1-1-1975

The Relation Between a Teacher Evaluation Checklist and Actual Student Course Gains

Thomas Michael Ozee

Eastern Illinois University

This research is a product of the graduate program in Psychology at Eastern Illinois University. Find out more about the program.

Recommended Citation
http://thekeep.eiu.edu/theses/3558

This Thesis is brought to you for free and open access by the Student Theses & Publications at The Keep. It has been accepted for inclusion in Masters Theses by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.
TO: Graduate Degree Candidates who have written formal theses.

SUBJECT: Permission to reproduce theses.

The University Library is receiving a number of requests from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow theses to be copied.

Please sign one of the following statements:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

[Signature]

Date

10/1/75

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because __________________________

______________________________

______________________________

______________________________

______________________________

Date

Author

pdm
THE RELATION BETWEEN A TEACHER EVALUATION CHECKLIST AND ACTUAL STUDENT COURSE GAINS

BY

THOMAS MICHAEL OZEE

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Arts in Psychology

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

1975

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING THIS PART OF THE GRADUATE DEGREE CITED ABOVE

Sept 30, 1975
DATE
Abstract

The relationship between student gain, as measured by pre and post achievement test measures, and a teacher evaluation checklist, the Student Evaluation of College Teaching Behaviors Instrument (SECTB), was examined in an attempt to validate the eight factors in the SECTB. Sixty seven students in three sections of introductory psychology were measured on both instruments. Results of this study do not support a relationship between student gain and the SECTB. The only significant correlate of gain was a factor measuring nature of presentation \( (r = -0.468) \) and this correlation was found in only one section. In this section there was a surprisingly large number of significant interfactor correlations (62%) in the SECTB. Only three of the 45 items discriminated between high and low gain scores. The relationship of gain and student ratings awaits refinements in measuring both characteristics.
TABLE OF CONTENTS

Introduction ................................................................. 1
Method ................................................................. 10
Subjects ................................................................. 10
Procedure ................................................................. 10
Results ................................................................. 13
Discussion ................................................................. 18
References ................................................................. 20
Appendix ................................................................. 22
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Correlation Matrix</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>Correlations Between Items and Gain</td>
<td>15</td>
</tr>
</tbody>
</table>
In the past few years, there has been an increasing interest in the evaluation of instruction as can readily be seen by the voluminous number of research articles written about this subject. For example, Feldman (1973) cited over 100 articles on student assessment of teaching effectiveness. His references contained information on college student's descriptions of effective or ideal teachers, student's ratings of college teachers they have had, and correlates of student assessments. deWolf (1975) produced a 220 item annotated bibliography of research reported since 1968 on student ratings of instruction in postsecondary institutions. This 220 item bibliography is arranged in alphabetical order, with a series of alphanumeric indices at the end of each entry describing the various topics covered by each article. This alphanumeric system covers 11 major areas and 315 subheading areas, dealing with the various aspects of student evaluation of instructors.

Students and faculty have developed a keen interest in the evaluation of instruction. Students have a stake in the quality of instruction that they receive, and feel that as consumers of education, they should have a voice in it (Costin, Greenough, and Menges, 1971). Faculty are interested in evaluations for reasons of promotion, tenure, and salary. The use of evaluations for these purposes has led to criticism of the evaluations on the part of the faculty. This criticism is centered around the high disagreement and varying results among the diverse assessment techniques used for evaluation.

Zelby (1974) indicates that student evaluations of faculty can
inhibit educational experimentation and development, especially if the opinions are used in formally determining salaries and promotions. He also suggests that the indiscriminate use of evaluations may increase the gap between first and second rate institutions, his reasonings being that the latter may use opinionnaires that emphasize popularity and mediocrity of education in an attempt to maintain a high level of enrollment. Zelby also feels that student evaluation of faculty in any form could permit administrators to shirk their responsibilities in the evaluation of teaching performance, and still offer the opinions as proof that something is being done.

Harari and Zedeck (1973) state that current forms of student evaluations are often ambiguous, verbose, disorganized, and arbitrarily developed. Further, they consist of global behavioral measures and vague trait descriptions, with the result, that the evaluation forms tend to be unreliable and susceptible to response biases. This is supported by McGuigan (1974), who states that there is a lack of sound methodologies in applying evaluations. He feels that evaluations are equated with the administration of casually constructed opinionnaires, and that little has been accomplished in the way of a systematic assessment of instructors and courses.

Rodewald and Carroll (1974) in using a paired - comparison stimuli in an attempt to scale student attitudes toward professors concluded that the, "data warn against the uncritical use of item - rating scales for evaluation of professional performance. When students cannot use
items on a questionnaire to express their general attitude in a consistent manner, the addition of a rating scale which assigns numbers to these attitudes is clearly a matter for concern if these numbers are given some later mathematical treatment". This is further supported in a review of the literature, by Kulik and Kulik (1974), that concluded that ratings are influenced by class size, upper vs. lower and elected vs. required status of course, and the discipline or department of the course. Subject matter differences within departments may also influence course ratings.

Proponents of student evaluations feel that student ratings could provide feedback to the instructor, norms against which faculty ratings could be judged, areas of strength or weakness in teaching areas, and a source of information for the students to aid him in selection of courses (Costin, Greenough, and Menges, 1971). However, these benefits of student evaluation of instruction can only be realized to the extent that the student evaluations reflect accurate and valid appraisals of classroom instruction received. In support of the reliability of student evaluations, Guthrie (cited in Costin et al 1971) found correlations of .87 and .89 between student's rankings of the quality of their teachers from one year to the next. Guthrie also discovered that such judgements were more stable than were faculty judgements of teaching quality. Centra (1973) reported only a modest correlation of .21 for the relation between teacher self ratings and ratings given by students, and that teachers as a group tended to give
themselves higher ratings than students did. Lovell and Haner (1955) using a forced choice rating scale found a correlation of .59 between ratings made two weeks apart.

Concerning research productivity and student evaluations, Siebring and Schaff (1974) stated in their review of the literature that most investigators conclude that student ratings provide valuable information if used wisely, but that results of investigations on the effect of instructor research productivity on teaching effectiveness are inconclusive. Aleamont and Yimer (1973) found that colleague and student ratings were not significantly related to instructor's research productivity. They did find that colleague ratings were significantly related to academic rank, which they feel gives some indication that the reputation of the instructors could be influencing colleague ratings. However, Maslow and Zimmerman (1956) found moderate relations in comparisons of student ratings of instruction, research productivity, and faculty (peer) ratings.

A review of the literature has revealed several methods of evaluating instruction. McKeachie and Solomon (1958) suggest that arousal of student interest in a field of study may be a useful criterion for measuring teaching effectiveness. They reported that there was a significant relationship between student's ratings of a psychology instructors teaching ability and the decision to take a further course in psychology in two out of five instances. Rodin and Rodin (1972) stated that there are two ways to evaluate teaching through the medium of students: (1) the subjective criterion of
teacher effectiveness based on student evaluations of their teachers, and (2) the objective criterion based on what students have learned, i.e., student gain. When final grade is used as the objective criterion, approximately half of the studies (Costin, Greenough, and Menges, 1971) showed positive correlations between grades and student evaluations, while the rest show negative or no correlations between the two. Kulik and Kulik (1974) concluded that the most striking thing about the studies that related student achievement to student ratings was the inconsistency of the results. They also concluded that overall, there was a slight tendency for students of highly rated teachers to outscore students of low rated teachers on final examinations. When a more precise measure has been used, such as scores on an achievement test, positive correlations have usually been reported (Morsh, Burgess, and Smith, 1956). Rodin and Rodin (1972), however, found a correlation of -0.75 between the rating of overall performance and final grade. Their conclusion was that students are unable to judge teaching effectiveness if the latter is measured by how much they have learned. This conclusion was challenged by Gessner (1973), who found correlations of 0.77 for performance on national normative exams and the student ratings of the content and organization of course instruction, and 0.69 for exam performance and the presentation of course instruction. Frey (cited in Sullivan 1974) found positive correlations of 0.95 between instructor’s rating and student’s examination performance for two mathematics courses. Sullivan and Skanes (1974) found out of ten
subject areas, where a correlation was made between instructor rating and final examination achievement, that nine were positive, and eight of those were above .32. Their conclusion was that there was a modest, but significant relationship between student evaluation of instruction and student achievement.

When the ratings of students in a given class were randomly paired, correlations ranging from .77 to .94 were reported by Guthrie (cited in Costin et al 1971) and Maslow and Zimmerman (1956). The mean odd item ratings on a forced choice instrument were found by Lovell and Haner (1955) to be correlated .79 with the mean even item ratings. Costin et al (1971) states, "... it would appear then, that students can rate classroom instruction with a reasonable degree of reliability" (p. 513). He further states, "... students are at least partially capable of distinguishing certain qualities of instruction which increase their knowledge or motivation" (p. 514).

McGuigan (1974) and Greenwood, Bridges, Ware, and Mclean (1973) stress the need for more empirically developed forms of teacher evaluation than are currently in use. McGuigan (1974) has developed what he terms the G statistic. The G statistic is a ratio between gain to possible gain, and therefore an index of amount learned relative to possible amount of learning. It therefore goes beyond a simple gain score, by eliminating the problem of artificial restriction of amount learned for students with a high pre test score. In this manner, he uses amount learned as a basis for evaluation.
According to a review by Kulik and Kulik (1974), the earliest factor analysis of student ratings were done on the Purdue Rating Scale. In three studies cited by Kulik and Kulik (1974) investigators found, in independent analysis, two factors in the ten items. The first factor reflected instructor competence, and the second factor suggested instructor empathy and rapport with students. They mention that these two factors emerge as two major dimensions in more recent factor analysis of student ratings, where more sophisticated methods and large item pools were used. Costin et al (1971) in a review of student rating of college instruction reported that Deshpande, Webb, and Marks derived 14 dimensions of teaching behavior, through factor analysis, representing 147 behavioral items.

Greenwood et al (1973) has developed the Student Evaluation of College Teaching Behaviors Instrument (SECTB). The SECTB is an eight factor evaluation instrument that deals with specific observable behaviors of the instructor, and permits the students to rate only those items which they consider relevant. The SECTB was developed by sampling students, faculty, and academic administration at a large southern state university. Respondents described six characteristics of the best and worst college instructor they have known. The responses were content analyzed into categories of characteristics. Behavioral statements were generated that represented each category of characteristics. The 134 statements were submitted to 76 classes and their instructors, where they were asked to indicate those
statements which were characteristic behaviors of either good or bad college instruction. The 85 surviving items were examined for behavioral specificity and modified if found lacking. The modified items were submitted to a sample of 425 students and to the entire teaching faculty of 1529. Subjects rated each item on a seven point scale according to whether it was related to bad college teaching, good college teaching, or unrelated to either good or bad college teaching. Results were analyzed separately for students and faculty. Means and standard deviations were calculated for each of the items. An interval was formed by adding and subtracting one standard deviation from the item mean. If the resulting interval included 0, the item was classified as neutral. Using this criterion, 20 items were rejected. Any item having a standard deviation greater than 1.24 was eliminated for having too much variability. Five items were eliminated in this manner, leaving a final set of 60 items. The 60 items were factor analyzed, using a principal axis solution and then rotated to the varimax criterion for both students and faculty. The sample was then combined and analyzed using the same procedure as for the separate analysis. The authors stated, that using a standard factor analysis criterion, eight factors were identified: (1) facilitation of learning, (2) obsolescence of presentation, (3) commitment to teaching, (4) evaluation, (5) voice communication, (6) openness, (7) currency of knowledge, and (8) rapport.

In light of the development of Greenwood's SECTB and McGuigan's
G factor this study is designed to examine the relationship of the SECTB to actual student gain, in an attempt to validate the eight factors in the SECTB.
Method

Subjects

The subjects were 67 undergraduate students (26 male, 41 female) at Eastern Illinois University, enrolled in a summer semester introductory psychology class. The subjects were divided into three sections under two instructors, section one (29 Ss) under instructor A, section two (15 Ss) and section three (23 Ss) under instructor B. 12 of the subjects had a prior psychology class in high school. Due to reasons of non attendance and dropping of the class, four subjects were lost in section one, three in section two, and eleven in section three, giving a final total of 49 Ss. Seven subjects were tested after the regular post test date due to non attendance on testing day. Each section met for a 50 minute period, five days a week. Section one met at 11:10 A. M., section two at 12:30 P. M., and section three at 1:40 P. M.

Procedure

The objective tests used as pre and post measures (see appendix) were administered at the beginning of the semester during the first full week of school, and again at the close of the semester, during the final week of classes before final examinations. The SECTB was presented the same day, following the post test, during the final week of classes (see table two). Students identified themselves on the SECTB form with the assurance that the ratings would not be exposed by name to their instructor.
Students were given the following instructions concerning the pre test. "This is a pre test for introductory psychology. It is being administered to determine the extent of your knowledge in the areas covered in introductory psychology. You are not being graded on these tests, but do the best that you can. If you have had any previous psychology courses in high school, please indicate so at the top of the answer sheet."

Instructions for the post test were the following. "This is a post test. It is being administered to determine how much you have gained in this class. You are not being graded on this exam, but do the best that you can." Instructions for the SECTB were, "This is the Student Evaluation of College Teaching Behaviors Instrument. It contains specific behavioral items relating to your instructor. If the statement applies to your instructor, check it. If the statement does not apply to your instructor, leave it blank. Write your name and section number at the top of the SECTB form. Your instructor will not see these evaluations, and your answers will in no way affect your grade."

Results of pre and post test measures were examined, and a gain score was established for each student using McGuigan's G statistic (McGuigan 1974). The G statistic being computed as follows:

\[ G = \frac{T_2 - T_1}{r - T_1} \]

where \( T_1 \) = pre test score
\( T_1 \) = post test score
\( r^2 \) = possible score
For purposes of this study, repeated items over factors on the SECTB were excluded where they had a low factor loading, and retained where they had a high factor loading. This resulted in the deletion of six items from the original form. Items deleted were; presented irrelevant material during lectures (factor three), tested primarily for isolated and/or obscure details (factor four), stated course objectives (factor four), established and kept office hours for individual conferences (factor four), would not deal with questions covering material beyond text (factor seven), and listened to student's problems (factor eight). This deletion of repeated items served to simplify the statistical analysis of item correlation to gain.

The pre and post test instrument is a 45 item test, prepared for evaluating large group instruction at Eastern Illinois University. Items were selected from a publishers pool (Morgan and King 1971), with questions from chapters covering the areas of development, learning, motivation, statistics, and psychopathology (see appendix for specific questions).

For purposes of multiple regresional analysis, data from sections two and three were combined, due to the low number of subjects in these sections. Biomedical (BMD02R) stepwise regression program was used for multiple regresional analysis (Dixon 1974).

Item analysis was accomplished by using a point biserial r with the following formula; 

\[ r_{pbi} = \frac{M_p - M_t}{\sqrt{\frac{P}{q}}} \]

where 

- \( M_p \) = mean gain for those checking 
- \( P \) = proportion of checks 
- \( M_t \) = mean for all gain scores 
- \( q \) = proportion of non checks 
- \( \sigma_t \) = standard deviation for gain scores
Results

Analysis of data revealed only one significant correlation between gain and factors in the combined sections two and three. \( r(23) = -0.468, P < 0.05 \). This correlation was between gain and factor two (obsolescence of presentation). No significant correlations between gain and factors were discovered in section one (see table one). Multiple regression analysis using the BMD02R stepwise regression program revealed only one factor (factor two, combined sections two and three) that contributed significantly to the overall correlation of factors with gain; other factor to gain correlations being too low to contribute significantly to the multiple regression.

Item analysis revealed two significant correlations between items and gain in section one. Item one in factor four (told students what was expected of them) \( r(24) = -0.53, P < 0.05 \) and item five in factor seven (presented material as an extension of the text) \( r(24) = 0.44, P < 0.05 \). One significant correlation was found in section two, item three, factor two (presented obsolete material) \( r(11) = -0.56, P < 0.05 \). No significant correlations between items and gain were found in section three (see table two for complete item analysis).

Approximately sixty percent of the inter-factor correlations in the combined sections two and three were significant at the 0.05 level, with five of these correlations significant at the 0.01 level. Approximately twenty one percent of the factor intercorrelations were significant at the 0.05 level in section one, with two of these being significant at the 0.01 level (refer to table one).
TABLE 1
CORRELATION MATRIX
SECTION I
FACTOR

<table>
<thead>
<tr>
<th>VARIABLE (GAIN)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>1.000</td>
<td>0.205</td>
<td>-0.314</td>
<td>0.124</td>
<td>-0.077</td>
<td>0.210</td>
<td>-0.108</td>
<td>0.317</td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
<td>1.000</td>
<td>0.310</td>
<td>0.156</td>
<td>0.040</td>
<td>0.196</td>
<td>0.419*</td>
<td>0.393*</td>
</tr>
<tr>
<td>3</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.015</td>
<td>-0.080</td>
<td>0.225</td>
<td>-0.285</td>
<td>-0.242</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
<td>0.616**</td>
<td>0.046</td>
<td>0.370</td>
<td>0.215</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
<td>0.240</td>
<td>0.460*</td>
<td>0.004</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
<td>0.553**</td>
<td>0.279</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
<td>0.403*</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SECTIONS II AND III

<table>
<thead>
<tr>
<th>VARIABLE (GAIN)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>1.000</td>
<td>-0.078</td>
<td>-0.468*</td>
<td>0.031</td>
<td>-0.044</td>
<td>0.288</td>
<td>0.011</td>
<td>-0.017</td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
<td>1.000</td>
<td>0.497*</td>
<td>0.531*</td>
<td>0.655**</td>
<td>0.432*</td>
<td>0.864**</td>
<td>0.512*</td>
</tr>
<tr>
<td>3</td>
<td>1.000</td>
<td>-0.165</td>
<td>-0.058</td>
<td>-0.319</td>
<td>-0.264</td>
<td>-0.248</td>
<td>-0.482*</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.000</td>
<td>0.331</td>
<td>0.527**</td>
<td>0.119</td>
<td>0.565**</td>
<td>0.470*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.000</td>
<td>0.486*</td>
<td>0.334</td>
<td>0.516**</td>
<td>0.479*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.000</td>
<td>0.192</td>
<td>0.491*</td>
<td>0.437*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.000</td>
<td>0.436*</td>
<td>0.210</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.000</td>
<td>0.472*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* P ≤ .05
** P ≤ .01
<table>
<thead>
<tr>
<th>ITEM</th>
<th>Sec. I</th>
<th>Sec. II</th>
<th>Sec. III</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gave organized answers to complicated questions in class</td>
<td>.02</td>
<td>-.30</td>
<td>-.22</td>
</tr>
<tr>
<td>I_2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Permitted students to express opinions which differed from his/her own</td>
<td>.08</td>
<td>-.20</td>
<td>0</td>
</tr>
<tr>
<td>I_3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encouraged students to ask questions</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
</tr>
<tr>
<td>I_4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dealt with student difficulties before they arose</td>
<td>.05</td>
<td>.40</td>
<td>-.05</td>
</tr>
<tr>
<td>I_5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilized background of students to aid in class activities</td>
<td>.20</td>
<td>.16</td>
<td>.38</td>
</tr>
<tr>
<td>I_6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encouraged class discussions</td>
<td>.24</td>
<td>.22</td>
<td>.40</td>
</tr>
<tr>
<td>I_7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explained the reasons for his/her criticisms</td>
<td>.12</td>
<td>.38</td>
<td>.20</td>
</tr>
<tr>
<td>I_8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivered orderly, logical presentations of the material</td>
<td>.01</td>
<td>.16</td>
<td>.15</td>
</tr>
</tbody>
</table>

**FACTOR II - OBSOLESCENCE OF PRESENTATION**

<table>
<thead>
<tr>
<th>ITEM</th>
<th>Sec. I</th>
<th>Sec. II</th>
<th>Sec. III</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Would not deal with questions covering material beyond text</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I_2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tested primarily for isolated and/or obscure details</td>
<td>0</td>
<td>-.34</td>
<td>-.10</td>
</tr>
<tr>
<td>I_3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presented obsolete material</td>
<td>0</td>
<td>-.56*</td>
<td>0</td>
</tr>
<tr>
<td>I_4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequently read aloud from the textbook</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I_5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presented facts without relating them to one another</td>
<td>-.32</td>
<td>0</td>
<td>-.05</td>
</tr>
<tr>
<td>I_6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class presentations were primarily reiterates of textbook</td>
<td>.08</td>
<td>.54</td>
<td>-.20</td>
</tr>
<tr>
<td>I_7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read extensively from his/her lecture notes</td>
<td>0</td>
<td>-.40</td>
<td>0</td>
</tr>
<tr>
<td>I_8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lacked knowledge of subject being presented</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I_9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presented irrelevant material during lectures</td>
<td>-.04</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**FACTOR III - COMMITMENT TO TEACHING**

<table>
<thead>
<tr>
<th>ITEM</th>
<th>Sec. I</th>
<th>Sec. II</th>
<th>Sec. III</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missed class often due to non-teaching responsibilities</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I_2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Permitted students to disrupt classroom activities</td>
<td>-.08</td>
<td>-.16</td>
<td>.13</td>
</tr>
<tr>
<td>I_3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spoke with poise</td>
<td>.32</td>
<td>-.18</td>
<td>.14</td>
</tr>
<tr>
<td>I_4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remained unruffled by student's questions</td>
<td>-.03</td>
<td>.02</td>
<td>.05</td>
</tr>
<tr>
<td>I_5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complained about his/her teaching assignment</td>
<td>.14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I_6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was late to class</td>
<td>-.07</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I_7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Came to appointments on time</td>
<td>.04</td>
<td>.14</td>
<td>.07</td>
</tr>
</tbody>
</table>
TABLE 2--Continued

<table>
<thead>
<tr>
<th>ITEM</th>
<th>FACTOR IV - EVALUATION</th>
<th>Sec. I</th>
<th>Sec. II</th>
<th>Sec. III</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>Told students what was expected of them</td>
<td>-.53*</td>
<td>.23</td>
<td>-.06</td>
</tr>
<tr>
<td>I_2</td>
<td>Explained how grading was done</td>
<td>0</td>
<td>0</td>
<td>-.22</td>
</tr>
<tr>
<td>I_3</td>
<td>Provided feedback on student work promptly</td>
<td>-.09</td>
<td>0</td>
<td>-.07</td>
</tr>
<tr>
<td>I_4</td>
<td>Gave tests which could be completed within the allotted time</td>
<td>-25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I_5</td>
<td>Gave clear, reasonable assignments</td>
<td>.18</td>
<td>.08</td>
<td>.22</td>
</tr>
<tr>
<td>I_6</td>
<td>Informed students of reports, term papers at the beginning of the course</td>
<td>-.06</td>
<td>.08</td>
<td>.06</td>
</tr>
<tr>
<td>I_7</td>
<td>Stated basis by which grades were determined</td>
<td>.25</td>
<td>0</td>
<td>-.23</td>
</tr>
<tr>
<td>I_8</td>
<td>Announced exams in advance</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ITEM</th>
<th>FACTOR V - VOICE COMMUNICATION</th>
<th>Sec. I</th>
<th>Sec. II</th>
<th>Sec. III</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>Changed pitch, volume or quality of speech</td>
<td>.06</td>
<td>.10</td>
<td>.17</td>
</tr>
<tr>
<td>I_2</td>
<td>Could be heard in all parts of the classroom</td>
<td>0</td>
<td>.23</td>
<td>.08</td>
</tr>
<tr>
<td>I_3</td>
<td>Spoke distinctly</td>
<td>0</td>
<td>.02</td>
<td>.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ITEM</th>
<th>FACTOR VI - OPENNESS</th>
<th>Sec. I</th>
<th>Sec. II</th>
<th>Sec. III</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>Listened to student's problems</td>
<td>.14</td>
<td>-.13</td>
<td>.28</td>
</tr>
<tr>
<td>I_2</td>
<td>Admitted being wrong when shown he/she was in error</td>
<td>-.08</td>
<td>.06</td>
<td>-.18</td>
</tr>
<tr>
<td>I_3</td>
<td>Laughed at his/her own mistakes</td>
<td>-.19</td>
<td>-.20</td>
<td>-.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ITEM</th>
<th>FACTOR VII - CURRENCY OF KNOWLEDGE</th>
<th>Sec. I</th>
<th>Sec. II</th>
<th>Sec. III</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>Dealt with questions covering material beyond text</td>
<td>.07</td>
<td>.23</td>
<td>-.07</td>
</tr>
<tr>
<td>I_2</td>
<td>Introduced new ideas and/or research findings in class</td>
<td>.27</td>
<td>.16</td>
<td>.38</td>
</tr>
<tr>
<td>I_3</td>
<td>Gave references to current publications</td>
<td>.14</td>
<td>.13</td>
<td>.06</td>
</tr>
<tr>
<td>I_4</td>
<td>Presented material as an extension of the text</td>
<td>.27</td>
<td>.19</td>
<td>-.15</td>
</tr>
<tr>
<td>I_5</td>
<td>Asked challenging and/or probing questions</td>
<td>.44*</td>
<td>.10</td>
<td>-.15</td>
</tr>
<tr>
<td>I_6</td>
<td>Stated course objectives</td>
<td>.19</td>
<td>-.21</td>
<td>-.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ITEM</th>
<th>FACTOR VIII - RAPPORT</th>
<th>Sec. I</th>
<th>Sec. II</th>
<th>Sec. III</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>Ridiculed students in front of class</td>
<td>-.07</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I_2</td>
<td>Students could understand professor's vocabulary</td>
<td>.06</td>
<td>.31</td>
<td>.21</td>
</tr>
<tr>
<td>I_3</td>
<td>Students could understand class presentation</td>
<td>-.01</td>
<td>.28</td>
<td>.15</td>
</tr>
<tr>
<td>I_4</td>
<td>Ignored student questions</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I_5</td>
<td>Established and kept office hours for individual conferences</td>
<td>.22</td>
<td>.05</td>
<td>-.30</td>
</tr>
</tbody>
</table>
TABLE 2--Continued

<table>
<thead>
<tr>
<th>ITEM</th>
<th>Sec. I</th>
<th>Sec. II</th>
<th>Sec. III</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_6</td>
<td>Insisted that his/her opinions were the only correct ones</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I_7</td>
<td>Permitted students to complete thought processes</td>
<td>.20</td>
<td>-.14</td>
</tr>
<tr>
<td>I_8</td>
<td>Evaluated each student as an individual</td>
<td>.13</td>
<td>-.54</td>
</tr>
</tbody>
</table>

*P < .05*
Discussion

Results of this study do not support a relationship between student gain, as measured by McGuigan's G factor (McGuigan 1974), and the SECTB. The failure of the multiple regression analysis to yield more than one significant correlation between gain and factors in the SECTB indicates that the combination of factors in the SECTB are not good predictors of gain. These results closely ally themselves with investigators who conclude that studies that relate student ratings of evaluation and gain are inconsistent (Costin, Greenough, and Menges, 1971, Kulik and Kulik 1974). Previous studies have yielded positive correlations (Morsh, Burgess, and Smith 1956, Gessner 1973, Frey 1973, Sullivan and Skanes 1974), negative correlations (Rodin and Rodin 1972, Bendig 1953 b) and no correlations (Bendig 1953 a) between student achievement and student ratings.

The lack of significant results in this study would appear to indicate that if, as McGuigan argues (McGuigan 1974) that student gain, as measured by the G statistic, is a valid measure of teacher effectiveness, then the SECTB, when measured against it, is not a valid instrument for the measure of teacher effectiveness. Also, the relatively large number (sixty percent) of significant correlations between factors on the SECTB in the combined sections two and three, would argue against the division of the evaluation into eight factors. The high agreement between factor one and factors three through eight (correlations ranging from .432 to .864) would indicate that the SECTB
in the case of this instructor could be limited to only two factors, factor one (facilitation of learning) and factor two (obsolescence of presentation). Factor two was negatively correlated with the other seven factors with correlations ranging from -.058 to -.482. However, in section one there were relatively few (approximately twenty one percent) correlations between factors. Further, the correlations between factor two and the remaining seven factors ranged from -.379 to .310. In the case of this instructor, the SECTB would appear to have some validation for the use of factors. These results should be viewed with caution, due to the small number of subjects used and the fact that the study only involved two instructors.

Results of this study support Rodin and Rodin's (1972) conclusion that students are unable to judge teaching effectiveness if the latter is measured by how much they have learned. Previous studies do imply, however, that student evaluations do validly measure student opinion (Guthrie 1949; Costin et al 1971; Ballard, Rearden, and Nelson 1975), regardless of any relationship to other measures.

It would appear, that before further research can be done relating student gain to teacher evaluations, that the inconsistencies in the relation of gain to teacher effectiveness and the relation of teacher evaluations to teacher effectiveness must be resolved. Student evaluations of instruction are now in use in many institutions, but care needs to be taken in their development, application, and use of results (Rodewald and Carroll 1974, Zelby 1974, McGuigan 1974, Harari and Sedeck 1973, Siebring and Schaff 1974).
REFERENCES


Bendig, A. W., Relation of level of course achievement of students, instructor and course ratings in introductory psychology. Educational and Psychological Measurement, 1953, 13, 437-488. (a)

Bendig, A. W., Student achievement in introductory psychology and student ratings of the competence and empathy of their instructors. Journal of Psychology, 1953, 36, 427-433. (b)


Maslow, A. H., and Zimmerman, W., College teaching ability, scholarly activity, and personality, *Journal of Educational Psychology*, 1956, 47, 185-189.


APPENDIX: PRE AND POST TEST INSTRUMENT

1. Psychology is defined as:
   1. Study of the brain
   2. Study of the mind
   3. Study of human and animal behavior
   4. Study of abnormal behavior

2. The test of a hypothesis lies in its:
   1. Common-sense appeal
   2. Precision
   3. Descriptive power
   4. Predictive power

3. Which of the following would be exposed to a special treatment or event?
   1. Control group
   2. Experimental group
   3. Encounter group
   4. Dependent group
   5. Independent group

4. Knowing or understanding something about people
   1. makes you a lay psychologist
   2. means you are aware and perceptive in interpersonal relationships
   3. means you will make a good grade in this course
   4. means you should be a psychologist

5. Which of the following is true concerning animals and psychology?
   1. Animal behavior is solely the province of zoologists.
   2. Animals do not have minds and therefore are of little relevance to psychology.
   3. Many of the same behavioral principles underlie animal and human behavior.
   4. Animal behavior is governed by the same laws as human behavior; therefore we do not need to experiment on humans.

6. The term "development" refers to:
   1. processes characterized by an increase in size
   2. processes characterized by continuous sequential changes
   3. the elaboration and learning of complex motor tasks
   4. increases in body size as one grows older

7. Which of the following most accurately describes the course of human development?
   1. Each individual displays a unique pattern of development
   2. It is totally unlike the development of other animals
   3. The sequence of development differs from culture to culture
   4. It follows an identifiable, orderly sequence
   5. There is no such thing as an orderly pattern of development
6. Critical periods of life have to do with
1. developing ability to do things well
2. practicing skills to gain perfection
3. birth order in the family
4. learning when the time is ripe

9. According to research, growth:
1. is less important in understanding children than maturation.
2. plays very little role in how the child sees himself.
3. proceeds in definite stages although children may widely vary.
4. proceeds in indefinite stages and has little effect on the child.

10. Which of the following is most closely connected to changes in behavior due to physical growth:
1. Imprinting
2. Nurture
3. Development
4. Maturation
5. Learning

11. "We tend to respond to stimuli that resemble those we have been conditioned to", This is an informal way of describing:
1. discrimination in classical conditioning,
2. spontaneous recovery in classical conditioning,
3. stimulus generalization in classical conditioning,
4. stimulus satiation in classical conditioning.

12. In classical conditioning CR means?
1. Classical response
2. Common Response
3. Classical reinforcement
4. Conditioned response
5. Conditioned reinforcement

13. Discrimination is most nearly the opposite of:
1. spontaneous recovery
2. extinction
3. stimulus generalization
4. secondary reinforcement

14. Psychologists define learning as:
1. the processing of stimulus input
2. a change in performance as a result of experience
3. identical with memorization
4. reduction of cognitive dissonance

15. The definition of learning implies:
1. change in behavior
2. practice
3. experience
4. all of the above

16. All learning implies:
1. retaining
2. conditioning
3. creative thinking
4. at least average intelligence
17. Positive transfer of training
   1. applied only to motor-skill learning.
   2. occurs if previous learning facilitates new learning.
   3. occurs only in operant conditioning.
   4. occurs only in classical conditioning.

18. The crowding of trials or practice sessions close together is called
   1. distributed practice
   2. whole learning
   3. massed practice
   4. Gestalt learning

19. A person tries to drive an automobile in England after learning to
    drive in the United States. In England the cars steering wheel is
    on the right, now on the left as in the United States. The person
    tries to shift gears and finds that he has turned on the turn
    signal. This may be an example of:
    1. positive transfer
    2. generalization
    3. negative transfer
    4. reversal shift

20. Incidental learning is considered to have taken place when
    1. several incidents are described to a subject and he remembers them.
    2. subjects are confused and give attention to the wrong aspect
       of a stimulus situation.
    3. learning takes place with no attention being paid to the material
       being learned.
    4. critical incidents provide the stimuli for learning.

21. Deprivation refers to:
    1. a lack of something or being blocked in attainment.
    2. forcing one's ideas upon others.
    3. mental starvation.
    4. learning.

22. One of Freud's outstanding contributions to psychology was his
    emphasis on the powerful role of
    1. ego-integrative motives
    2. unconscious motives
    3. cultural influences and impact
    4. achievement motives
    on personality development

23. The motivation occurring when several ideas are out of harmony with
    each other is known as:
    1. cognitive dissonance
    2. negative goal reduction
    3. cognitive enhancement
    4. dissonant ideation

24. The term for the tendency of the body to maintain a balance among
    internal physiological conditions is:
    1. homodaiety
    2. staticality
    3. homeostasis
    4. physiostasis
25. Motives are:
   1. directly observed
   2. inferred from behavior
   3. independent of behavior
   4. unlearned; drives are learned

26. The tendency to respond positively or negatively to certain persons, objects, or situations is called a/an
   1. goal-directed tendency
   2. secondary emotion
   3. attitude
   4. ambient valence

27. Which of the following is the relationship between frustration and conflict?
   1. Conflict causes frustration
   2. Frustration causes conflict
   3. Conflict and frustration are not related
   4. The terms are synonymous

28. "Lie detectors" detect which of the following most directly?
   1. fear
   2. guilt
   3. anxiety
   4. autonomic changes

29. To be classed as a conflict, the simultaneous arousal of two or more incompatible motives result in
   1. vacillation
   2. a situation to which there is no solution
   3. defense mechanisms
   4. unpleasant emotions

30. The condition that exists when a goal-response suffers interference is:
   1. Aggression
   2. Inhibition
   3. Displacement
   4. Frustration

31. A rough indication of the degree of correlation can be obtained by plotting a/an
   1. frequency polygon
   2. scattergram
   3. inferential diagram
   4. frequency histogram

32. Values on the vertical axis (ordinate) of a histogram are most often associated with
   1. the number of observations or relative frequency
   2. class intervals of the independent variable
   3. the correlation of the dependent and independent variables
   4. the accuracy of the measuring technique

33. The Normal Curve indicates
   1. no correlations
   2. positive correlations
   3. abnormal behavior
   4. relative standing within a group
34. John received the following scores:
   - Math 84 percentile
   - Science 50 percentile
   - Reading 0 standard deviation
   - English -1 standard deviation

In which subject did he do best?
1. Math
2. Science
3. Reading
4. English

35. A histogram is
   1. a measure of central tendency
   2. a bar graph
   3. a measure of variation
   4. a broken line graph

36. Used in moderation, defense mechanisms:
   1. are harmful
   2. lead to neurosis
   3. help to reduce tension
   4. increase anxiety
   5. help to decrease fear

37. Defense mechanisms are not:
   1. mental processes
   2. attempts to relieve anxiety
   3. a form of self-deception
   4. conscious

38. The underlying and fundamental dynamic force in psychoanalytic theory is
   1. unconscious motivation
   2. defensive drive
   3. superiority drive
   4. conflict between the oral and anal stages of development

39. An individual who explains his behavior in such a way as to assign a socially approved motive to it is employing the defense mechanism known as:
   1. repression
   2. reaction formation
   3. projection
   4. rationalization

40. Defense mechanisms
   1. are voluntary, and a person is aware that he is using them.
   2. function to protect the self-concept.
   3. are damaging, because the person is deliberately lying to himself about his feelings.
   4. produce feelings of isolation.

41. Mr. Z sees snakes climbing all over the walls and floor. This is a(n)
   1. delusion
   2. hallucination
   3. conflict
   4. defense mechanism
42. The most frequently occurring psychosis is the
1. affective reaction 3. schizophrenic reaction
2. paranoid reaction 4. involutional reaction

43. Behavior disorders:
1. are usually the result of a specific traumatic event.
2. have natural and understandable origins.
3. begin by using appropriate defense mechanisms.
4. are organic in origin.

44. Whether a person will or will not develop the symptoms of abnormal
behavior depends chiefly on
1. the kinds of frustrations he encounters.
2. the kinds of conflicts he experiences.
3. individual differences in use of defense mechanisms.
4. individual differences in tolerance for stress.

45. A mental illness arising from a gunshot wound in the head would be
characterized as a(n)
1. organic disorder
2. psychoneurosis
3. functional disorder
4. severe psychogenic disorder