluation Committee was a non-voting advisory member of the committee.

The faculty information provided to the Salary and Promotion Committee was to be used by them as one source of data for making salary and promotion decisions. The members of the committee were instructed by the Evaluation Committee to:

\(^1\) The grand mean was a weighted mean because the computation took into account the number of responses upon which the separate class means were based.
Table 1

An example of the data summarization for a given faculty member.

<table>
<thead>
<tr>
<th>Course Number</th>
<th>Class Size</th>
<th>Mean 4.88CC</th>
<th>Number Clesticinaires</th>
<th>SD 0.3264</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>14</td>
<td>4.88CC</td>
<td>14</td>
<td>0.3264</td>
</tr>
<tr>
<td>23</td>
<td>10</td>
<td>5.00CC</td>
<td>0</td>
<td>0.00CC</td>
</tr>
<tr>
<td>179</td>
<td>24</td>
<td>4.8778</td>
<td>19</td>
<td>0.3655</td>
</tr>
<tr>
<td>176</td>
<td>19</td>
<td>4.5117</td>
<td>17</td>
<td>0.2236</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competence</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Course Number</th>
<th>Class Size</th>
<th>Mean 4.6888</th>
<th>Number Clesticinaires</th>
<th>SD 0.3655</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>14</td>
<td>4.6888</td>
<td>14</td>
<td>0.3655</td>
</tr>
<tr>
<td>23</td>
<td>10</td>
<td>5.00CC</td>
<td>0</td>
<td>0.00CC</td>
</tr>
<tr>
<td>179</td>
<td>24</td>
<td>4.8778</td>
<td>19</td>
<td>0.3655</td>
</tr>
<tr>
<td>176</td>
<td>19</td>
<td>4.5117</td>
<td>17</td>
<td>0.2236</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competence</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Committee not to look at only the grand mean of each faculty member in reaching their decisions. The committee was instructed to look at the separate class means for each faculty member. If there were large discrepancies between the class means, the committee was advised to read the written comments made by students in all classes taught by that faculty member. Similarly, if the grand standard deviation were large, the committee was advised to read the written comments made by students in all classes taught by that faculty member. In addition to reading the written comments in the situations mentioned above, the members of the Salary and Promotion Committee reported that they read the written comments of all the faculty with low grand means and of a few faculty with high grand means.

After taking into consideration written comments, class size, one extremely low or extremely high class mean, etc., the Salary and Promotion Committee generally assigned each faculty member to one of five rating categories on the basis of their grand mean. However, a few faculty members were assigned to a higher or a lower rating category than that indicated by the grand mean. These ratings, then, constituted the final determination of student evaluation of faculty teaching performance.

The Salary and Promotion Committee was also provided with a "service" rating for each faculty member. These ratings were made by the faculty member's dean and an elected peer and were a rating of service to the University and the community. The ratings were made along a scale from 3 to -1 with 1/2 unit steps. The Salary and Promotion Committee was also assigned the

---

1 To help them resist the temptation of doing so, I refused to provide them with a rank ordering of faculty members on the basis of the grand means.

2 The rating categories used were 3, 2, 1, 0, and -1. Faculty members with grand means from 4.70 to 5.00 were assigned to category 3; faculty members with grand means from 4.40 to 4.69 were assigned to category 2; etc.
task of reviewing these ratings. The final overall evaluation of a given faculty member was a weighted sum of the teaching rating and the service rating. The teaching rating was weighted 60% and the service rating was weighted 40%.
An extensive analysis of the evaluation data was performed. Much of the data analysis was focused on grand means and class means since, in general, these means were the major determinant of the final teaching rating for virtually all faculty members. The data analysis was conducted to determine the adequacies and inadequacies of using only a 2-item evaluation instrument. Ideally, the adequacy of the instrument should have been studied before it was used in making salary and promotion decisions. However, due to external pressures to develop a process for making salary and promotion decisions on a merit basis which emphasized student evaluations, the present system was put into effect. The data were collected in such a way that it was possible to answer many questions about the adequacy of the process.

Reliability

The reliability of a given measurement instrument refers to the amount of consistency or agreement among measurements obtained from repeated application of the instrument to the same person. The notion of reliability is closely related to the notion of error. The more error in a given measurement process, the less agreement among repeated measurements of the same person and, consequently, the less reliability the process is said to have. The reliability of a measurement process is commonly determined by measuring a group of persons on two separate occasions. If the scores obtained on the two occasions are quite similar for most persons, the process is said to be reliable; if the scores on the two occasions are unrelated, the process is said to be unreliable. The actual degree of reliability may be assessed
by means of a correlation coefficient.¹

The reliability of the evaluation process was determined by looking at the amount of agreement between the mean evaluations for different classes and at the amount of agreement between the mean evaluations for the winter and spring quarters.

The reliability of class means was determined as follows: since most faculty members taught two classes a quarter, one winter quarter class was arbitrarily chosen for each faculty member and called class 1. The second class chosen was called class 2. In the spring quarter one class was arbitrarily chosen and called class 3 and the second class chosen was called 4. It was then possible to correlate the class 1 mean evaluation with the class 2 evaluation, class 1 with class 3, etc. The results of intercorrelating the class means and the number of pairs of classes each correlation was based upon are presented below where the subscripts of r

¹ A correlation coefficient is an index of the degree of relationship between two variables. The coefficient, which is symbolized by r, may take on values from -1.0 to 0.0 to +1.0. If the coefficient is found to be 0.0 or very near 0.0, there is said to be no correlation between the two variables. A greater relationship between two variables is indicated to the extent that the coefficient departs from 0.0. A perfect relationship is indicated by either -1.0 or +1.0. The sign of the coefficient indicates the direction of the relationship: a negative sign means there is an inverse relationship between the two variables while a positive sign indicates a direct relationship. For example, consider the relationship between height and income. If the correlation were found to be $r = .45$, this indicates that as height increases, income also has a tendency to increase. If the correlation had been $r = -.45$, this indicates that as height increases, income tends to decrease. If the correlation had been $r = .70$, rather than $r = .45$, it shows that height and income have a tendency to increase together but that the tendency is stronger if $r = .70$ than if $r = .45$. 
indicate the two classes that were correlated:

\[ r_{1,2} = .55, \quad n = 55; \]
\[ r_{1,3} = .59, \quad n = 75; \]
\[ r_{1,4} = .65, \quad n = 49; \]
\[ r_{2,3} = .36, \quad n = 55; \]
\[ r_{2,4} = .53, \quad n = 42; \]
\[ r_{3,4} = .49, \quad n = 49. \]

All the correlations were significant at the .01 level. Correlations of this magnitude indicate that there was moderate agreement among evaluations of courses in the same and different quarters.

A mean evaluation for the winter quarter and mean evaluation for the spring quarter were computed for each faculty member. The correlation for these means was found to be \( r = .62 \). This significant correlation was based upon 75 cases.

Thus, it was found that the reliabilities of the class means were generally above .50 while the reliability of a quarter mean was .62. Notice that the reliability of the quarter mean was higher than that of the course means. This is because the quarter mean was based upon more information (and more information generally reduces the error in the measurement process). Actually, the reliability of interest is the reliability of the grand mean which was based upon all courses in both quarters. This reliability will be higher

1 Significance is a statistical concept that refers to the fact that the results of any research may be due to chance. Results are said to be significant if the probability is very small that they are due to chance. If it is likely that the results are due to chance, the results are said to be insignificant. The above results could have occurred less than 1 time out of 100 by chance.

2 Correlations of this magnitude are said to be moderate to high correlations (Guilford, 1956). That is, there was moderate to high agreement between evaluations faculty received in different classes and in different quarters.
than that of a course mean or a quarter mean because it is based upon even more information. This reliability coefficient cannot be computed because only one grand mean was available for each faculty member. However, a technique for projecting this reliability coefficient on the basis of the reliability of the quarter mean was used and the projected reliability of the grand mean was found to be .77. Finally, if this evaluation procedure were to be used in subsequent years, the grand means would be based upon evaluations from all courses from three quarters rather than two. The projected reliability of a grand mean based upon data from three quarters was found to be .83.

Standard Error of Measurement. Although, as was mentioned above, reliability and error are inversely related, it is sometimes illuminating to look at the amount of error in a given measurement process in addition to looking at the reliability of the process. A common indicator of the amount of error in a measurement process is the standard error of measurement. The standard error of measurement might be thought of as a "sort of average" of the errors in the measurements obtained from a given measurement process.

Due to the error in a given measurement process, any actual measurement obtained is thought to be only an approximation to or an estimate of the

---

1 The Spearman - Brown Technique was used to make these projections. It seems reasonable to assume that the assumptions underlying the appropriate use of this technique were met. See Magnusson (1966) for a discussion of the Spearman - Brown Technique.

2 The standard error of measurement is computed directly from the reliability coefficient as follows:

\[ \text{SE meas} = \text{SD} \sqrt{1 - r} \]

where SD is the standard deviation of the measurements and r is the reliability coefficient. It may be seen from the formula that as the reliability coefficient increases, the standard error of measurement decreases.
hypothetical "error-free" or "true" value that the measurement process is
designed to measure. The standard error of measurement is commonly used to
set up a confidence interval or margin of error such that there is reason-
able confidence that the interval includes the true value.

The standard error of measurement for the grand means was computed
using the projected reliability of the grand mean. The standard error was
found to be .13. Using this standard error, it can be said that the hypo-
thesical "true" grand mean for any faculty member will be included within an
interval defined by the faculty member's obtained grand mean plus or minus .13
about 68% of the time. Similarly, there is 95% confidence that the interval
defined by the obtained grand mean plus or minus .26 would include the true
grand mean.

When the grand means were initially computed, there was some concern that
they would not discriminate among the faculty because the range of the grand
means was small. Although it is probably partially correct to insist that the
range of measurements resulting from a given measurement process should be
large, this ignores the magnitude of the error in the process. It is more
accurate to say that the range should be large relative to the standard error. 2

---

1 For example, 68% of the time the interval defined by the obtained measure-
ment plus or minus one standard error of measurement will include the true
value. Similarly, 95% of the time the interval defined by the obtained measure-
ment plus or minus twice the standard error will include the true value. A
very conservative estimate would be that only 75% of the time the interval
defined by plus or minus twice the standard error will include the true value.
This estimate makes no assumption concerning the shape of the distribution
of obtained scores around the true score. However, the shape of the dis-
tribution is probably close enough to symmetrical that the confidence level
is very close to 95%.

2 Actually, the concern about the magnitude of the range probably stemmed from
a tacit fear that the error in the measurement process was large relative to
the range.
In order for the grand means to make discriminations among the faculty on the basis of real differences as opposed to discrimination based solely upon the error in the measurement process, it may be roughly stated that the standard error must be at least less than $1/4$ of the range. This is the same as saying that the $95\%$ confidence interval must be less than the range.

The range of the grand means was found to be 1.5 since the largest grand mean was 5.0 and the smallest was 3.5. Consequently, the standard error of measurement was less than $1/10$ of the range which, of course, is less than $1/4$. Similarly, since the $95\%$ confidence interval for any faculty member is the obtained grand mean plus or minus .26, the size of the confidence interval was .52, which is less than the range.

The preceding discussion of the adequacy of the discrimination between faculty members and the size of the range relative to the standard error is just another way of saying that the measurement process had moderate to high reliability.

Another concern that was frequently expressed when the grand means were initially computed was that all of the means fell above the midpoint of the rating scale. That is, students were asked to rate the faculty along a 5-point scale from 1 to 5 with a midpoint of 3 and all of the faculty grand means fell above 3. The literature on faculty evaluations, however, shows that this is to be expected. When rating faculty, students have a tendency to almost exclusively use the upper end of the rating scale (Hildebrand, 1972).

The Standard Error of the Mean. The standard error of measurement is a generalized indicator of the amount of error in all of the measurements obtained from a given measurement process. All of the measurements are taken into account in its computation and, consequently, the standard error of measurement is the same for every obtained measurement. In the present case
this means that the grand mean for every faculty member was taken into account in the computation of the standard error of measurement and that the standard error was the same for every faculty member.

Since the measurements obtained from the evaluation process were means (the grand means) rather than unitary scores (such as would have been obtained if height or IQ had been measured for each faculty member), it was possible to compute a separate standard error for each faculty member which was based only upon the data used to compute the grand mean for that faculty member. This standard error is called the standard error of the mean. The standard error of the mean for a given faculty member takes into account the number of students that responded to the evaluation forms. The more responses, the smaller the standard error of the mean. It also takes into account the amount of agreement or disagreement among the students in their evaluations. The more agreement, the smaller the standard error of the mean.

The standard error of the mean was computed for all 75 faculty members for whom grand means were computed. A frequency distribution of the standard errors obtained by the faculty are presented in table 2 on page 17. The cumulative frequency and the cumulative percentage columns in the table indicate the frequency and percentage, respectively, of the faculty members with standard errors smaller than or equal to the standard error in that row. It may be seen that over half the faculty had standard errors less than or equal to .09 and that nearly 90% of the faculty had standard errors less than or equal to .15. Since the standard error of measurement was equal to .13, the table indicates that the standard error of the mean was less than the standard error of measurement for over 75% of the faculty.

Using the standard error of the mean a confidence interval was set up for each faculty member in a manner virtually identical to that described above.
<table>
<thead>
<tr>
<th>Standard Error of the Mean</th>
<th>Frequency</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>.22 - .24</td>
<td>2</td>
<td>75</td>
<td>100%</td>
</tr>
<tr>
<td>.19 - .21</td>
<td>2</td>
<td>73</td>
<td>97%</td>
</tr>
<tr>
<td>.16 - .18</td>
<td>5</td>
<td>71</td>
<td>95%</td>
</tr>
<tr>
<td>.13 - .15</td>
<td>10</td>
<td>66</td>
<td>88%</td>
</tr>
<tr>
<td>.10 - .12</td>
<td>18</td>
<td>56</td>
<td>75%</td>
</tr>
<tr>
<td>.07 - .09</td>
<td>26</td>
<td>38</td>
<td>51%</td>
</tr>
<tr>
<td>.04 - .06</td>
<td>12</td>
<td>12</td>
<td>16%</td>
</tr>
<tr>
<td>.00 - .03</td>
<td>0</td>
<td>0</td>
<td>00%</td>
</tr>
</tbody>
</table>

Table 2

Frequency distribution for standard errors of the mean.
except that the standard error of the mean was used instead of the standard error of measurement. Since the standard error of the mean was less than the standard error of measurement for 75% of the faculty, obviously the confidence intervals based on the standard error of the mean were less than the corresponding confidence intervals based on the standard error of measurement for 75% of the faculty.

Since reliability and error are inversely related, this suggests that the reliability of the grand mean for the majority of faculty members was actually greater when error was determined for each faculty member separately than was previously indicated in the discussion of the standard error of measurement.

Validity

The validity of a measurement process refers to the extent to which the process measures that which it purports to measure. A common procedure for determining the validity of a measurement process is to relate the obtained measurements to some generally agreed upon criterion or definition of that which the process claims to measure. For example, a college aptitude test could be validated by relating it to a common criterion, college grades. The scores on the test could be correlated with college grades. If the correlation were fairly high, the test would be said to be a valid indicator of potential for success (as measured by grades) in college.

Unfortunately, it is not as easy to validate the evaluation process as it is to validate a college aptitude test due to the absence of a generally agreed upon criterion or definition of good teaching. In my opinion, it will be impossible to find a generally agreed upon criterion of good teaching.
There are probably as many definitions of good teaching as there are teachers. Consequently, in light of the reliability evidence, it appears that we are measuring "something" about faculty members and the way they conduct their classes; however, in the absence of a criterion, it isn't clear if that "something" is good teaching or what it is. It should be emphasized that this difficulty is not unique to the present evaluation process. In the absence of a generally agreed upon criterion of good teaching, it will be impossible to directly validate even the most sophisticated evaluation process.

Some data were obtained which, it was hoped, would have a bearing on the validity issue. The evidence consisted of correlating the grand means obtained from the present evaluation process with measures obtained from other procedures which have been used to evaluate teaching at the University or at other institutions. One of the procedures employed a more sophisticated, research based evaluation form while the other procedure involved a committee approach to the evaluation of teaching. Of course, it should be realized that neither of these procedures had been validated in the sense that they had been related to some generally agreed upon criterion of good teaching. However, if the three techniques were correlated, this type of corroborative evidence would suggest that there is some implicit agreement about the nature of good teaching and what kinds of things are related to it even though an explicit definition is not available.\(^1\)

**Davis Form.** The more sophisticated evaluation form mentioned above was administered by students in my Advanced Test Theory course to nearly all classes

\(^1\) This is essentially a shift from a criterion related approach to validity to a construct validity approach (Cronbach & Meehl, 1955).
in the University during the winter quarter and to a random sample of 45 classes during the spring quarter. The form was constructed by the Berkeley Center for Research and Development in Higher Education and was extensively researched over a three year period at the University of California, Davis (Hildebrand, Wilson, & Dienst, 1971). It is commonly called the Davis form.

The form was constructed by asking students at the University of California, Davis to identify the "best" and the "worst" teacher they had had that year. If a teacher were identified as best three or more times, he was placed in the "best" category and if a teacher were identified as worst three or more times, he was placed in the "worst" category. It was found that there was almost complete agreement from year to year as to the identification of best and worst teachers. In fact, there was evidence that the agreement persisted over a five year period. Faculty members were also asked to identify the best and worst teachers among their peers using the same procedure. It was found that there was virtually complete agreement among faculty and student perceptions of best and worst teachers.

A list of 236 items which were descriptive of a wide range of different aspects of teaching was prepared and students were asked to indicate which items were descriptive of the teachers that had been categorized as best or worst. This resulted in a shorter 91-item form composed of items that were said to "discriminate" best from worst teachers. An item was said to discriminate best from worst if the item was descriptive of, say, 75% or more of the teachers categorized as best but was descriptive of only, say, 25% or fewer of the teachers categorized as worst. (Note that this was essentially a procedure for validating each item separately, against an implicit definition of good teaching.)

A factor analysis of the 91-item form was then performed. Factor analytic techniques acknowledge the fact that ostensibly different items
may actually be measuring the same thing (dimension or factor). For example, two items like "Comes to class prepared" and "Presents material systematically" might both be measuring the same dimension such as, say, an organizational dimension. A factor analysis is used to determine the number of basic dimensions measured by the items on a form and to determine which items are measuring which dimensions. The factor analysis of the 91-item form revealed that the 91 items were actually measuring only five basic dimensions. A thorough study was made of the items measuring each of the five dimensions and a new item was written for each dimension which summarized all of the items measuring that dimension. These five items and the original 91 items were administered to a new sample of students to evaluate faculty performance and it was found that each of the five new items correlated very highly with the items that they were supposed to summarize. Thus, the five new items, as a group, summarize what was being measured by the original 91-item form. These five items make up the Davis form. A copy of the Davis form is included in appendix D.

A score was obtained for each faculty member on the Davis form by finding the mean of the five items across all students in all classes in a given quarter. The Davis means were correlated with the grand means obtained from the present evaluation process in the winter quarter and in the spring quarter. The correlations in the two quarters, respectively, were:

\[ r = .66, \quad n = 86; \]
\[ r = .77, \quad n = 45; \]

which were moderately high correlations. Both correlations were significant at the .01 level. This strongly suggests that the simple 2-item grand mean and the mean based upon the more sophisticated 5-item Davis form were measuring
essentially the same thing or something very similar. Actually, this isn't too surprising, if it is realized that the "validation" procedure for the Davis form consisted of asking for a general identification of best and worst teachers, while one of the two items on which the grand means were based asked for a general rating of faculty along a good teaching dimension. That is, both forms asked for a global evaluation of teachers with respect to "good teaching". It could be argued that this type of validity, content validity, might be the best to hope for in assessing student opinion of good teaching.

**Retention Committee Ratings.** An earlier plan to validate the evaluation process was to correlate the grand means with ratings of faculty performance made by the Retention Committee. Retention decisions about faculty hired in the fall of 1971 were made by a seven man Retention Committee. The decisions were made on the basis of information in faculty files. Most faculty placed a self evaluation in their file which described the work they had done, their evaluation of the work, and their plans for the future. Also most faculty placed in their file student evaluations of their teaching and advising. In many cases these evaluations were open-ended written statements. The faculty were rated along a 9-point scale on five dimensions. Only the teaching dimension will be considered here.

Before the grand means were correlated with the Retention Committee's teaching ratings, it was decided to intercorrelate the ratings made by the committee members to determine the amount of agreement (reliability) among the raters. Only five committee members actually rated the faculty so there were ten intercorrelations. In general, there was no agreement among the members of the Retention Committee.\(^1\) The failure to obtain reliable teaching

\(^1\) There were positive and negative correlations. Only one correlation was significant at the .01 level.
ratings indicates that the ratings are not valid since reliability is a necessary requirement for validity. Consequently, the teaching rating was not used to attempt to validate the present evaluation process.

I would like to emphasize that the Retention Committee's teaching ratings were unreliable. Many faculty people have argued for written qualitative student evaluations of faculty which would be read and interpreted by a committee. The present data illustrate what is commonly found to be a problem with such a procedure: the lack of agreement or reliability of the interpretations of the qualitative statements. If student data are to be used to make quantitative decisions (salaries are quantitative), it seems more reasonable to do the quantification at the student level rather than have an intermediate committee make quantitative decisions on qualitative input from students. The latter procedure allows for the subjective biases of the committee members to enter into the decision-making process.

Class Size

It was felt by some administrators and faculty members that faculty who had taught small classes had had a distinct advantage. It was argued that in a small class it is easier to develop good rapport, that it is easier to devote more time to students on an individual basis, and that students reciprocate by giving a higher evaluation. It was even recommended that the evaluation of faculty with small classes should be adjusted downward to correct for this. To test this claim, correlations between mean evaluations and the number of cases the evaluation was based upon were computed.

It will be recalled that arbitrarily chosen classes were numbered 1 and 2

---

1 A good discussion of the relationship between reliability and validity is presented in Guilford (1954).
in the winter quarter and 3 and 4 in the spring quarter. The correlations between the class mean and class size were computed for each of these four groups of classes. The four correlations and the number of cases each was based upon were:

- **group 1:** $r = -0.06$, $n = 75$
- **group 2:** $r = 0.01$, $n = 55$
- **group 3:** $r = -0.13$, $n = 75$
- **group 4:** $r = -0.07$, $n = 49$

None of these correlations were significant at the .01 level.

The correlation was computed between the quarter mean and the number of cases the mean was based upon in both the winter and spring quarters. The two correlations, respectively, were:

- $r = -0.04$, $n = 75$
- $r = 0.12$, $n = 75$

Neither correlation was significant at the .01 level.

Finally, the correlation between the grand mean and the number of responses it was based upon was found to be:

- $r = 0.05$, $n = 75$

This correlation was not significant at the .01 level.

In light of the evidence, it appears that there was no simple relationship between class size and evaluation and that the suggestion to adjust downward the evaluation of faculty members with small classes was not justified. This finding was consistent with that generally reported in the evaluation literature (Costin, Greenough, & Menges, 1971).

**Lecture vs. Discussion**

A claim was made by some faculty members that teachers who lectured had
an advantage over teachers who conducted discussions. It was felt that most
students were accustomed to traditional teaching techniques and felt more
comfortable in a structured classroom situation. Consequently, the more
structured lecture courses would receive higher evaluations.

The evaluation literature fails to support this contention. The litera-
ture shows that if a difference in evaluation does exist between lecture and
discussion classes, that discussion courses have a small, but significant,
advantage (Costin, Greenough, & Menges, 1971; Hildebrand, Wilson, & Dienst,
1971). An analysis of the present evaluation data was performed to deter-
mine if the finding from the evaluation literature was supported.

When students in my classes administered the Davis form, faculty in
each class were asked to complete a questionnaire. One of the items on the
questionnaire asked the faculty to describe their class along a 6-point
lecture-discussion scale from (1) Formal lecture to (6) Unstructured dis-
cussion. The correlations between the rating on the lecture-discussion dimen-
sion with the class mean were computed for courses numbered 1 and 2 in the
winter quarter and for the random sample of 45 classes given the Davis form
in the spring quarter. The correlations, respectively, were:

\[ r = -0.21, \quad n = 59; \]
\[ r = -0.23, \quad n = 31; \]
\[ r = -0.06, \quad n = 35. \]

None of the correlations was significant at the .01 level. The findings
support the general findings in the literature that lecture classes do not
have an advantage over discussion classes.

Innovation

One of the mandates of the University is to develop innovative teaching.
One of the reservations expressed by the faculty about the evaluation process was that it may discourage innovation. Innovation is risky, at least, with respect to evaluation. A faculty member who teaches a new course or uses a new format in a course or a new teaching style runs the risk of failure (at least the first time through) and, consequently, low evaluations. The safer course is to avoid innovation so the argument goes.

One of the items on the faculty questionnaire which accompanied the Davis form asked faculty to rate the degree of innovation of the course along a 7-point scale. The correlation between the innovation ratings and the class means for courses numbered 1 and 2 in the winter quarter and the random sample of 45 courses given the Davis form in the spring quarter were found to be

\[ r = 0.01, \quad n = 63; \]
\[ r = -0.14, \quad n = 35; \]
\[ r = 0.02, \quad n = 39. \]

None of the correlations was significant at the .01 level.

Faculty were also asked to specify the number of times they had taught each course. This, then, was correlated with the class means. The correlations were

\[ r = 0.06, \quad n = 65; \]
\[ r = 0.28, \quad n = 35; \]
\[ r = -0.13, \quad n = 39. \]

None of these correlations was significant at the .01 level.

Finally, the public affairs colloquium (PAC) courses are a general type of innovative, non-traditional course at the University. They are concerned with current issues and public related issues. The mean evaluation for PAC courses and the mean evaluation for non-PAC courses was computed for courses
numbered 1 and 2 in the winter quarter and the random sample of 45 courses in the spring quarter. The means did not differ significantly.

Thus, there were three different lines of evidence that failed to support the contention that "trying something new will result in a lower evaluation." This issue is discussed in the literature (Hildebrand, 1972).

**Faculty Characteristics**

It was of interest to learn if certain characteristics of faculty members were related to the evaluations which they received. Consequently, correlations between the grand mean and other information about the faculty were computed.

**Academic Rank.** The correlation between academic rank and the grand mean was computed. The ranks were coded as follows: professor, 4; associate professor, 3; assistant professor, 2; instructor, 1. The correlation was

\[ r = .09, \quad n = 75 \]

which was not significant at the .01 level.

**Salary.** The correlation between monthly income and grand mean was computed. The correlation was found to be

\[ r = .04, \quad n = 75. \]

It was not significant at the .01 level.

**Experience.** The correlation between number of years of experience and the grand mean was found to be

\[ r = -.07, \quad n = 73. \]

This correlation was not significant at the .01 level. The number of years of experience used in computing the correlation was the total number of years of experience stated in faculty files. Actually, it is the number of years of teaching experience that are of interest in the present paper. Unfortunately,
that information was not readily available. However, a correlation which has bearing on the years of teaching issue was presented earlier in the section on innovation. It was reported there that the number of times the course was taught was unrelated to the class mean. This evidence suggests that experience may not be the best teacher.

**Education.** The correlation between amount of education and the grand mean was obtained. Since the actual number of years of education was not readily available, the highest degree received was used as the indicator of amount of education. The highest degree received was coded as follows: Doctorate, 4; All-but-dissertation, 3; Masters, 2; Baccalaurette, 1; No academic degree, 0. The correlation was found to be

\[ r = 0.21, \quad n = 74 \]

which was not significant at the .01 level.

Thus, it appears that there was no significant direct relationship between the grand mean and academic rank, salary, experience, and educational level. It was hoped to compare the evaluations received by faculty in different programs. However, given the multidisciplinary nature of the University, it was difficult to categorize the faculty into mutually exclusive disciplinary program groups. To have done so would have required either the arbitrary forcing of many faculty into questionable categories or a large "miscellaneous" category. Consequently, the analysis was not performed.

**Student Characteristics**

Some administrators and faculty felt that students should sign their names to the evaluation forms. It was felt that having students sign the forms would encourage student responsibility. On the other hand, faculty involved in the development of the evaluation process were concerned that having students
sign might inhibit critical responses which would result in generally high evaluations for all faculty (a ceiling effect). It was finally decided by the Evaluation Committee to make signing the forms optional.

The question was then raised as to whether there was any difference between the evaluation given by students who signed the forms and students who did not sign the forms. The question was answered by drawing a random sample of 50 classes from the spring quarter classes and separating the evaluation forms for each class into those that were signed and those that were not signed. The mean evaluation given by the signers was 4.64 and the mean evaluation given by the non-signers was 4.43. The non-signer's mean evaluation was significantly lower than the signer's mean evaluation.

In light of the evidence it might be recommended that students be required not to sign the evaluation forms. The evidence suggests that this might decrease the ceiling effect and increase the range of the grand means. Increasing the range should increase the reliability of the grand mean and decrease the error of measurement (McNemar, 1969). Fortunately, there is also evidence which suggests that the general ordering of faculty on the basis of the grand means would be unaltered regardless of whether students were required to sign or not. The correlation between the mean evaluations given by signers and non-signers was found to be

\[ r = .71, \quad n = 48 \]

which was significant at the .01 level. Thus, signers and non-signers appear to order faculty the same way.

It should not necessarily be concluded on the basis of the data that had

\[ t = 4.95 \]

which was significant at the .01 level.
The evaluation given the faculty member by a given student was the mean of the ratings on the competency item and the teaching item.

Grades and Evaluation. A commonly expressed concern of many faculty members was that students who were to receive a low grade had retaliated and given a low evaluation; that is, many faculty felt there was a correlation between the student's grade and the evaluation given. This contention was tested as follows: a random sample of 150 students was drawn from all students registered in the winter quarter. Since most students were enrolled for, at least, two or more courses, one course was randomly chosen for each student. A search through the evaluation forms for that course was conducted to find the evaluation form signed by that student. Due to absenteeism and the fact that many students did not sign the forms, the signed evaluation forms for only 39 students from the original sample of 150 were found. Also since many students had registered for the pass-no credit option, grades of A, B, and C were available for only 29 of the 39 students. Although the sample of 29 students was biased, the correlation between the grade the student received and the evaluation the student gave the teacher was computed. The correlation of \( r = .27 \) was not significant at the .01 level. In spite of the fact that the sample was biased, the failure to find a significant correlation is consistent with the reports in the evaluation literature. These studies have

---

1 The evaluation given the faculty member by a given student was the mean of the ratings on the competency item and the teaching item.
generally found no relationship between grades and evaluation; the positive relationships that have been reported were typically low (Costin, Greenough, & Menges, 1971; Hildebrand, Wilson, & Dienst, 1971).

Age, Sex, and Undergraduate-Graduate Classification. Using the same sample, there was no significant relationship found between the age, sex, or undergraduate-graduate status of a student and the evaluation the student gave the faculty member. Again the results from this biased sample were consistent with the reports in the literature (Costin, Greenough, & Menges, 1971; Hildebrand, Wilson, & Dienst, 1971).
Salary and promotion decisions were made on a merit basis at Sangamon State University for the 1971-72 academic year. The two factors upon which the merit decision was based were a teaching rating and a service rating. The teaching factor was determined by student ratings of faculty on a questionnaire which was administered to all classes during the spring and winter quarters. The service to the University and community rating was determined jointly by the faculty member's dean and an elected peer. The information on teaching and service was turned over to the Salary and Promotion Committee. The Committee reviewed the information on a University-wide basis and made final ratings on teaching and service for every faculty member in the University. An overall rating was obtained for each faculty member which was a weighted summary of the teaching and service rating with teaching weighted 60% and service weighted 40%. Salary and promotion decisions, then, were generally based upon the final overall rating.

In December, 1971, the services of a consultant were obtained to construct a form to be used by students to evaluate faculty teaching performance. The consultant's work resulted in a 20-item evaluation form which was administered to all classes in the winter quarter. The form was shortened to an 8-item form, which was administered in the spring quarter. However, after a careful reading of the responses to the 20-item form and the 8-item form, it was felt that many items were ambiguous and were misinterpreted by students and that many items discriminated against faculty with certain teaching styles or philosophy or who were in a certain discipline. Consequently, the evaluation of teaching was based only upon three of the items in the winter and spring evaluation forms. The three items, respectively, asked about the
faculty member's competency and teaching ability and asked if the faculty member should be retained. The Salary and Promotion Committee was presented with the mean rating and standard deviation for the first two items combined for each class taught by a given faculty member. The grand mean and standard deviation over all classes taught by the faculty member was also presented as was a frequency distribution of the responses to the retention item. In general, this was the information used by the Salary and Promotion Committee in arriving at the teaching rating.

An extensive research analysis of the data obtained from the evaluation process was performed. In general, it was found that the measures obtained from the evaluation process were reliable. There was agreement between means of different classes taught by a given faculty member. There was agreement between the mean evaluations obtained in the winter and spring quarters. The projected reliability coefficients for the grand mean which was based upon information from two quarters and for a grand mean to be based upon information from three quarters were high.

The difficulty of validating the evaluation process (and any evaluation process) due to the lack of a generally agreed upon definition or criterion of good teaching was discussed. However, the grand means were shown to have a high correlation with means obtained from a more sophisticated, research-based, 5-item evaluation questionnaire, the Davis form. Finally, an attempt to use the Retention Committee's teaching ratings as a criterion of good teaching revealed the unreliability of these ratings. A discussion of the problems of a committee approach to evaluation where committee members read written qualitative evaluations was presented.

No relationship was found between class size and evaluation. There was no
relationship between ratings faculty made of their classes along a lecture-discussion dimension and the evaluation received in the classes. There was no evidence that innovation penalized evaluations: ratings faculty made of the innovativeness of their classes were unrelated to evaluations; the number of times a course had been taught by a faculty member was unrelated to evaluation; evaluations given in public affairs colloquia (PAC) courses did not differ from evaluations given in non-PAC courses.

It was found that academic rank, salary, experience, and highest degree held were unrelated to evaluation.

Students who did not sign the evaluation forms gave significantly lower evaluations than did students who signed the forms.

Using a small biased sample, it was found that there was no relation between the grade that a student received in a class and the evaluation the student gave the faculty member. Using the same sample, there was no relationship found between age, sex, and undergraduate-graduate classification and the evaluations given by the student. Although the sample was small and biased, the findings were consistent with those reported in the evaluation literature.

In summary, the research has shown that the evaluation process used had moderate to high reliability and had a moderately high correlation with a more sophisticated evaluation process. It has also been shown that certain biases which many faculty members feared were present in the evaluation instrument either did not exist or had only an insignificant effect.
REFERENCES


SANGARON STATE UNIVERSITY

FACULTY EVALUATION

STUDENT FORM

There are a variety of functions or roles that a teacher may fill in different learning situations. Not all of those mentioned in this form may be appropriate to the learning situation or faculty individual you are being asked to evaluate. This form focuses on four different ways in which teachers may aid your learning and growth as an individual. The final sections provide for a more specific evaluation focusing on the goals of your teacher and your own goals in this learning situation. You are asked to consider each of these sections in relation to the instructor you are evaluating.

This form is intended to be used in two different ways:

1. Numerical Ratings: For each question, rate your instructor on the seven point scale on the right using 1 to indicate definite relationship between the question and this learning situation and instructor. For the reverse use 7. If the question is not appropriate to use in this situation, leave it blank.

2. Written Evaluations: Each question is followed by a space for your written response. Please use this space to clarify and explain the rating you gave. You may wish to give specific examples or indicate ways that this faculty member has functioned in this or other learning situations which would extend your evaluation beyond the mere rating.

Learning Situation Evaluated  Instructor Evaluated  Date
Section 1: Teacher as Facilitator

In taking stock of developments in yourself during this academic term, consider to what extent this teacher may have served as a catalyst or facilitator in the learning situation.

1. Did your skill in written communication increase during this course? (1) (2) (3) (4) (5) (6) (7)

2. Did your ability to express yourself verbally in raising questions and discussing issues improve during this course? (1) (2) (3) (4) (5) (6) (7)

3. Did your ability to think critically develop during this course? (1) (2) (3) (4) (5) (6) (7)

4. Was the instructor aware of when students did not understand material or were bored? (1) (2) (3) (4) (5) (6) (7)
Section I (continued)

Did you become more self reliant and independent in your learning during this course? (1)(2)(3)(4)(5)(6)(7)

Did the instructor stimulate your interest in the subject matter by his own enthusiasm? (1)(2)(3)(4)(5)(6)(7)

Did your perspective change during this course: did you become more sensitive to new phenomena and ideas? (1)(2)(3)(4)(5)(6)(7)

Did you develop as a person during this course? Are you more responsive to others and/or do you know and accept yourself more fully? (1)(2)(3)(4)(5)(6)(7)
Section 2: Teacher as Diagnoser and Advisor

One way that a teacher may function in aiding your learning is by choosing and evaluating problems that you have and advising you a program of study.

Were assignments, tests, etc., of the right number and at the right time to foster your learning? (1) (2) (3) (4) (5) (6) (7)

Were the comments and evaluations of your papers, tests, etc., helpful as feedback in your learning? (1) (2) (3) (4) (5) (6) (7)

Was this teacher helpful in giving you guidance and support when you needed it? (1) (2) (3) (4) (5) (6) (7)

Were you given freedom to develop your own ideas and encouraged to try out your own solutions to problems? (1) (2) (3) (4) (5) (6) (7)

Does this teacher know you well enough so that you would seek out his/her advice in areas other than this learning situation? (1) (2) (3) (4) (5) (6) (7)
faculty members of a university can be viewed as one of the resources available to students in their learning and development.

- After being involved in this learning situation do you believe that you have mastered a significant area of knowledge and/or developed skills useful in dealing with this subject matter?

- Would you feel it worthwhile to seek out this teacher if you have questions about the subject matter or resources in his field of specialization?

- Did working with this faculty member help you integrate material and develop broader generalizations in this subject matter?

- Did this instructor help by clarifying or developing relevant examples to make readings or other resources more understandable for you?

- Was the instructor able to promote class discussions and student participation by raising challenging questions?
cause of Sangamon's special foci on innovation and public
facing, faculty members can serve as agents for fostering
of these institutional goals.

Were there significant attempts made to innovate either in the
kinds of material covered or the teaching-learning methods used?

 Were the goals of the course developed jointly with students
or adjusted to take account of student interests?

 Was there consideration of the possible relevance of the
content or skills developed in this course to their potential
implications and use in public affairs?

 Would you consider this faculty member a potentially useful
advisor for you on your applied study quarter?
EVALUATION FORM

Your candid and thoughtful evaluation of this course and teacher will help to improve instruction at Sangamon State, since the forms will assist in bringing out the strengths and weaknesses of both courses and instructors. Student evaluation forms will also be used as one source of information in determining faculty merit pay increases and retention.

Four of the possible roles of a teacher at Sangamon State are utilized as the core ideas around which questions for this form have been constructed. These are: Teacher as 1) facilitator; 2) advisor; 3) resource; 4) orienter toward public affairs.

In Section I, each question will have a numbered scale. Descriptive words above the numbers indicate their meaning. Please write your own response to each of the questions, explaining your views, giving examples, and so on, in addition to circling a number. If a question is not appropriate in this situation, please circle NA (not applicable.)

Section II is an open space in which to respond to this instructor and learning situation relative to gaining a sense of yourself in your personal and social history.

Section III asks for your opinions on this evaluation form itself.

Course title: ___________________________ Instructor: ___________________________

Date: _______________ Student (optional): ____________________________
Section I.

1. Did your ability to think critically develop, that is, the capacity to see the point at issue, formulate meaningful questions, and understand the problems involved?

<table>
<thead>
<tr>
<th></th>
<th>very well</th>
<th>average</th>
<th>poorly</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Explain:

2. Did your ability to express yourself orally increase?

<table>
<thead>
<tr>
<th></th>
<th>greatly</th>
<th>moderately</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Explain:

3. Did your ability to communicate in writing increase?

<table>
<thead>
<tr>
<th></th>
<th>very much</th>
<th>somewhat</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Explain:

4. Did the teacher enable you to become more self-reliant and independent in your learning?

<table>
<thead>
<tr>
<th></th>
<th>yes</th>
<th>somewhat</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Explain:
5. Was your enthusiasm for learning stimulated?

<table>
<thead>
<tr>
<th>greatley</th>
<th>moderately</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA 5</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Explain:

6. Was there freedom to develop your own ideas and encouragement to try out your own solutions to problems?

<table>
<thead>
<tr>
<th>definitely yes</th>
<th>uncertain</th>
<th>definitely no</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA 5</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Explain:

7. Do you feel that you have an understanding of what the course was about?

<table>
<thead>
<tr>
<th>definitely yes</th>
<th>uncertain</th>
<th>definitely no</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA 5</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Explain:

8. Were the assignments (of whatever form) valuable for your learning?

<table>
<thead>
<tr>
<th>greatley</th>
<th>moderately</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA 5</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

9. Were the comments and evaluation on the formal assignments helpful?

<table>
<thead>
<tr>
<th>most helpful</th>
<th>somewhat helpful</th>
<th>not at all helpful</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA 5</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Explain:
10. Were the comments and evaluations on your oral participation helpful?

<table>
<thead>
<tr>
<th>Level</th>
<th>Most Helpful</th>
<th>Somewhat Helpful</th>
<th>Not at All Helpful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most helpful</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Somewhat helpful</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Not at all helpful</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Explain:

11. Was the teacher available and useful in giving you guidance and support when you needed it?

<table>
<thead>
<tr>
<th>Level</th>
<th>Definitely Yes</th>
<th>More or Less</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitely yes</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>More or less</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Explain:

12. Do you think this teacher is competent in the content or matter offered in this course?

<table>
<thead>
<tr>
<th>Level</th>
<th>Exceptionally Competent</th>
<th>Satisfactory</th>
<th>Incompetent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exceptionally competent</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Incompetent</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Explain:

13. Did the teacher encourage you to utilize resources other than himself, assigned materials and the classroom situation?

<table>
<thead>
<tr>
<th>Level</th>
<th>Definitely Yes</th>
<th>Moderately Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitely yes</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Moderately yes</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Explain:
14. Was the content of the class experience informative beyond what you gained through utilizing the assigned materials?

<table>
<thead>
<tr>
<th>definitely yes</th>
<th>perhaps</th>
<th>definitely no</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA 5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Explain:

15. Did the class move intellectually through discussions?

<table>
<thead>
<tr>
<th>definitely yes</th>
<th>somewhat</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA 5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Explain:

16. If the teacher offered films, recordings, field experiences, and so on, were they valuable for your learning?

<table>
<thead>
<tr>
<th>most valuable</th>
<th>of some value</th>
<th>of no value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA 5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Explain:

17. Were lab sessions valuable?

<table>
<thead>
<tr>
<th>definitely yes</th>
<th>somewhat</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA 5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Explain:
18. Was there consideration of the possible relevance of the content or skills developed in this course to their potential implications and use in public affairs?

<table>
<thead>
<tr>
<th>Definitely Yes</th>
<th>To Some Degree</th>
<th>Not At All</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Explain:

19. Overall, do you consider this person a good teacher?

<table>
<thead>
<tr>
<th>Excellent</th>
<th>Good</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Explain:

20. After careful consideration, do you think that this teacher should be:

_____ retained

_____ retained, but encouraged to improve in the following areas:

_____ not retained.
Section II.

To what extent has your participation in this teacher's course helped to clarify your own understanding of your past and present, and how you view yourself in the future?

Section III.

Do you have any comments on this evaluation form itself?

Thank you.
EVALUATION FORM

Your candid and thoughtful evaluation of this course and teacher will help to improve instruction at Sangamon State, since the forms will assist in bringing out the strengths and weaknesses of both courses and instructors. Student evaluation forms will also be used as one source of information in determining faculty merit pay increases and retention.

The form is divided into two sections --

Section I: Descriptive words above a numbered scale indicate the meaning of the numbers. Please circle the number which most accurately reflects your opinion. If a question seems completely inappropriate, circle NA (not applicable).

Section II: Please write all your comments in this portion of the form, not after the questions themselves.

Course title: ___________________________ Instructor: ___________________________

Date: ___________________________ Student (optional): ___________________________
Section I.

1. Do you think this teacher is competent in the content or matter offered in this course?

   exceptionally competent  satisfactory  incompetent
   NA  5  4  3  2  1

2. Was your enthusiasm for learning stimulated:

   greatly  moderately  not at all
   NA  5  4  3  2  1

3. Was the teacher available and useful in giving you guidance and support when you needed it?

   definitely yes  more or less  no
   NA  5  4  3  2  1

4. Was there freedom to develop your own ideas and encouragement to try out your own solutions to problems?

   definitely yes  uncertain  definitely no
   NA  5  4  3  2  1

5. Do you feel that you have an understanding of what the course was about?

   definitely yes  uncertain  definitely no
   NA  5  4  3  2  1

6. Was the content of the class experience informative beyond what you gained through utilizing the assigned materials?

   definitely yes  perhaps  definitely no
   NA  5  4  3  2  1

7. Overall, do you consider this person a good teacher?

   excellent  good  poor
   NA  5  4  3  2  1

8. After careful consideration, do you think that this teacher should be:

   retained
   retained, but encouraged to improve in the following areas:

   not retained
Each of these statements describes a basic component of teaching. Give the instructor an overall rating for each component, reserving the highest scores for unusually effective performance.

<table>
<thead>
<tr>
<th>Component</th>
<th>LOW SCORE</th>
<th>HIGH SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Has command of the subject, presents material in an analytic way, contrasts points of view, discusses current developments, and relates topics to other areas of knowledge.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>2. Makes himself clear, states objectives, summarizes major points, presents material in an organized manner, and provides emphasis.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>3. Is sensitive to the response of the class, encourages student participation, and welcomes questions and discussion.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>4. Is available to and friendly toward students, is interested in students as individuals, is himself respected as a person and is valued for advice not directly related to the course.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>5. Enjoys teaching, is enthusiastic about his subject, makes the course exciting, and has self-confidence.</td>
<td>2 3 4 5 6 7</td>
<td></td>
</tr>
</tbody>
</table>