CALIFORNIA STATE UNIVERSITY, NORTH RIDGE

STUDENT PERSONALITY AND TEACHER EVALUATION

A thesis submitted in partial satisfaction of the requirements for the degree of Master of Arts in Psychology

by

Kirk Patrick Barrett

January, 1974
The thesis of Kirk Patrick Barrett is approved:

Committee Chairman

California State University, Northridge

December, 1973
ACKNOWLEDGMENTS

The author wishes to thank Dr. Benjamin Mehlman for his patient support and encouragement, and for his valuable help in preparing this thesis.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT.</td>
<td>vi</td>
</tr>
<tr>
<td>INTRODUCTION.</td>
<td>1</td>
</tr>
<tr>
<td>METHOD.</td>
<td>24</td>
</tr>
<tr>
<td>RESULTS</td>
<td>30</td>
</tr>
<tr>
<td>DISCUSSION.</td>
<td>33</td>
</tr>
<tr>
<td>REFERENCES.</td>
<td>39</td>
</tr>
<tr>
<td>APPENDIX.</td>
<td>42</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Average Evaluation, on a Five Point Scale, Received by Each Instructor from the Other Four Instructors on Six Items</td>
<td>26</td>
</tr>
<tr>
<td>2.</td>
<td>Intercorrelations of Four Items Measuring Student's Perception of Instructor on an Authoritarian-Democratic Continuum</td>
<td>28</td>
</tr>
<tr>
<td>3.</td>
<td>Difference in Means for High and Low Groups on Scores of Four Items Measuring Student Perception of Instructor on an Authoritarian-Democratic Continuum</td>
<td>31</td>
</tr>
<tr>
<td>4.</td>
<td>Differences in Means for High and Low Groups on Scores of Overall Rating of Instructors</td>
<td>32</td>
</tr>
</tbody>
</table>
ABSTRACT

STUDENT PERSONALITY AND TEACHER EVALUATION

by

Kirk Patrick Barrett

Master of Arts in Psychology

January, 1974

This study investigated one aspect of the problem of teaching effectiveness. Two approaches, the comparison of teaching methods and student evaluations of teachers, were examined. Studies in both areas have been conflicting and confusing (Anderson, 1959; McKeachie, 1957).

This study measured 118 students in five introductory psychology courses on five personality variables of the Edwards Personal Preference Schedule (1959). Students were placed in high or low groups for each personality variable, based on the score for that variable. Instructors were equated in terms of scoring in the low range of the Rokeach Dogmatism Scale (Rokeach, 1960), a measure of authoritarianism. All students rated their instructor on four items measuring the student's perception of the instructor's place on an authoritarian-democratic continuum. Students also gave an overall rating of the
instructor. High and low groups on each of the variables were compared.

No differences were found between high and low groups for any of the variables in either the perception of the instructor or the overall rating of the instructor.

Possible reasons for the lack of significant differences were discussed, including the problem of controlling teacher variables, and the need for research on broader kinds of personality variables, as opposed to the narrow single and separate variables used.
INTRODUCTION

A persistent problem for teachers has been that of determining what is good or effective teaching. Effective teaching is the primary goal and responsibility of teachers, and yet it has been a most difficult area to define and measure. One generally accepted measure of how well a teacher has taught, is the amount the student has learned. However, the complexity of the problem quickly becomes apparent when the measurement is attempted. The question arises as to how many variables besides the actual teaching are involved, and the relative importance of each variable. Can it be established that some quality of the instructor, of the course, or of the student, has not been partially responsible for the student's learning?

Additionally, the measurement of student learning itself is open to question. Is an exam the only criterion, or even the best criterion, of learning? It may be that more is learned than an exam reveals. Some feel that a student's attitude toward and interest in the subject must be included in the definition of teaching effectiveness.
It may be that these aspects are ultimately more relevant to the effectiveness of teaching than the storing of facts. But whatever the criterion of effective teaching, its measurement has been elusive.

Despite the complexity of the problem, its importance has generated a great deal of thought and research. One major approach to this problem has been to pit different teaching methods against each other and attempt to measure differences in student learning. Another widely used measure of teaching effectiveness has been the use of student ratings of teacher performance.

In searching the literature on these two approaches to teaching effectiveness, the lack of agreement as to their relative effectiveness is obvious. Conflicting conclusions are abundant, and the inability to replicate results is common. It is fair to say that, despite the interest in the subject, no universally acceptable measure of teaching effectiveness has been found. A review of the literature dealing with teaching methods and student evaluations helps to illuminate the complexity of the problem of defining effective teaching.

It is evident that the literature on teaching effectiveness deals almost exclusively with questions of
teaching method and teacher variables. Teacher variables include personality variables as well as issues of knowledge and skill. Issues of student variables, especially personality variables, have been largely ignored. It may be that the failure of both approaches to measure teaching effectiveness is due in part to the fact that the student's response to instruction is determined by a large number of variables, and cannot be attributed solely to the instruction. In particular, it seems likely that student personality variables interact with the instruction. 

This thesis will attempt to investigate the role of student personality variables as they relate to selected teaching variables.

Nachman and Opochinsky (1958) point out that many studies comparing teaching methods show little or no difference in the amount of knowledge gained. It is noted that in most properly conducted experiments, measurements pertain to the experimental period. This is not true of experiments on teaching. It is obviously impossible to rule out other important variables when the teaching experiment is carried out over weeks or months. The authors speculate that if there are actual differences, they may be masked in measurement. For instance, a final exam, which is often used as the dependent measurement, is
affected by many variables other than teaching variables. This source of variance would include motivation, amount of outside study, personality, and environmental pressures, to name some of the more obvious ones. Using class size as their independent variable, Nachman and Opochinsky (1958) predicted that small classes would learn more than large classes during the actual class period, but that these differences would not be apparent on a final exam, due to the effect of variables mentioned above. In order to measure this difference, it was necessary to take measurements immediately after the treatment was administered. Surprise quizzes were given to both groups, and as predicted, the smaller group did better than the larger group on the quizzes, but no difference was found on the final exam.

Nachman and Opochinsky's (1958) method was suggested as a useful paradigm for dealing with the problems inherent in studying teaching methods. However, the method of testing immediately after the treatment is not practical when the teaching variable becomes more abstract than class size. For example, one of the most widely studied of the basic differences in methodological approach has been the authoritarian-democratic variable. This is also known as the directive-nondirective or student centered-
teacher centered continuum. Anderson (1959) describes the authoritarian leader as one who is impersonal, punishing to class members who deviate from his wishes, decides the agenda for the class, decides how labor is to be divided and how work will be done, and judges the soundness of ideas. The democratic leader is one who acts in a friendly, personal manner, allows the class to plan its own agenda and to choose individual tasks, allows students to talk without asking permission, and accepts suggestions from the class on how work shall be done. In brief, the authoritarian teacher makes decisions for the class and remains remote from them. The democratic teacher shares decision-making with the class and is more warm and open with them.

It has been hypothesized by many (Johnson and Smith, 1953; Bills, 1952; Asch, 1951) that student centered (democratic) classes would produce better learning than teacher centered classes. Results have not confirmed this.

Asch (1951) taught an unstructured group which was allowed to determine its own program. No direction was imposed, but suggested readings were offered. Students determined their own grades. Three control groups were used to check against three different hypotheses. These
groups had the usual lecture format, and were assigned homework, tests and a term paper. One of the control groups, which was matched with the experimental group on the basis of entrance tests and GPA, was to control for course content by means of a final exam administered to both. The final exam did not count at all toward the grade in the student centered group. Contrary to Asch's prediction, this control group did better on the final exam than the experimental group. Asch concluded that factual material is not learned as well in nondirective settings. This conclusion is suspect, though not necessarily untrue. The experimental group was made aware that the final exam would in no way affect their grade, and so it is quite evident that the motivational qualities were not the same for both groups. Asch also makes the claim that the nondirective group gained more knowledge of diversified subject matter and did more independent reading.

The remaining two control groups in the Asch (1951) study were intended to examine changes in social attitudes and in emotional adjustment. It was thought that the nondirective group would show greater understanding of others and more liberal social attitudes. However, no differences in social attitudes were found. The Minnesota Multiphasic Personality Inventory (MMPI) was used to
measure emotional adjustment. The nondirective group showed more improvement on the post MMPI than the control. Asch concluded that self-understanding and adjustment are the main benefits of nondirective classes. He notes that some students benefit more than others from nondirective classes, and that it might be helpful to know what student characteristics are most likely to be beneficial in that setting. This would seem to imply that there are also student characteristics for which a directive atmosphere would produce better learning. There has been little research into these student personality variables.

Bills (1952) also tried to show that student centered teaching would be more effective than the usual lecture-discussion format. He matched the experimental and control groups on the basis of IQ. Both groups were given the same objective test based on the content of the text. No significant difference was found in the amount of learning. It was claimed that the student centered class showed more positive attitudes toward both the class and the subject matter, and that the course was of more personal value.

In another study, Johnson and Smith (1953) used two democratically run classes to test four separate hypotheses. The first two hypotheses were that students would
evaluate democratic classes higher than the traditional lecture classes and that students would more readily accept decisions made democratically than those made by the instructor. On both counts results were divided, with one class supporting the hypotheses and the other showing the opposite reaction. Even though one teacher was clearly more successful with the democratic paradigm than the other teacher, the possibility of teacher differences is discounted by the authors. They make no attempt to justify this position. Based solely on what the authors report, the possibility of teacher differences contributing to these mixed results seems clearly indicated.

Johnson and Smith (1953) also predicted that students in the democratic class would develop more democratic attitudes, and that they would learn more. No effect was found for either of the hypotheses.

In each of the studies so far cited, it was pointed out that results of earlier authoritarian-democratic studies have been mixed and confusing. Perhaps the most comprehensive examination of this confusing situation is a review by Anderson (1959) of forty-nine authoritarian-democratic studies. The criteria for effectiveness which Anderson examined were productivity and morale. These have generally been the dependent variables in such
studies. Productivity was defined as the quantity and quality of work done, or, when appropriate, "changes in behavior." Morale was defined as the extent to which group members found the group personally satisfying, and to which members believed that the class or group successfully approached its goals. Anderson notes that the definitions of productivity and morale vary greatly in the studies he reviews. This has obviously been a major problem for researchers trying to compare results of various studies, and for generalizing to other classes or situations.

The studies reviewed by Anderson were not all concerned with classroom application of the authoritarian-democratic problem. (For Anderson's definitions of this continuum, see p. 5.) However, no general findings could be made as to which method produced greater learning or productivity. Though some have found significant statistical differences in favor of one or another, the author thinks it doubtful that differences have any practical or social significance.

Of the studies reviewed by Anderson which were concerned with education, eleven favored student centered classes, eight favored teacher centered classes, and
thirteen found no difference between the two.

The findings as to morale tend also to be confusing. It appears that generally, morale is dependent on situational factors. The general conclusion drawn by Anderson is that:

Democratic leadership is associated with high morale when a primary group goal is social, as in the case of the recreational and fraternal groups, or emotional catharsis, as in the case of the therapy groups. Morale is higher under authoritarian leadership, however, in groups which are primarily committed to some task goal rather than a social-emotional goal (p. 204).

As specifically applied to the educational research, however, it is generally found that morale is higher in learner centered groups. When anxiety is present due to grades being based on a final exam, the generalization does not hold.

Anderson mentions three other factors which contribute to the confusion found in the results. First, there is a general lack of methodological rigor and adequate research design. Also, many of the studies seem to be isolated. Authors do not seem to be familiar with other studies in the area. Second, the operational definitions of leadership styles are inadequate and imprecise. There are many variables involved in leadership, but the
authoritarian-democratic construct is not adequate as an exclusive concept of leadership behavior, and is thus an inadequate basis for research. And third, the words "authoritarian" and "democratic" carry surplus meaning. Anderson believes that the moralistic connotations of the labels have led to poor studies. In particular, a biased portrayal of authoritarian qualities has prevailed. The "bad" qualities of being overly harsh, critical and cold have been overdone. Anderson believes that this bias has operated in most studies.

The overall conclusion of this review, and one which applies to the literature in general, is that there is no evidence that one style of leadership or teaching method is better than another.

A second major approach to this important problem of evaluating teaching effectiveness has been to ask students to evaluate their teachers. Studies in this area have considered two important questions. First, the question of the reliability of student evaluations. This problem deals with the question of whether or not students can agree on a teacher's qualities. Secondly, it must be determined if the student evaluations are valid. That is, whether the qualities evaluated are related to a teacher's
effectiveness. There is some confusion associated with this second question.

Stalnaker and Remmers (1928) asked students to rank teachers on ten teacher traits from the Purdue Scale, and found a correlation greater than +0.9 for three classes. This seemed to show that students agree on important traits in teaching.

However, the authors' conclusion that the evidence shows the Purdue Scale to be valid and reliable is not warranted. While it is clear that student ratings are good measures of student opinion of teacher qualities, the question remains as to whether or not those opinions are valid judgments of teaching effectiveness. It is one of the problems of research on ratings that validity of the rating scales is inferred from their reliability (Vermillion, Leftwich and Remmers, 1963).

Still, researchers have used rating instruments in an attempt to discover some stable dimensions of teaching effectiveness. Stalnaker and Remmers (1928) intended to show this independence of dimensions when they reported that no halo effect could be found for the ten subscales of the Purdue Scale. Bendig (1953) does report a halo effect for nine of the ten subscales, based on a factor analysis.
In a larger study of the Purdue Scale (Heilman and Armentrout, 1936), forty-six teachers were rated by 2,115 students. Teacher ratings on the ten subscales were highly variable which tends to support the belief in the independence of subscales.

Still, the authors were unsuccessful in their search for teacher variables which would correlate with the subscale ratings. No reliable differences were found for either the length of teaching experience or for the age difference of teachers. Additionally, ratings were not significantly affected by class size, severity of grading, measured personality traits, student interest in the course, sex of the teacher or the maturity of the rater.

The problem of validating student ratings of instructors has led to much speculation as to suitable criteria. It has been suggested that one criterion of effectiveness is the continuing influence on the student after he has left the class. McKeachie and Soloman (1958) point out that exams are not the only criterion for teaching effectiveness. Broadened interest in the subject may be a useful yardstick. With this in mind, an attempt was made to validate instructor ratings using as the criterion, the number of students who elected to take further courses
in the area. As in so many other studies, results were confusing.

For five consecutive semesters, students in elementary psychology courses were asked to rate the effectiveness of the instructor and the course. Students for each elementary class were monitored for an additional semester to determine whether they enrolled in any of the available undergraduate courses in psychology. Each instructor was ranked according to the percentage of students who continued in psychology. Instructors were also ranked according to the rating given by students in the elementary course. A correlation was computed between those two rankings. Correlations were significant for only two of the five semesters. Two others were positive but not significant, and one was negative and not significant.

Isaacson, McKeachie, Milholland, Lin, Hofeller, Baerwaldt, and Zinn (1964) identified six stable factors contributing to teaching effectiveness. This was accomplished by a factor analysis of 145 items obtained from many different rating instruments. The six factors identified were skill, overload (difficulty), structure, feedback, group interaction and student-teacher rapport (warmth).
McKeachie, Lin and Mann (1971) devised a rating scale tapping these six dimensions. They used that scale to conduct five validity studies against various criteria. Despite the use of criteria ranging from objective tests to a measure of attitudinal sophistication, results were not clear cut. It was reported that four of the five studies showed that (1) high skill was effective with women, (2) structure was effective with women, but not with men, and (3) warmth was effective in promoting "thinking" as opposed to knowledge.

In spite of the fact that results have been inconclusive and replication of success has not been the rule, most researchers continue to regard student evaluations of teachers as a useful tool in assessing teaching effectiveness. One justification for this is the view that one of the goals of education should be the favorable reaction of the student (McKeachie, 1954). This view points out that it is important that the student be interested enough to continue his education. An inherent danger in this position is the possibility that a teacher's effectiveness might become equated with his popularity. This is discussed by Rodin and Rodin (1972), who conclude that students are not good judges of teaching effectiveness.
They point out that there appears to be a relation between a student's grade and the rating he gives the teacher.

The real problem of defining teaching effectiveness, according to the authors, lies in determining how much the student has learned.

Rodin and Rodin (1972) studied a large, 293 student calculus class and used as a measure of learning forty paradigm problems which covered the entire domain of the course. Students had up to six attempts at each paradigm, and grades were based only on the number correct.

At the end of the course, instructors were rated on a four point scale. With students' initial ability partialed out, a correlation was obtained between the rating and the objective measure of $r = .746$. The correlation without partialing was $r = -.754$.

It is interesting to note that this correlation between grades and ratings was in the reverse direction of previous studies which had shown low positive correlations. The authors speculate that this difference may have been due to the fact that instructors in previous studies may have introduced unconscious bias in constructing their objective tests, and that grading was more subjective than the authors'. Their conclusion is that evaluations by students tend to reflect those qualities which are personal.
and social rather than the instructor's effectiveness.

It is worth noting that, if Rodin and Rodin (1972) have successfully measured teaching effectiveness objectively as the amount learned, then, contrary to their suggestion that ratings are useless, the strong negative correlation would make the teacher ratings quite useful. Furthermore, it would be of interest to consider the possibility that their results may be particular to the subject matter. Carney and McKeachie (1966) have reported more favorable student ratings for "life oriented" courses than for "science oriented" courses. It does seem intuitively reasonable that a directive approach might be more effective in a mathematics class than it would be in a humanities class. The unusually large size of the class is also a factor which might affect results. Neither variable was controlled in this study, and both deserve investigation.

In view of these considerations, and of the great difference in results reported by Rodin and Rodin (1972) from results of other studies, it is essential to replicate their findings before speculating further.

It has been suggested (Asch, 1951; McKeachie, et al, 1971) that much of the confusion may be due to differing
goals of teachers and students. In order to determine teaching effectiveness it is necessary to know what the objectives of the teaching are, and which students are being taught.

Much research has gone into the identification of teacher qualities and teaching method, with the mixed results here presented. There are, of course, a huge number of variables involved in the problem. Student variables comprise one area of variability. Most of the student variables which have been considered in the literature have been demographic characteristics. McKeachie (1954), summarizing some of this information, reports that no differences are found for most demographic variables such as age, sex, veteran - non veteran, course grade or year in college. But very little can be found in the literature as to the personality characteristics of students as they interact with instructor method. One study (Carney and McKeachie, 1966) which promised to investigate the relationship of two personality measures to student ratings, gave only one ambiguous sentence about the results.

One aspect of the problem of the interaction of student personality variables with teaching method, is apparent in the observation that different students can have
different perceptions of a teacher's qualities. While one student may believe a teacher to be fair and open, another may conclude that the same teacher is unfair and cold. Since the same teacher is observed under the same external circumstances by all students, it seems clear that a major factor underlying this difference in perception is difference in the variables the student brings to the situation. It is also quite possible that these differences in perception may account for much of the variability of studies aimed at discovering the most effective means of teaching.

The purpose of this thesis will be to test the idea that students who differ on selected personality variables will (a) perceive an instructor differently on an authoritarian-democratic continuum, and (b) also have different opinions of the overall effectiveness of the teacher.

As an operational definition of the student personality characteristics to be explored, five scales of the Edwards Personal Preference Schedule (EPPS) (Edwards, 1959) were used. Those scales are labeled Achievement (ach), Deference (def), Order (ord), Autonomy (aut), and Sussorance (suc).
The EPPS gives the following definition of Achievement:

**Achievement:** To do one's best, to be successful, to accomplish tasks requiring skill and effort, to be a recognized authority, to accomplish something of great significance, to do a difficult job well, to solve difficult problems and puzzles, to be able to do things better than others, to write a great novel or play (p. 11).

A student who scores high on the ach variable should, according to this definition, be one who is at his best when he is in an atmosphere where he can excel, exhibit his knowledge, and where his extra efforts will be acknowledged and appreciated. This seems more likely to occur in a student centered atmosphere than in a teacher centered atmosphere. Based on these assumptions, the first hypothesis is:

1. Students who score high in Achievement will tend to rate their instructor as more democratic than will those students who are rated low in Achievement, and will also give a higher overall rating of the instructor.

The EPPS gives the following definition of Deference:

**Deference:** To get suggestions from others, to find out what others think, to follow instructions and do what is expected, to praise others, to tell others that they have done a good job, to accept the leadership of others, to read about great men, to conform to custom and avoid the unconventional, to let others make decisions (p. 11).
A student who scores high on the def variable should be most comfortable in a class which is highly structured, in which he can follow instructions which explain clearly what is expected of him, and which does not require that he make decisions. Thus, this student should be more comfortable in an authoritarian class than in a democratic class. This leads to the second hypothesis which is:

2. Students who score high in Deference will tend to rate their instructor as less democratic than will students who are low in Deference, and will also give a higher overall evaluation of the instructor.

The EPPS defines Order as:

Order: To have written work neat and organized, to make plans before starting on a difficult task, to have things organized, to keep things neat and orderly, to make advance plans when taking a trip, to organize details of work, to keep letters and files according to some system, to have meals organized and a definite time for eating, to have things arranged so that they run smoothly without change (p. 11).

A student who scores high on the ord variable should be more comfortable with the structure and discipline offered in an authoritarian class, and be uncomfortable in a less structured democratic class. Hypothesis number three is:

3. Students who score high in Order will tend to rate their instructor as less democratic than will students
rated low in Order, and will also give a higher overall rating of the instructor.

The EPPS defines Autonomy as:

Autonomy: To be able to come and go as desired, to say what one thinks about things, to be independent of others in making decisions, to feel free to do what one wants, to do things that are unconventional, to avoid situations where one is expected to conform, to do things without regard to what others may think, to criticize those in positions of authority, to avoid responsibilities and obligations (p. 11).

A student who scores high on the aut variable is one who does not enjoy being bound by rules. He is most comfortable in an atmosphere where he can freely express himself, even if this puts him in conflict with his teachers. This student would prefer a democratic class to an authoritarian class. The fourth hypothesis is:

4. Students who score high in Autonomy will tend to rate their instructor as more democratic than will students rated low in Autonomy, and will also give a higher overall rating of the instructor.

The EPPS defines Succorance as:

Succorance: To have others provide help when in trouble, to seek encouragement from others, to have others be kindly, to have others be sympathetic and understanding about personal problems, to receive a great deal of affection from others, to have others do favors cheerfully, to be helped by others when
depressed, to have others feel sorry when one is sick, to have a fuss made over one when hurt (p. 11).

A student who scores high on the suc variable would be most comfortable in an atmosphere where the teacher is warm, friendly and shows a sympathetic interest in his problems, and where the teacher is encouraging and supportive. This student would prefer a democratic class to an authoritarian class. The final hypothesis, number five, is:

5. Students who score high in Succorance will tend to rate their instructor as more democratic than will students rated low in Succorance, and will also give a higher overall rating of the instructor.
Subjects were 118 students, forty-three male and seventy-five female, in five Introductory Psychology courses (Psychology 150) at California State University, Northridge. Participation by students was voluntary, but participants received credit in partial fulfillment of course requirements for experimental participation. Procedures took place during regular class sessions.

Also serving as subjects were the instructors of the five classes. All of the instructors, three female and two male, were second year graduate students in psychology at California State University, Northridge, who were participating in a teacher training program. Three of the instructors had taught the same course in the previous semester. Two were teaching the course for the first time, and none of the five had any other teaching experience. As part of the teacher training program, instructors regularly visited each others' classes, and a weekly meeting was held to discuss teaching performance.

As a premeasure of authoritarian-democratic qualities,
the Rokeach Dogmatism Scale Form E (Rokeach, 1960) was administered to all instructors. Scores ranged from +13 to +71, and all were within the low dogmatism range. At the end of the semester, each instructor was asked to rate the other four instructors, based on previous observations, on a six item questionnaire (see appendix). Table 1 shows the average of the four ratings for each instructor for each of the six items.

Considering both instructor measurements, it was assumed that there were no important differences in the style, method or quality of instruction presented to the five classes.

In order to operationally define student perception of the instructor's position on an authoritarian-democratic continuum, four items relating to qualities, as defined by Anderson (1959), were presented to students. The four items presented were:

1. Is he actively helpful when students have difficulty?

2. Does he appear sensitive to students' feelings and problems?

3. Is he flexible?

4. Does he make students feel free to ask questions, disagree, express their beliefs, etc.?
### TABLE 1

**AVERAGE EVALUATION, ON A FIVE POINT SCALE, RECEIVED BY EACH INSTRUCTOR FROM THE OTHER FOUR INSTRUCTORS ON SIX ITEMS**

<table>
<thead>
<tr>
<th>Item</th>
<th>Instructor 1</th>
<th>Instructor 2</th>
<th>Instructor 3</th>
<th>Instructor 4</th>
<th>Instructor 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scholarship</td>
<td>4.5</td>
<td>3.5</td>
<td>3.75</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Organization/Clarity</td>
<td>3.75</td>
<td>3.75</td>
<td>2.75</td>
<td>4.25</td>
<td>2.5</td>
</tr>
<tr>
<td>Instructor-Group</td>
<td>4.5</td>
<td>3.75</td>
<td>3.25</td>
<td>4.5</td>
<td>4.0</td>
</tr>
<tr>
<td>Interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor-Individual</td>
<td>4.0</td>
<td>3.75</td>
<td>3.25</td>
<td>4.75</td>
<td>3.0</td>
</tr>
<tr>
<td>Interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamism/Enthusiasm</td>
<td>4.5</td>
<td>3.25</td>
<td>2.75</td>
<td>4.75</td>
<td>3.5</td>
</tr>
<tr>
<td>Overall Impression</td>
<td>4.0</td>
<td>4.0</td>
<td>2.5</td>
<td>4.5</td>
<td>3.0</td>
</tr>
</tbody>
</table>
An intercorrelation matrix, Table 2, shows that all items are positively correlated, for this sample of subjects, with a range of $r = +0.25$ to $r = +0.71$.

A fifth item was also used to determine the student's overall evaluation of the teacher. That item was:

5. Considering all of the above qualities which are applicable (including any that you've added), how would you rate this teacher?

All of the items were contained in the rating scale developed by McKeachie (1969) which was based on the six factors identified by Isaacson, et al. (1964), described earlier. The entire rating scale was presented to students.

The Edwards Personal Preference Schedule (1959) was administered to students early in the semester. Students were told that their participation was voluntary, and that the results would be used in research for a Master of Arts degree.

Student scores were inspected for five of the personality scales on the EPPS. Those scales, which have already been described, are Achievement, Deference, Order, Autonomy, and Succorance. Students who scored at or above the seventieth percentile on any scale were placed in the
TABLE 2
INTERCORRELATIONS OF FOUR ITEMS MEASURING STUDENT'S PERCEPTION OF INSTRUCTOR ON AN AUTHORITARIAN-DEMOCRATIC CONTINUUM

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.50</td>
<td>.36</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.25</td>
<td>.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
high group for that personality scale. Students who scored at or below the thirtieth percentile on any scale were placed in the low group for that scale.

During the final week of the semester, students were asked to fill out a rating form on their instructor. Ratings on each item were on a ten point scale. Ratings on the four critical items of the teacher evaluation score were summed and that total used as a measure of the student's perception of his teacher's position on an authoritarian-democratic continuum.
RESULTS

It was intended that t tests would be used to test the differences between the means of high and low groups on each of the five variables, both for the scores on the four critical items, and for the scores on the overall rating of the instructors. However, visual inspection of the means makes it very clear that no differences were found. Means for high and low groups in each of the five personality variables are shown in Tables 3 and 4.

The largest difference between means appeared on the Succorance variable. This was only a 1.9 point difference on a possible scale of forty points. A t test was run on this variable yielding a t of 1.09 which was not significant at any level.
TABLE 3

DIFFERENCE IN MEANS FOR HIGH AND LOW GROUPS ON SCORES OF FOUR ITEMS MEASURING STUDENT PERCEPTION OF INSTRUCTOR ON AN AUTHORITARIAN-DEMOCRATIC CONTINUUM

<table>
<thead>
<tr>
<th>Item</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACHIEVEMENT</td>
<td>35.13</td>
<td>35.29</td>
</tr>
<tr>
<td></td>
<td>n=30</td>
<td>n=24</td>
</tr>
<tr>
<td>DEFERENCE</td>
<td>35.39</td>
<td>34.82</td>
</tr>
<tr>
<td></td>
<td>n=28</td>
<td>n=44</td>
</tr>
<tr>
<td>ORDER</td>
<td>35.48</td>
<td>34.69</td>
</tr>
<tr>
<td></td>
<td>n=33</td>
<td>n=39</td>
</tr>
<tr>
<td>AUTONOMY</td>
<td>35.16</td>
<td>35.61</td>
</tr>
<tr>
<td></td>
<td>n=44</td>
<td>n=13</td>
</tr>
<tr>
<td>SUCCORANCE</td>
<td>35.80</td>
<td>33.89</td>
</tr>
<tr>
<td></td>
<td>n=36</td>
<td>n=19</td>
</tr>
</tbody>
</table>
TABLE 4
DIFFERENCES IN MEANS FOR HIGH AND LOW GROUPS ON SCORES OF OVERALL RATING OF INSTRUCTORS

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACHIEVEMENT</td>
<td>8.23</td>
<td>8.58</td>
</tr>
<tr>
<td>DEFERENCE</td>
<td>8.61</td>
<td>8.66</td>
</tr>
<tr>
<td>ORDER</td>
<td>8.66</td>
<td>8.51</td>
</tr>
<tr>
<td>AUTONOMY</td>
<td>8.57</td>
<td>8.92</td>
</tr>
<tr>
<td>SUCCORANCE</td>
<td>8.51</td>
<td>8.74</td>
</tr>
</tbody>
</table>
DISCUSSION

The results do not support the prediction that students differing on selected personality variables will perceive an instructor differently and give a different overall rating of the instructor. It would be useful to examine some of the possible reasons for this.

One area of concern has to do with Anderson's (1959) assertion that the authoritarian-democratic construct is not an adequate concept of leadership, or, particularly when used exclusively, adequate to describe the teacher-student interaction. In spite of the fact that it is so generally accepted as a continuum of leadership style, it is possible that other variables are equally or more important than the authoritarian-democratic variable. For example, the instructor qualities of scholarship and enthusiasm for the subject could certainly be exhibited in any instructor, from the most authoritarian to the most democratic. It is possible that those variables would be so much more important in terms of student satisfaction with the instructor that they would overshadow the
importance of authoritarian-democratic qualities.

Another question is whether or not the instructors were in fact highly similar on the authoritarian-democratic continuum. It is possible that the Rokeach Dogmatism Scale (Rokeach, 1960) is not a sensitive enough measure of this dimension. It is conceivable that the graduate psychology students participating in a closely supervised teaching program in which there was considerable openness to the supervisor and to fellow student-teachers, might have been hypersensitive to a paper and pencil measure of their personal tendencies toward authoritarianism.

A related problem, which is relevant to the whole problem of studying teaching effectiveness, is the problem of controlling teacher variables. This has clearly been a most persistent and difficult problem.

Even if teachers are equated on specific variables, e.g., intelligence, education, etc., there are always many other variables which cannot be equated, and whose effects may be important. For instance, asking five instructors to teach using a democratic method would very likely produce five different interpretations of democratic teaching. A narrow operational definition of democratic teaching would be difficult considering the multitude of possibilities arising from the
instructor-class interaction. If the same teacher were used to present different methods, the problem of unconscious bias exists, and also the probability that, even without bias, the instructor will not be equally good with both methods.

One suggestion for increasing control in studies where method is the variable understudy, is to introduce independent observations of what actually occurred in the classroom. This was not attempted in the present study, but it might have been useful in defining the authoritarian-democratic quality of the classes.

The increasing use of television and other automated teaching aids opens up the possibility of better control over teacher variables for experimental purposes. However, these machines bring new problems. The problem of equating content when presenting different methods is still present. The problem of the lack of student-teacher interaction when these devices are used may introduce problems not yet recognized. It becomes clear that the number of variables involved in teaching is so large that adequate experimental control is frequently not possible at present. A great deal of research is needed in this area to identify and then to isolate and control the variables important to effective teaching.
A second area of concern when considering the lack of support for the hypotheses is the relevance of personality measures in a classroom situation. It seems reasonable that personality variables dealing with achievement and with ways of relating to authority should be relevant to a student's behavior in the classroom. It is possible, however, that classroom behavior is specific to the classroom situation. The lecture-discussion classroom experience may be so well defined, formalized and separated from other aspects of the student's life that predictions from personality variables would not hold. Or, to put it another way, it is possible, particularly in a formalized, well-defined classroom situation, that student personality variables are relatively unimportant. It seems generally true that the more precisely and completely the behavior is defined for a situation by society or the environment, the less likely there will be that wide variation of behavior we often identify as personality.

One question which was not investigated in this thesis was the possibility that particular combinations of personality variables may be more meaningful than single variables, as far as describing the student's personality.

Zuckerman (1958) describes the personality concept of...
rebelliousness, and uses as a measure, a combination of the Autonomy, Dominance and Aggression scales of the EPPS. Bernardin and Jessor (1957) describe a dependency measure as a score at or above the seventieth percentile on the Deference scale of the EPPS, combined with a score at or below the fiftieth percentile on the Autonomy scale. These, or other similar variables which are more encompassing than the single scales, may be more important than the single, isolated EPPS variables studied here. This problem deserves more study.

It is possible that there were differences between high and low groups which did not show up because the groups were not extreme enough. Differences might become more apparent if a larger sample were taken and groups were picked which were at greater extremes. However, this might have little practical value even if differences were found, if the results only applied to the extremes of the continuum.

The results of this study do not support the specific hypotheses dealt with in this study using the specific instruments and population of this study. However, because of the contradictions and confusion in the literature on teaching effectiveness, and the possibilities for improving education inherent in the identification of variables
relating to teaching effectiveness, it is important that research in this area continue.
REFERENCES


Heilman, J. D., and Armentrout, W. D. The rating of college teachers on ten traits by their students. Journal of Educational Psychology, 1936, 27, 197-216.


McKeachie, W. J. Student ratings of faculty: A research review. Improving College and University Teaching, 1957, 5, 4-8.


APPENDIX

Classroom Visit Report Form
for Evaluation of Instructor by Other Instructors
APPENDIX

CLASSROOM VISIT REPORT FORM
FOR EVALUATION OF INSTRUCTOR BY OTHER INSTRUCTORS

Instructors have unique styles. These ratings should be completed with full recognition of the tremendous diversity in teaching approaches. The Scales listed below may not all be equally applicable to this particular instructor, to this particular course, to the objectives of this particular period, etc.

<table>
<thead>
<tr>
<th>Highest</th>
<th>Average</th>
<th>Lowest</th>
<th>Insufficient Opportunity to Observe</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

1. Analytic/Synthetic-relates to scholarship, with emphasis on breadth, analytic ability, and conceptual understanding.

2. Organization/Clarity-relates to skill at presentation, but is subject-related, not student-related, and not concerned merely with rhetorical skill.

3. Instructor-Group Interaction-relates to rapport with the class as a whole, sensitivity to class response, and skill at securing active class participation.

4. Instructor-Individual Interaction-relates to mutual respect and rapport between the instructor and the individual student.

5. Dynamism/Enthusiasm-relates to the flair and infectious enthusiasm that comes with confidence, excitement for the subject, and pleasure in teaching.

6. If you were a student in the instructor's class, how would you evaluate your learning experience?